AIMS on Applied Mathematics Vol. 8

Hyperbolic Problems: Theory, Numerics, Applications

Fabio Ancona
Alberto Bressan
Pierangelo Marcati
Andrea Marson
Editors

American Institute of Mathematical Sciences

AIMS Series on Applied Mathematics Volume 8

Hyperbolic Problems: Theory, Numerics, Applications

Proceedings
of the Fourteenth International Conference
on Hyperbolic Problems
held in Padova,
June 25-29, 2012

Fabio Ancona,
Alberto Bressan,
Pierangelo Marcati,
Andrea Marson
Editors

American Institute of Mathematical Sciences

Fabio Ancona,
Alberto Bressan,
Pierangelo Marcati,
Andrea Marson
Editors

AMS 2000 subject classifications: Primary 35B25, 35B27, 35Lxx, 35M10, 35Q30, 35Q60, 35R35, 49K20, 65C20, 65L06, 65Mxx, 65Nxx, 65Txx, 65Yxx, 65Z05, 74B20, 74Jxx, 76Rxx, 76Txx, 80A32, 80Mxx

ISBN-10: 1-60133-017-0; ISBN-13: 978-1-60133-017-8

Printed in People's Republic of China.

aimsciences.org

PREFACE

This volume contains the Proceedings of the HYP2012 International Conference devoted to Theory, Numerics and Applications of Hyperbolic Problems, held in Padova, June 24–29, 2012. This was the fourteenth in a highly successful series of bi-annual meetings, which brought together several leading experts in the field, practitioners, and young researchers, discussing the latest theoretical advances and the most relevant applications.[1]

Hyperbolic conservation laws is a mathematical discipline deeply rooted in the tradition of classical continuum mechanics, and yet replenished with challenging open problems. It has experienced continued growth in recent years, thanks to the introduction of new ideas and techniques, and a wealth of new applications. The HYP2012 meeting highlighted a number of topics where recent progress has been particularly significant: singular limits and dispersive equations in mathematical physics, nonlinear wave patterns in several space dimensions, particle dynamics, multiphase flow and interfaces, transport in complex environments, control problems for hyperbolic PDEs and related Hamilton-Jacobi equations, general relativity and geometric PDEs.

The conference was attended by 340 participants from 30 different countries. The social program included a boat excursion to the historical Villa Pisani and to Villa Foscari - La Malcontenta, and a conference banquet in the great hall of the 13-th century building "Palazzo della Ragione" in Padova, which was once the seat of the City Council, with frescoes from the Giotto school. During the dinner, Professor Constantine Dafermos, Professor James Glimm, and Professor Tai Ping Liu were honored with the "Galileo medal" for scientific excellence conferred by the Mayor of Padova, Flavio Zanonato. A keynote speech was delivered by Professor James Glimm. Professor Glimm is credited with many pioneering contributions in the general area of the theory and numerics of hyperbolic equations. His speech provided an overview of the field, from its early days to the present time, with an outlook toward the role of hyperbolic PDE models in interdisciplinary science. The conference banquet also featured the brilliant performance of the Marco Castelli quartet, one of the most talented Italian jazz groups, introduced by Silvia Faggian from the University of Venice.

The present volume of proceedings contains 7 papers from plenary speakers, 9 from invited speakers, and 100 papers related to contributed talks. These contributions cover a wide range of topics. A very partial list includes: new methods for constructing turbulent solutions to multi-dimensional systems of conservation laws based on Baire category, transport equations with non-Lipschitz velocity fields, relative entropy functionals and the stability of fluid systems, numerical methods for hyperbolic systems with stiff relaxation terms and for multiphase flow, new advances

[1] The detailed program and the slides of all speakers as well as most of the video of the plenary speakers of the HYP2012 conference can be found on the website http://www.hyp2012.eu/ .The HYP2012 website will be accessible at this address untill 2020.

in homogenization theory, optimal sensor location for solutions to multidimensional wave equations, singularities in general relativity.

We believe that this volume will provide a timely survey of the state of the art, and a stimulus for further progress in this exciting field.

We take this opportunity to thank the members of the HYP2012 Scientific Committee (listed at `http://www.hyp2012.eu/organization/scientific-committee`) for their expertise in the selection of the plenary and invited speakers of the conference and for their contribution in reviewing the papers of the volume. We would like also to express our warm appreciation to all other members of the HYP2012 Organizing Committee (listed at `http://www.hyp2012.eu/organization/organizing-committee`) that in various ways have contributed to the successful realization of this event. Finally, we are extremely thankful to the many graduate students and post-docs of the Dipartimento di Matematica of Università di Padova, coordinated by Khai T. Nguyen and Fabio S. Priuli, for their assistance and dedicated work throughout the conference.

We gratefully acknowledge financial support from the following sponsors:

- ERC Starting Grant 2009 “Hyperbolic Systems of Conservation Laws: singular limits, properties of solutions and control problems”
- Research Project MIUR PRIN 2009 “Systems of Conservation Laws and Fluid Dynamics: methods and Applications”
- GDRE CONEDP European Research Group on PDE control
- FP7-PEOPLE-2010-ITN European Research Network “Sensitivity Analysis for Deterministic Controller Design”
- Progetto di Eccellenza Fondazione Cariparo 2009-2010 Nonlinear PDEs: models, analysis, and control-theoretic problems
- ERC Starting Grant 2010 Traffic Management by Macroscopic Models
- Dipartimento di Matematica, Università di Padova
- Dipartimento di Matematica Pura ed Applicata, Università degli Studi dell’Aquila
- University of Padova
- University of Zurich
- University of Basel

Fabio Ancona
Alberto Bressan
Piero Marcati
Andrea Marson

CONTENTS

Preface i

Part 1: Plenary Lectures 1–107

Camillo De Lellis, Elisabetta Chiodaroli and Ondřej Kreml
Surprising solutions to the isentropic Euler system of gas dynamics 1

Eduard Feireisl
Relative entropies, dissipative solutions, and singular limits of complete fluid systems 11

David Gérard-Varet and Nader Masmoudi
Recent progress in the theory of homogenization with oscillating Dirichlet data 29

Franz Fuchs, Andrew McMurry, Nils Henrik Risebro and Siddhartha Mishra
Explicit and implicit finite volume schemes for radiation MHD and the effects of radiation on wave propagation in stratified atmospheres 41

Sebastiano Boscarino and Giovanni Russo
Implicit-explicit Runge-Kutta schemes for hyperbolic systems with stiff relaxation and applications 61

Robert Saye and James Sethian
New interface methods for tracking multiphase multiphysics 81

Yannick Privat, Emmanuel Trélat and Enrique Zuazua
Optimal sensor location for wave and Schrödinger equations 89

Part 2: Invited Lectures 109–232

Gianluca Crippa
Ordinary differential equations and singular integrals 109

Jean-Luc Guermond and Bojan Popov
Entropy viscosity for the Euler equations and questions regarding parabolic regularization 119

Soohyun Bae, Sun-Ho Choi and Seung-Yeal Ha
Dynamic instability of the Vlasov-Poisson-Boltzmann system in high dimensions 125

Rinaldo M. Colombo and Michael Herty
Nodal conditions for hyperbolic systems of balance laws 147

Corrado Lattanzio and Athanasios E. Tzavaras
Relative entropy methods for hyperbolic and diffusive limits 163

Pierre Degond, Amic Frouvelle and Jian-Guo Liu
A note on phase transitions for the Smoluchowski equation with dipolar potential 179

Katrin Grunert, Helge Holden and Xavier Raynaud
Lipschitz metric for the two-component Camassa–Holm system 193

Moritz Reintjes
Points of shock wave interaction are 'regularity singularities' in spacetime 209

Michael Shearer
Two fluid flow in porous media 221

Part 3: Contributed Talks 233–1066

Eduardo Abreu
Numerical simulation of three-phase flow in heterogeneous media with spatially varying nonlinear hyperbolic-parabolic flux functions 233

Franz Achleitner, Sabine Hittmeir and Christian Schmeiser
On nonlinear conservation laws regularized by a Riesz-Feller operator 241

Fabio Ancona, Olivier Glass and Khai T. Nguyen
On quantitative compactness estimates for hyperbolic conservation laws 249

Boris Andreianov
One-dimensional conservation law with boundary conditions: General results and spatially inhomogeneous case 259

Denise Aregba-Driollet
The Riemann problem for Kerr equations and non-uniqueness of selfsimilar entropy solutions 269

Thomas Auphan, Philippe Angot and Olivier Guès
Penalty methods for edge plasma transport in a tokamak 277

Hantaek Bae and Konstantina Trivisa
On the Doi model for the suspensions of rod-like molecules in compressible fluids 285

Luís Almeida and Patrizia Bagnerini
Multi-scale tissular-cellular model for wound healing 293

Joshua Ballew
Low Mach number limits to the Navier-Stokes-Smoluchowski system 301

Caroline Bauzet
Method for solving a stochastic conservation law 309

Stefan Berres, Frank Huth, Hartmut Schwandt and Günter Bärwolff
A multi-phase model for pedestrian flow with strategic and tactic component .317

Luigi C. Berselli
Some results on the two-dimensional dissipative Euler equations 325

Raul Borsche and Jochen Kall
ADER-schemes on networks of scalar conservation laws 333

Laurent Boudin and Julien Mathiaud
Asymptotic behavior of a diffusive scheme solving the inviscid one-dimensional pressureless gases system 341

Benjamin Boutin, Frédéric Coquel and Philippe G. LeFloch
Coupling techniques for nonlinear hyperbolic equations 349

Alberto Bressan
Globally optimal and Nash equilibrium solutions for traffic flow on networks . 357

M. Briani, D. Aregba-Driollet and R. Natalini
How to improve the decay of the numerical error for large times: The case of dissipative BGK systems 365

Buğra Kabil
On the asymptotics of solutions to resonator equations 373

Raimund Bürger, Pep Mulet and Luis M. Villada
Adaptive mesh refinement for spectral WENO schemes for efficient simulation of polydisperse sedimentation processes 381

Sunčica Čanić and Boris Muha
A nonlinear moving-boundary problem of parabolic-hyperbolic-hyperbolic type arising in fluid-multi-layered structure interaction problems 389

Giovanni Alberti, Stefano Bianchini and Laura Caravenna
Reduction on characteristics for continuous solutions of a scalar balance law . . 399

Filipe Carvalho and Ana Jacinta Soares
Detonation wave problems: Modeling, numerical simulations and linear stability 407

Pablo Castañeda, Dan Marchesin and Frederico Furtado
On singular points for convex permeability models 415

Fausto Cavalli
Linearly implicit schemes for convection-diffusion equations 423

Nikolai Chemetov and Fernanda Cipriano
The inviscid limit for slip boundary conditions 431

Koottungal Revi Arun, Guoxian Chen and Sebastian Noelle
A finite volume evolution Galerkin scheme for acoustic waves in heterogeneous media 439

Fabio Ancona and Giuseppe Maria Coclite
On the asymptotic stabilization of a generalized hyperelastic-rod wave equation 447

Armando Coco and Giovanni Russo
Boundary treatment in ghost point finite difference methods for compressible gas dynamics in domain with moving boundaries 455

Rinaldo M. Colombo, Mauro Garavello, Magali Lécureux-Mercier and Nikolay Pogodaev
Conservation laws in the modeling of moving crowds 467

Andrea Corli and Haitao Fan
A hyperbolic model for phase transitions in porous media 475

Kristian Debrabant and Espen R. Jakobsen
Semi-Lagrangian schemes for linear and fully non-linear Hamilton-Jacobi-Bellman equations 483

A.I. Delis and I.K. Nikolos
On a solution reconstruction and limiting procedure for unstructured finite volumes 491

Maria Laura Delle Monache and Paola Goatin
A strongly coupled PDE-ODE system modeling moving density

constraints in traffic flow 501

Bruno Després, Christophe Buet and Emmanuel Franck
A priori analysis of asymptotic preserving schemes with the modified equation 509

Liviu Florin Dinu
Wave-wave interactions of a gasdynamic type 515

Donatella Donatelli and Pierangelo Marcati
Analysis of oscillations and defect measures in plasma physics 525

Renjun Duan
Asymptotic stability of kinetic plasmas for general collision potentials 533

Arnaud Duran, Fabien Marche, Christophe Berthon and Rodolphe Turpault
Numerical discretizations for shallow water equations with source terms on unstructured meshes 541

Jan Ernest
Schemes with Well-Controlled Dissipation (WCD) for scalar conservation laws with pseudo-parabolic regularization 551

Adriano Festa and Maurizio Falcone
L^1 convergence of a SL scheme for the eikonal equation with discontinuous coefficients 559

Hermano Frid
A brief note on recent developments on divergence-measure fields 567

Raul Borsche, Rinaldo M. Colombo and Mauro Garavello
Mixed systems with boundaries 575

Mohamed-Gazibo Karimou
Degenerate convection-diffusion equation with a Robin boundary condition 583

Christoph Gersbacher
Explicit higher order schemes for the coupling of dimensionally heterogeneous free-surface flow models 591

Jan Giesselmann and Athanasios E. Tzavaras
On cavitation in elastodynamics 599

Christophe Chalons, Mathieu Girardin and Samuel Kokh
Operator-splitting-based asymptotic preserving scheme for the gas dynamics equations with stiff source terms 607

Christian Bourdarias, Marguerite Gisclon and Stéphane Junca
Blow-up or not blow-up at the hyperbolic boundary for a system from chemistry? 615

Hyunkyung Lim, Yijie Zhou, Valmor F. de Almeida and James Glimm
Fully developed turbulent mixing in an annular sector 623

Frédéric Coquel, Edwige Godlewski and Nicolas Seguin
Fluid systems: Approximation by relaxation and coupling 631

Debora Amadori, Rinaldo M. Colombo, Graziano Guerra and Wen Shen
Slow erosion of granular flow: Continuous and discontinuous profiles 641

Maren Hantke and Ferdinand Thein
Numerical solutions to the Riemann problem for compressible isothermal Euler equations for two phase flows with and without phase transition 651

Itsuko Hashimoto
Asymptotic behavior of solutions for damped wave equations with non-convex convection term on the half line 659

Boris Haspot
Porous media equations, fast diffusions equations and the existence of global weak solution for the quasi-solutions of compressible Navier-Stokes equations . 667

Christiane Helzel and Marsha J. Berger
Cartesian grid embedded boundary methods for hyperbolic problems 675

Zhenying Hong, Guangwei Yuan and Xuedong Fu
Time and space discrete scheme to suppress numerical solution oscillation for the neutron transport equations 685

François James and Nicolas Vauchelet
Numerical simulation of a hyperbolic model for chemotaxis after blow up 693

Quansen Jiu, Yi Wang and Zhouping Xin
Global well-posedness of 2D compressible Navier-Stokes equations with large data and vacuum 701

Pierre Castelli and Stéphane Junca
Oscillating waves and optimal smoothing effect for one-dimensional nonlinear scalar conservation laws 709

Dante Kalise
A WENO-TVD finite volume scheme for the approximation of atmospheric phenomena 717

Friedemann Kemm
An enhancement to the AUFS flux splitting scheme by Sun and Takayama ... 725

K. Sudarshan Kumar, G. D. Veerappa Gowda and C. Praveen
On polymer flooding problem in oil reservoir simulation 733

Christian Klingenberg and Ujjwal Koley
An error estimate for viscous approximate solutions to degenerate anisotropic convection-diffusion equations 741

Michael Benfield, Helge Kristian Jenssen and Irina A. Kogan
1-D conservative systems: A geometric approach 749

Ujjwal Koley and Nils Henrik Risebro
Convergence of a finite difference scheme for 2 × 2 Keyfitz-Kranzer system ... 759

Mirko Kränkel
A local discontinuous Galerkin scheme for compressible phase field flow 767

Gerhard Dziuk, Dietmar Kröner and Thomas Müller
Existence and uniqueness for scalar conservation laws on moving hypersurfaces 775

Alexei A. Mailybaev
Renormalization and universal structure of blowup in 1D conservation laws . . . 783

Vítor Matos, Pablo Castañeda and Dan Marchesin
Classification of the umbilic point in immiscible three-phase flow in porous media . . . 791

Thomas Müller and Axel Pfeiffer
Well-balanced simulation of geophysical flows via the shallow water equations with bottom topography: Consistency and numerical computation . . . 801

Martin Nolte
Approximation of the effective Hamiltonian through a degenerate elliptic problem . . . 809

Shinya Nishibata, Masashi Ohnawa and Masahiro Suzuki
Nonlinear stability of boundary layer solutions to the Euler-Poisson equations in plasma physics . . . 817

Sigrun Ortleb and Andreas Meister
A well-balanced DG scheme with unconditionally positive implicit time integration . . . 823

Evgeniy Yu. Panov
On decay of periodic renormalized solutions to scalar conservation laws . . . 831

Marica Pelanti and Keh-Ming Shyue
A mixture-energy-consistent numerical approximation of a two-phase flow model for fluids with interfaces and cavitation . . . 839

Carlo Morosi, Mario Pernici and Livio Pizzocchero
A posteriori estimates for Euler and Navier-Stokes equations . . . 847

Ramón G. Plaza
On the stability of degenerate viscous shock profiles . . . 857

Giacomo Dimarco, Lorenzo Pareschi and Vittorio Rispoli
Asymptotically implicit schemes for the hyperbolic heat equation . . . 865

Volker Elling and Joseph Roberts
Steady self-similar inviscid flow . . . 873

Panters Rodríguez-Bermúdez and Dan Marchesin
Loss of strict hyperbolicity for vertical three-phase flow in porous media . . . 881

Rinaldo M. Colombo, Paola Goatin and Massimiliano D. Rosini
On the management of vehicular and pedestrian flows . . . 889

Olga Rozanova
Exact solutions with singularities to ideal hydrodynamics of inelastic gases . . . 899

Ingenuin Gasser, Martin Rybicki and Winnifried Wollner
Modelling, simulation and optimization of gas dynamics in an exhaust pipe . . . 907

Jochen Schütz, Michael Woopen and Georg May
A combined hybridized discontinuous Galerkin / hybrid mixed method for viscous conservation laws . . . 915

Gabriella Puppo and Matteo Semplice
Finite volume schemes on 2D non-uniform grids . . . 923

Maria Lukáčová-Medvid'ová and Nikolaos Sfakianakis
Theoretical study of entropy dissipation of moving meshes 931

Zhi-Qiang Shao
Almost global existence of classical discontinuous solutions to general quasilinear hyperbolic systems of conservation laws with small BV initial data 941

Julio Daniel Silva and Dan Marchesin
Riemann solutions without intermediate constant states for a system of conservation laws 949

Gianluca Crippa, Carlotta Donadello and Laura V. Spinolo
A note on the initial-boundary value problem for continuity equations with rough coefficients 957

Stefano Spirito and Luigi C. Berselli
On inviscid limits for the Navier-Stokes equations with slip boundary conditions involving the vorticity 967

Marta Strani
Metastable dynamics of interfaces for the hyperbolic Jin-Xin system 975

Andrea Terracina
Entropy formulation for forward–backward parabolic equations 983

Dmitry Tkachev and Aleksander Blokhin
Stability of a supersonic flow about a wedge with a weak shock wave satisfying the Lopatinski condition 991

Yuri Trakhinin
Existence and stability of relativistic plasma-vacuum interfaces 999

Alessandro Morando, Paola Trebeschi and Yuri Trakhinin
The linearized plasma-vacuum interface problem in ideal incompressible MHD 1007

Knut Waagan
Relaxation-based approximate Riemann solvers for elastic rods 1015

Jean-Michel Coron and Zhiqiang Wang
Control of a scalar conservation law with a nonlocal velocity 1023

Nils Henrik Risebro and Franziska Weber
Burgers' equation with a filtered velocity 1031

Chi-Kun Lin and Kung-Chien Wu
Anelastic approximation of the Gross-Pitaevskii equation for general initial data 1039

Dmitry Kotov, H.C. Yee and Bjorn Sjögreen
Comparative study of high-order positivity-preserving WENO schemes 1047

Lorenzo Zanelli
Mather measures in semiclassical analysis 1059

Part 1

Plenary Lectures

SURPRISING SOLUTIONS TO THE ISENTROPIC EULER SYSTEM OF GAS DYNAMICS

Camillo De Lellis, Elisabetta Chiodaroli
and Ondřej Kreml

Institut für Mathematik, Universität Zürich
Winterthurerstrasse 160
8056 Zürich, Switzerland

Abstract. In a recent paper, jointly with Elisabetta Chiodaroli and Ondřej Kreml we consider the Cauchy problem for the isentropic compressible Euler system in 2 space dimensions, with initial data which assume two different constant values and have a discontinuity across a line. If we consider selfsimilar solutions we then encounter a classical 1-dimensional Riemann problem for the corresponding hyperbolic system of conservation laws. We show that for some suitable choice of the pressure and of the initial data there exist infinitely many bounded admissible solutions which are not selfsimilar and indeed are genuinely 2-dimensional. We also show that some of these Riemann data are generated by a 1-dimensional compression wave. Our theorem leads therefore to Lipschitz initial data for which there are infinitely many global bounded admissible weak solutions. Each of these solutions coincide as long as the classical (Lipschitz) solution exists and they differentiate themselves immediately after the first blow-up time. Our approach is heavily influenced by a work of László Székelyhidi which provides a similar result in the case of the classical vortex-sheet problem for the incompressible Euler equations.

1. **Introduction.** Consider the isentropic compressible Euler equations of gas dynamics in n space dimensions. This system consists of $n+1$ scalar equations, which state the conservation of mass and linear momentum. The unknowns are the density ρ and the velocity v and the system takes the the form:

$$\begin{cases} \partial_t \rho + \mathrm{div}_x(\rho v) \;=\; 0 \\ \\ \partial_t(\rho v) + \mathrm{div}_x\,(\rho v \otimes v) + \nabla_x[p(\rho)] \;=\; 0 \end{cases} \tag{1}$$

The pressure p is a function of ρ determined from the constitutive thermodynamic relations of the gas under consideration and it is assumed to satisfy $p' > 0$ (this hypothesis guarantees also the hyperbolicity of the system on the regions where ρ is positive). A common choice is the polytropic pressure law $p(\rho) = \kappa\rho^\gamma$ with constants $\kappa > 0$ and $\gamma > 1$. The classical kinetic theory of gases predicts exponents $\gamma = 1 + \frac{2}{d}$, where d is the degree of freedom of the molecule of the gas.

2000 *Mathematics Subject Classification.* Primary: 36L65; Secondary: 36L67, 35L45.

Key words and phrases. Hyperbolic systems of conservation laws, Riemann problem, admissible solutions, ill-posedness, convex integration.

Camillo De Lellis has been supported by the SNF grant 129812. Ondřej Kreml's research has been financed by the SCIEX Project 11.152.

A lot of attention has been devoted in the literature to the *Cauchy problem* which consists of solving (1) on a domain of the form $\mathbb{R}^2 \times [0, T[$ (where T might also be infinite), subject to an initial condition of type

$$\left\{\begin{array}{lll} \rho(\cdot, 0) & = & \rho^0 \\ \\ v(\cdot, 0) & = & v^0 \, . \end{array}\right. \tag{2}$$

It is well known that, even starting from extremely regular initial data, the solutions of the Cauchy problem for the system (1) develop singularities in finite time. It is also well-known that after the appearance of the first singularity weak solutions (i.e. solutions in the usual distributional sense, see Definition 2.1 for the precise formulation) are not unique: the standard example is provided by "non-physical" shocks, which can however be ruled out imposing that the weak solutions satisfy some further admissibility condition. Much effort has been put in understanding how this approach can give well-posedness results after the appearance of the first singularity, leading to a quite mature and successful theory in one space dimension (we refer the reader to the monographs [1],[8] and [19]).

Here we consider the case of two space dimensions and restrict our attention to bounded weak solutions of (1) which satisfy the following additional inequality in the sense of distributions (called usually *entropy inequality*, although for the specific system (1) this is rather a weak form of the energy balance):

$$\partial_t \left(\rho\varepsilon(\rho) + \rho\frac{|v|^2}{2}\right) + \mathrm{div}_x \left[\left(\rho\varepsilon(\rho) + \rho\frac{|v|^2}{2} + p(\rho)\right) v\right] \;\leq\; 0 \tag{3}$$

where the internal energy $\varepsilon : \mathbb{R}^+ \to \mathbb{R}$ is given through the law $p(r) = r^2\varepsilon'(r)$. Indeed, admissible solutions are required to satisfy a slightly stronger condition, i.e. a form of (3) which involves also the initial data, see Definition 2.2.

Starting from the work [12] it was observed that (3) is in this case not enough to restore uniqueness of admissible *bounded* solutions. The methods used in [12], inspired by techniques developed in the theory of differential inclusions, show a rather surprising abundance of admissible solutions to the Cauchy problem with certain particular initial data. However those specific initial data were rather irregular, leaving open the question whether this fact alone was responsible for such behavior.

The investigations of [12] have been pushed further in [5] and in [6]: in the latter paper, we have shown that the same nonuniqueness result holds even for *Lipschitz* initial data, therefore leading to the following theorem.

Theorem 1.1. *Let $p(\rho) = \rho^2$. Then there are* Lipschitz *initial data ρ^0 and v^0, with $\rho^0 \geq c_0 > 0$ for which there are* infinitely many *admissible bounded weak solutions (ρ, v) of the Cauchy problem (1)-(2), with $\inf \rho > 0$. All these solutions coincide with the classical one as long as it exists and differ immediately after the formation of the first singularity.*

2. **Main results.** We recall here the usual definitions of weak and admissible solutions to (1).

Definition 2.1. By a *weak solution* of (1)-(2) on $\mathbb{R}^2 \times [0, \infty[$ we mean a pair $(\rho, v) \in L^\infty(\mathbb{R}^2 \times [0, \infty[)$ such that the following identities hold for every test functions $\psi \in C^\infty_c(\mathbb{R}^2 \times [0, \infty[, \mathbb{R})$, $\phi \in C^\infty_c(\mathbb{R}^2 \times [0, \infty[, \mathbb{R}^2)$:

$$\int_0^\infty \int_{\mathbb{R}^2} [\rho\partial_t\psi + \rho v \cdot \nabla_x\psi] \, dxdt + \int_{\mathbb{R}^2} \rho^0(x)\psi(x, 0)dx \;=\; 0 \tag{4}$$

$$\int_0^\infty \int_{\mathbb{R}^2} [\rho v \cdot \partial_t \phi + \rho v \otimes v : D_x \phi + p(\rho) \operatorname{div}_x \phi] + \int_{\mathbb{R}^2} \rho^0(x) v^0(x) \cdot \phi(x,0) dx \;=\; 0. \tag{5}$$

Definition 2.2. A bounded weak solution (ρ, v) of (1)-(2) is admissible if it satisfies the following inequality for every nonnegative test function $\varphi \in C_c^\infty(\mathbb{R}^2 \times [0,\infty[)$:

$$\begin{aligned}\int_0^\infty \int_{\mathbb{R}^2} &\left[\left(\rho\varepsilon(\rho) + \rho\frac{|v|^2}{2}\right)\partial_t\varphi + \left(\rho\varepsilon(\rho) + \rho\frac{|v|^2}{2} + p(\rho)\right) v \cdot \nabla_x \varphi\right] \\ &+ \int_{\mathbb{R}^2} \left(\rho^0(x)\varepsilon(\rho^0(x)) + \rho^0(x)\frac{|v^0(x)|^2}{2}\right)\varphi(x,0)\, dx \;\geq\; 0\,. \end{aligned}\tag{6}$$

The following is then the theorem proved in [12]:

Theorem 2.3 (De Lellis - Székelyhidi)**.** *For any $p \in C^1$ with $p' > 0$ there are pairs $\rho^0, v^0 \in L^\infty$ such that there are infinitely many bounded admissible solutions (ρ, v) of (1)-(2) with $\inf \rho > 0$.*

As already mentioned, the initial data constructed in [12] were however very irregular. It was then proved by Chiodaroli that indeed the ill-posedness of Theorem 2.3 still holds even if ρ^0 is regular. More precisely

Theorem 2.4 (Chiodaroli)**.** *For any $p \in C^1$ with $p' > 0$ and any $\rho^0 \in C^1$ with $\inf \rho^0 > 0$ there is $v^0 \in L^\infty$ such that there are infinitely many bounded admissible solutions (ρ, v) of (1)-(2) with $\inf \rho > 0$.*

In [6] we consider first initial data of a very particular form. We denote the space variable as $x = (x_1, x_2) \in \mathbb{R}^2$ and set

$$(\rho^0(x), v^0(x)) := \begin{cases} (\rho_-, v_-) & \text{if } x_2 < 0 \\ (\rho_+, v_+) & \text{if } x_2 > 0, \end{cases}\tag{7}$$

where $\rho_\pm, v_\pm$ are constants.

It is well-known that for some special choices of these constants there are solutions of (1) which are *rarefaction waves*, i.e. self-similar solutions depending only on t and x_2 which are locally Lipschitz for positive t and constant on lines emanating from the origin (see [8, Section 7.6] for the precise definition). Reversing their order (i.e. exchanging $+$ and $-$) the very same constants allow for a *compression wave* solution, i.e. a solution on $\mathbb{R}^2 \times]-\infty, 0[$ which is locally Lipschitz and converges, for $t \uparrow 0$, to the jump discontinuity of (7). When this is the case we will then say that the data (7) are *generated by a classical compression wave.*

It follows from the usual treatment of the 1-dimensional Riemann problem that for data as in (7) uniqueness holds if the admissible solutions are also required to be self-similar, i.e. of the form $(\rho, v)(x,t) = \left(r\left(\frac{x_2}{t}\right), w\left(\frac{x_2}{t}\right)\right)$, and to have locally bounded variation. In fact in this case the solutions are obtained "gluing" together rarefaction waves and jump discontinuities across interfaces of type $\{(x,t) : x_2 = \nu t\}$.

In the paper [6] we show the existence of bounded admissible solutions which are *not* selfsimilar. Although we expect such solutions to exist for very general pressure laws, we show them only for some particular choice of the pressure p.

Theorem 2.5 (Chiodaroli-De Lellis-Kreml)**.** *There are smooth pressures p with $p' > 0$ and constants $\rho_\pm, v_\pm$ for which, if (ρ^0, v^0) are as in (7), then there are infinitely many bounded admissible solutions (ρ, v) of (1)-(2) with $\inf \rho > 0$.*

Among the pressure laws of Theorem 2.5 there is also the quadratic law $p(\rho) = \rho^2$. The strongest results of [6] are indeed proved fur such law. More precisely we have the following strengthened version of Theorem 2.5.

Theorem 2.6 (Chiodaroli-De Lellis-Kreml)**.** *Assume $p(\rho) = \rho^2$. Then there are constants $\rho_\pm, v_\pm$ for which the conclusion of Theorem 2.5 holds and such that (ρ^0, v^0) are generated by a classical compression wave.*

Theorem 1.1 is then a simple corollary of Theorem 2.6: the solutions of Theorem 1.1 are simply obtained "patching" a classical compression wave with the nonstandard solutions of Theorem 2.6.

3. ***h*-principle and differential inclusions.** The proof of Theorem 1.1 relies heavily on the works of the first author and László Székelyhidi, who in the paper [11] introduced methods from the theory of differential inclusions to explain the existence of compactly supported nontrivial weak solutions of the *incompressible* Euler equations (discovered in the pioneering work of Scheffer [20]; see also [21]). Indeed the paper [12] is based on the observation that these methods could be applied to the compressible Euler equations and lead to the ill-posedness of bounded admissible solutions, see [12].

The link with the incompressible Euler equations is provided by the following elementary remark.

Remark 1. Assume $\Omega \subset \mathbb{R}^2 \times \mathbb{R}$ and let (ρ, v) be a distributional solution of (1) with constant density ρ. Then the pair $(v, 0)$ is a weak solution of the incompressible Euler equations

$$\left\{\begin{array}{l} \partial_t v + \operatorname{div} v \otimes v + \nabla q = 0 \\ \\ \operatorname{div} v = 0\,. \end{array}\right. \tag{8}$$

Or in other words a "pressureless" solution, where $q = 0$: note however that q could be set to be any given constant.

Although classical solutions of the incompressible Euler equations with constant pressure are rather rare, the methods of [11] show that there are many such weak solutions. In fact the constraints posed by the equations for weak solutions are so much weaker than those posed for classical solutions, that all these irregular ones can be constructed to satisfy the additional constraint $|v| =$ const.. In particular these methods yield the following crucial lemma (cf. with [6, Lemma 3.7]; here $\mathcal{S}_0^{2\times 2}$ denotes the set of symmetric traceless 2×2 matrices and Id is the identity matrix).

Lemma 3.1. *Let $(\tilde{v}, \tilde{u}) \in \mathbb{R}^2 \times \mathcal{S}_0^{2\times 2}$ and $C > 0$ be such that*

$$\tilde{v} \otimes \tilde{v} - \tilde{u} < \frac{C}{2} \mathrm{Id}\,. \tag{9}$$

For any open set $\Omega \subset \mathbb{R}^2 \times \mathbb{R}$ there are infinitely many maps $(\underline{v}, \underline{u}) \in L^\infty(\mathbb{R}^2 \times \mathbb{R}, \mathbb{R}^2 \times \mathcal{S}_0^{2\times 2})$ with the following property

(i) $\underline{v}$ and $\underline{u}$ vanish identically outside Ω;
(ii) $\operatorname{div}_x \underline{v} = 0$ and $\partial_t \underline{v} + \operatorname{div}_x \underline{u} = 0$;
(iii) $(\tilde{v} + \underline{v}) \otimes (\tilde{v} + \underline{v}) - (\tilde{u} + \underline{u}) = \frac{C}{2}\mathrm{Id}$ a.e. on Ω.

For the relevance of the condition (9) and the techniques used to prove these type of theorems we refer the reader to the survey article [13]: we give here just a brief comment. Observe that, inside Ω, the pair $(v,u) = (\tilde{v} + \underline{v}, \tilde{u} + \underline{u})$ solves the linear identities

$$\begin{cases} \partial_t v + \operatorname{div} u = 0 \\ \\ \operatorname{div} v = 0 \end{cases} \tag{10}$$

and the algebraic constraint

$$u = v \otimes v - \frac{C}{2}\mathrm{Id}\,. \tag{11}$$

Since C is a constant, plugging (11) into (10) we actually conclude that v is a solution (in Ω) of (8) with constant pressure (such constant being free for us to decide). Moreover, since u is trace-free, we conclude that $|v|^2$ equals the constant C. So, (9) can be interpreted as a relaxation of (11), i.e. an inequality which would be automatically satisfied by any weak limit of sequences of solutions as above. The methods of [11]-[12] essentially show that for any such "candidate weak limit" there are indeed many sequences of *exact solutions* converging to it (although such solutions are rather irregular).

4. **The main geometric idea.** Coming back to *compressible Euler* consider now any constant $\rho_0 > 0$. The pair $(\rho, v) = (\rho_0, \tilde{v} + \underline{v})$ is then a weak solution of (1) in Ω, as it can be easily verified by the identities

$$\partial_t \rho_0 + \operatorname{div}(\rho_0 v) = \rho_0 \operatorname{div} v$$

and

$$\partial_t(\rho_0 v) + \operatorname{div}(\rho_0 v \otimes v) + \nabla[p(\rho_0)] = \rho_0(\partial_t v + \operatorname{div} v \otimes v)\,.$$

In fact it also an admissible solution: since

$$\rho_0 \varepsilon(\rho_0) + \rho_0 \frac{|v|^2}{2}$$

and

$$\rho_0 \varepsilon(\rho_0) + \rho_0 \frac{|v|^2}{2} + p(\rho_0)$$

are both constants, (3) amounts to $\operatorname{div} v = 0$.

Observe however that the pair (ρ, v) ceases to give a solution of compressible Euler on the *whole* space-time. Assume now to chop $\mathbb{R}^2 \times \mathbb{R}$ into finitely many open subsets Ω_i and repeat on each Ω_i the construction of the previous section, starting from arbitrary constants for v, u and ρ. Define a resulting pair (ρ, v) by setting it equal, in each separate Ω_i, to the various functions given by Lemma 3.1 (in particular we are free to set the constants ρ_i for the value of the pressure). Although (ρ, v) is an admissible solution of (1) in each separate open set, it might fail to do so on the entire space-time. However, a careful computation shows that in order to be an admissible solution on the entire space-time, we just need to satisfy some compatibility conditions at the interfaces, which are reminiscent of the Rankine-Hugoniot conditions. We observe that some care is needed: due to the oscillatory nature of the solutions, the traces of $|v|^2$ do not coincide with the moduli squared of the traces of v!

The relevant computations show that these compatibility conditions depend only on the chosen "starting" constants. We are therefore ready to give an outline of the main idea behind the construction in [6], which indeed stems out of several conversations of the authors with László Székelyhidi and it is inspired by his work [23].

Consider first some data as in (7). We then partition the upper half space $\{t > 0\}$ in regions contained between half-planes meeting all at the line $\{t = x_2 = 0\}$, see Definition 5.1 and cf. Figure 1. We then define the density function $\rho = \overline{\rho}$ to be constant in each region: this density function will indeed give the final ρ for all the solutions we construct and it is therefore required to take the constant values $\rho_\pm$ in the outermost regions $P_\pm$.

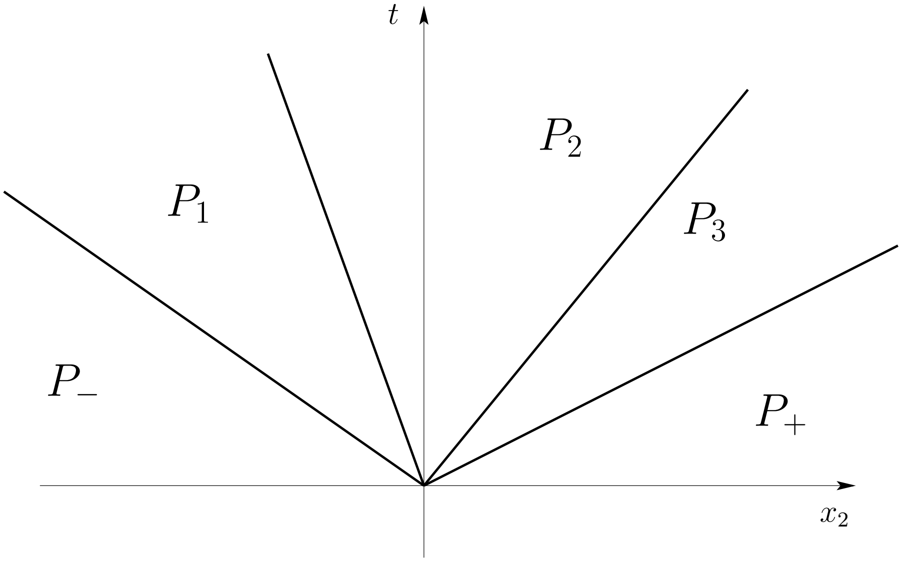


Figure 1. A "fan partition" in five regions.

We then solve the compressible Euler equations (1) in each region $P_1, \ldots, P_N$ using Lemma 3.1, so imposing that the modulus of the velocity is constant (in each region): its square will be denoted by C_i.

The corresponding constant values (ρ_i, v_i, u_i) will then give a globally defined (piecewise constant) function $(\overline{\rho}, \overline{v}, \overline{u})$, which will be called a *fan subsolution* of the compressible Euler equations. We then wish to choose our subsolution so that, after solving (1) in each region P_i with the methods of [12], the resulting globally defined (ρ, v) are admissible *global* solutions of (1). This leads to a suitable system of PDEs for the piecewise constant functions $(\overline{\rho}, \overline{v}, \overline{u})$ which are summarized in the Definitions 5.2 and 5.3.

5. **Subsolutions.** The approach sketched in the previous section leads to the following rigorous definitions (cf. Definitions 3.3, 3.4 and 3.5 in [6]).

Definition 5.1 (Fan partition)**.** A *fan partition* of $\mathbb{R}^2\times]0,\infty[$ consists of finitely many open sets $P_-, P_1, \ldots, P_N, P_+$ of the following form

$$P_- = \{(x,t) : t > 0 \quad \text{and} \quad x_2 < \nu_- t\} \tag{12}$$
$$P_+ = \{(x,t) : t > 0 \quad \text{and} \quad x_2 > \nu_+ t\} \tag{13}$$
$$P_i = \{(x,t) : t > 0 \quad \text{and} \quad \nu_{i-1} t < x_2 < \nu_i t\} \tag{14}$$

where $\nu_- = \nu_0 < \nu_1 < \ldots < \nu_N = \nu_+$ is an arbitrary collection of real numbers.

Definition 5.2 (*Fan Compressible* subsolutions)**.** A *fan subsolution* to the compressible Euler equations (1) with initial data (7) is a triple $(\overline{\rho}, \overline{v}, \overline{u}) : \mathbb{R}^2\times]0,\infty[\to (\mathbb{R}^+, \mathbb{R}^2, \mathcal{S}_0^{2\times 2})$ of piecewise constant functions satisfying the following requirements.

(i) There is a fan partition $P_-, P_1, \dots, P_N, P_+$ of $\mathbb{R}^2\times]0,\infty[$ such that

$$(\overline{\rho}, \overline{v}, \overline{u}) = \sum_{i=1}^{N} (\rho_i, v_i, u_i)\mathbf{1}_{P_i} + (\rho_-, v_-, u_-)\mathbf{1}_{P_-} + (\rho_+, v_+, u_+)\mathbf{1}_{P_+}$$

where ρ_i, v_i, u_i are constants with $\rho_i > 0$ and $u_\pm = v_\pm \otimes v_\pm - \frac{1}{2}|v_\pm|^2\mathrm{Id}$;

(ii) For every $i \in \{1, \dots, N\}$ there exists a positive constant C_i such that

$$v_i \otimes v_i - u_i < \frac{C_i}{2}\mathrm{Id}\,. \tag{15}$$

(iii) The triple $(\overline{\rho}, \overline{v}, \overline{u})$ solves the following system in the sense of distributions:

$$\partial_t \overline{\rho} + \mathrm{div}_x(\overline{\rho}\,\overline{v}) \;=\; 0 \tag{16}$$

$$\partial_t(\overline{\rho}\,\overline{v}) + \mathrm{div}_x\,(\overline{\rho}\,\overline{u}) + \nabla_x \left(p(\overline{\rho}) + \frac{1}{2}\left(\sum_i C_i \rho_i \mathbf{1}_{P_i} + \overline{\rho}|\overline{v}|^2 \mathbf{1}_{P_+\cup P_-} \right) \right) = 0 \tag{17}$$

Definition 5.3 (Admissible fan subsolutions)**.** A fan subsolution $(\overline{\rho}, \overline{v}, \overline{u})$ is said to be *admissible* if it satisfies the following inequality in the sense of distributions

$$\partial_t \left(\overline{\rho}\varepsilon(\overline{\rho})\right) + \mathrm{div}_x \left[\left(\overline{\rho}\varepsilon(\overline{\rho}) + p(\overline{\rho})\right)\overline{v}\right] + \partial_t \left(\overline{\rho}\frac{|\overline{v}|^2}{2}\mathbf{1}_{P_+\cup P_-} \right) + \mathrm{div}_x \left(\overline{\rho}\frac{|\overline{v}|^2}{2}\overline{v}\mathbf{1}_{P_+\cup P_-} \right)$$
$$+ \sum_{i=1}^{N} \left[\partial_t \left(\rho_i \frac{C_i}{2}\mathbf{1}_{P_i} \right) + \mathrm{div}_x \left(\rho_i\,\overline{v}\,\frac{C_i}{2}\mathbf{1}_{P_i} \right) \right] \;\le\; 0\,. \tag{18}$$

The discussion of the previous section can then be summarized in the following proposition.

Proposition 5.4. *Let p be any C^1 function and $(\rho_\pm, v_\pm)$ be such that there exists at least one admissible fan subsolution $(\overline{\rho}, \overline{v}, \overline{u})$ of (1) with initial data (7). Then there are infinitely many bounded admissible solutions (ρ, v) to (1)-(7) such that $\rho = \overline{\rho}$.*

6. **The algebra.** As already mentioned, the various conditions given in the above definitions can be easily reduced to Rankine-Hugoniot conditions on the (flat) interfaces dividing the various regions. As shown in [6] it suffices consider fan subsolutions with a fan partition consisting of only three sets, namely P_-, P_1 and P_+: this rather restrictive assumption is already enough to show that subsolutions exist.

We introduce therefore the real numbers $\alpha, \beta, \gamma, \delta, v_{-1}, v_{-2}, v_{+1}, v_{+2}$ such that

$$v_1 = (\alpha, \beta), \tag{19}$$
$$v_- = (v_{-1}, v_{-2}) \tag{20}$$
$$v_+ = (v_{+1}, v_{+2}) \tag{21}$$
$$u_1 = \begin{pmatrix} \gamma & \delta \\ \delta & -\gamma \end{pmatrix}. \tag{22}$$

We are now ready to report the algebraic conditions that such numbers must satisfy and which correspond to Proposition 5.1 in [6].

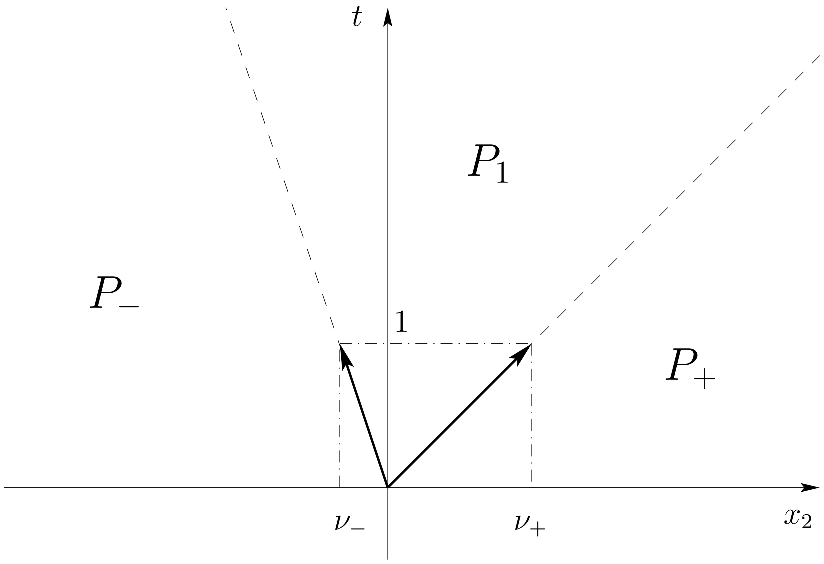


FIGURE 2. The fan partition in three regions.

Proposition 6.1. *Let $N = 1$ and P_-, P_1, P_+ be a fan partition as in Definition 5.1. The constants $v_1, v_-, v_+, u_1, \rho_-, \rho_+, \rho_1$ as in (19)-(22) define an admissible fan subsolution as in Definitions 5.2-5.3 if and only if the following identities and inequalities hold:*

- *Rankine-Hugoniot conditions on the left interface:*

$$\nu_-(\rho_- - \rho_1) = \rho_- v_{-2} - \rho_1 \beta \tag{23}$$

$$\nu_-(\rho_- v_{-1} - \rho_1 \alpha) = \rho_- v_{-1} v_{-2} - \rho_1 \delta \tag{24}$$

$$\nu_-(\rho_- v_{-2} - \rho_1 \beta) = \rho_- v_{-2}^2 + \rho_1 \gamma + p(\rho_-) - p(\rho_1) - \rho_1 \frac{C_1}{2}\,; \tag{25}$$

- *Rankine-Hugoniot conditions on the right interface:*

$$\nu_+(\rho_1 - \rho_+) = \rho_1 \beta - \rho_+ v_{+2} \tag{26}$$

$$\nu_+(\rho_1 \alpha - \rho_+ v_{+1}) = \rho_1 \delta - \rho_+ v_{+1} v_{+2} \tag{27}$$

$$\nu_+(\rho_1 \beta - \rho_+ v_{+2}) = -\rho_1 \gamma - \rho_+ v_{+2}^2 + p(\rho_1) - p(\rho_+) + \rho_1 \frac{C_1}{2}\,; \tag{28}$$

- *Subsolution condition:*

$$\alpha^2 + \beta^2 < C_1 \tag{29}$$

$$\left(\frac{C_1}{2} - \alpha^2 + \gamma\right)\left(\frac{C_1}{2} - \beta^2 - \gamma\right) - (\delta - \alpha\beta)^2 > 0\,; \tag{30}$$

- *Admissibility condition on the left interface:*

$$\begin{aligned} &\nu_-(\rho_- \varepsilon(\rho_-) - \rho_1 \varepsilon(\rho_1)) + \nu_- \left(\rho_- \frac{|v_-|^2}{2} - \rho_1 \frac{C_1}{2}\right) \\ &\le [(\rho_- \varepsilon(\rho_-) + p(\rho_-)) v_{-2} - (\rho_1 \varepsilon(\rho_1) + p(\rho_1))\beta] + \left(\rho_- v_{-2} \frac{|v_-|^2}{2} - \rho_1 \beta \frac{C_1}{2}\right)\,; \end{aligned} \tag{31}$$

- *Admissibility condition on the right interface:*

$$\nu_+(\rho_1\varepsilon(\rho_1)-\rho_+\varepsilon(\rho_+))+\nu_+\left(\rho_1\frac{C_1}{2}-\rho_+\frac{|v_+|^2}{2}\right)$$
$$\leq [(\rho_1\varepsilon(\rho_1)+p(\rho_1))\beta-(\rho_+\varepsilon(\rho_+)+p(\rho_+))v_{+2}]+\left(\rho_1\beta\frac{C_1}{2}-\rho_+v_{+2}\frac{|v_+|^2}{2}\right). \quad (32)$$

Although there seems to be an abundance of constants satisfying the requirements of the proposition above, it has proved rather difficult to find an efficient way of finding them. A large portion of the paper [6] is spent to give two different methods to generate *some* constants fulfilling the inequalities and identities (23)-(32).

7. **Specific solutions.** The first of these methods makes the specific choice $p(\rho)=\rho^2$. It is with this specific pressure law that we reach Theorem 2.6 and hence our main result Theorem 1.1. More precisely we show that there are constants satisfying the requirements of Proposition 6.1 for which, in addition, the initial data (7) is generated by a compression wave. Such data are in fact easy to characterize, following classical computations.

Lemma 7.1. *Let* $0<\rho_-<\rho_+$, $v_+=(-\frac{1}{\rho_+},0)$ *and* $v_-=(-\frac{1}{\rho_+},2\sqrt{2}(\sqrt{\rho_+}-\sqrt{\rho_-}))$. *Then there is a pair* $(\rho,v)\in W^{1,\infty}_{loc}\cap L^\infty(\mathbb{R}^2\times]-\infty,0[,\mathbb{R}^+\times\mathbb{R}^2)$ *such that*

(i) $\rho_+\geq\rho\geq\rho_->0$;

(ii) The pair solves the hyperbolic system

$$\begin{cases}\partial_t\rho+\mathrm{div}_x(\rho v) = 0\\ \partial_t(\rho v)+\mathrm{div}_x(\rho v\otimes v)+\nabla_x[p(\rho)] = 0\end{cases} \quad (33)$$

with $p(\rho)=\rho^2$ *in the classical sense (pointwise a.e. and distributionally);*

(iii) for $t\uparrow 0$ *the pair* $(\rho(\cdot,t),v(\cdot,t))$ *converges pointwise a.e. to* (ρ^0,v^0) *as in* (7);

(iv) $(\rho(\cdot,t),v(\cdot,t))\in W^{1,\infty}$ *for every* $t<0$.

A clever choice of some of the constants combined with some careful algebraic computations show then the following

Lemma 7.2. *Let* $p(\rho)=\rho^2$. *There exist* $\rho_\pm,v_\pm$ *satisfying the assumptions of Lemma 7.1 and* $\rho_1,C_1,v_1,u_1,\nu_\pm$ *satisfying the algebraic identities and inequalities* (23)-(32).

REFERENCES

[1] A. Bressan, "Hyperbolic systems of conservation laws. The one-dimensional Cauchy problem", Oxford University Press, Oxford, 2000.

[2] T. Buckmaster, C. De Lellis and L. J. Székelyhidi, *Transporting microstructure and dissipative Euler flows*, Preprint (2013)

[3] G.Q. Chen and H. Frid, *Uniqueness and asymptotic stability of Riemann solutions for the compressible Euler equations*, Trans. Amer. Math. Soc., **353** (2001), 1103–1117.

[4] G.Q. Chen and H. Frid, *Extended divergence-measure fields and the Euler equations for gas dynamics*, Comm. Math. Phys., **236** (2003), 251–280.

[5] E. Chiodaroli, *A counterexample to well-posedeness of entropy solutions to the compressible Euler system*, Preprint (2011)

[6] E. Chiodaroli, C. De Lellis and O. Kreml, *Global ill-posedness of the isentropic system of gas dynamics*, Preprint (2013)

[7] D. Cordoba, D. Faraco and F. Gancedo, *Lack of uniqueness for weak solutions of the incompressible porous media equation*, Arch. Ration. Mech. Anal., **200** (2011), 725–746.

[8] C. M. Dafermos, "Hyperbolic conservation laws in continuum physics", Third edition. Springer, Berlin, 2010.

[9] S. Daneri, *Cauchy problem for dissipative Hölder solutions to the incompressible Euler equations*, Preprint (2013).

[10] C. De Lellis, *Notes on hyperbolic systems of conservation laws and transport equations*, in "Handbook of differential equations: evolutionary equations. Vo. III" (eds. C. M. Dafermos and E. Dafermos), (2007), 277–382.

[11] C. De Lellis and L. J. Székelyhidi, *The Euler equations as a differential inclusion*, Ann. Math., **170** (2009), 1417–1436.

[12] C. De Lellis and L. J. Székelyhidi, *On admissibility criteria for weak solutions of the Euler equations*, Arch. Ration. Mech. Anal., **195** (2010), 225–260.

[13] C. De Lellis and L. J. Székelyhidi, *The h-principle and the equations of fluid dynamics*, Bull. Amer. Math. Soc. (N.S.), **49** (2012), 347–375 (2012)

[14] C. De Lellis and L. J. Székelyhidi, *Continuous dissipative Euler flows*, Preprint (2012). To appear in *Inv. Math.*

[15] C. De Lellis and L. J. Székelyhidi, *Dissipative Euler flows and Onsager's conjecture*, Preprint (2012).

[16] R. DiPerna, *Uniqueness of solutions to hyperbolic conservation laws*, Indiana Univ. Math. J., **28** (1979), 137–188.

[17] P. Isett *Hölder continuous Euler flows in three dimensions with compact support in time*, Preprint (2012).

[18] B.L. Keyfitz and H. C. Kranzer, *Existence and uniqueness of entropy solutions to the Riemann problem for hyperbolic systems of two nonlinear conservation laws*, J. Differential Equations, **27** (1978), 444–476.

[19] D. Serre, "Systems of conservation laws. 1. Hyperbolicity, entropies, shock waves", Cambridge University Press, Cambridge, 1999.

[20] V. Scheffer, *An inviscid flow with compact support in space-time*, J. Geom. Anal., **3** (1993), 343–401.

[21] A. Shnirelman, *On the nonuniqueness of weak solution of the Euler equation*, Comm. Pure Appl. Math., **50** (1997) 1261–1286.

[22] R. Shvidkoy, *Convex integration for a class of active scalar equations*, J. Amer. Math. Soc., **24** (2011), 1159–1174.

[23] L. J. Székelyhidi, *Weak solutions to the incompressible Euler equations with vortex sheet initial data*, C. R. Acad. Sci. Paris Ser.I, **349** (2011), 1063–1066.

[24] L. J. Székelyhidi, *Relaxation of the incompressible porous media equation*, Preprint (2011). To appear in *Ann. l'ENS.*

E-mail address: `camillo.delellis@math.uzh.ch`
E-mail address: `elisabetta.chiodaroli@math.uzh.ch`
E-mail address: `ondrej.kreml@math.uzh.ch`

RELATIVE ENTROPIES, DISSIPATIVE SOLUTIONS, AND SINGULAR LIMITS OF COMPLETE FLUID SYSTEMS

Eduard Feireisl

Institute of Mathematics of the Academy of Sciences of the Czech Republic
Žitná 25, 115 67 Praha 1, Czech Republic
and
Charles University in Prague, Faculty of Mathematics and Physics
Mathematical Institute
Sokolovská 83, 186 75 Praha 8, Czech Republic

Abstract. We discuss the role of relative entropies in the analysis of complete fluid systems. The relative entropy, or rather relative energy functional measures the "distance" between a weak solution of a given system of equations and any other trajectory ranging in the same function space. We introduce a relative entropy functional for the full Navier-Stokes-Fourier system based on the ballistic free energy and discuss possible applications in the mathematical analysis of singular limits.

1. **Introduction.** The method of relative entropies has been widely used in rather different areas of the modern theory of partial differential equations, see Berthelin and Vasseur [3], Carrillo [5], Dafermos [7], Saint-Raymond [31], among others. To introduce the concept of *relative entropy*, we consider an abstract (infinite-dimensional) dynamical system generated by the solution operator of the evolutionary problem

$$\frac{\mathrm{d}}{\mathrm{d}t} U(t) = \mathcal{A}(t, U(t)),\ t > 0,\ U(0) = U_0, \tag{1}$$

where $\mathcal{A}$ is a (non-linear) generator. We suppose that the problem (1) admits a (not necessarily) unique solution U ranging in a Banach space X. Here, we suppose that U is a kind of generalized (weak) solution and the space X chosen as large as possible. In the applications studied in the present paper, the system (1) will be a system of partial differential equations governing the time evolution of a fluid, while U is its distributional solution. Besides, we introduce a target space for regular (smooth) solutions $Y \subset X$.

We say that a functional

$$\mathcal{E}\left(U \middle| V\right) : X \times Y,\ Y \subset X \to R \tag{2}$$

is a *relative entropy* for the problem (1) if $\mathcal{E}$ enjoys the following properties:

2000 *Mathematics Subject Classification.* Primary: 35Q30, 35B25; Secondary: 35Q31.

Key words and phrases. Relative entropy, Navier-Stokes-Fourier system, low Mach number limit, inviscid limit.

The author acknowledges the support of the project LL1202 in the programme ERC-CZ funded by the Ministry of Education, Youth and Sports of the Czech Republic.

- **Distance property.** We have $\mathcal{E}(U|V) \geq 0$ and
$$\mathcal{E}\left(U\Big|V\right) = 0 \text{ only if } U \equiv V.$$
- **Lyapunov functional.** Let V be an *equilibrium solution* of the system (1), meaning
$$\mathcal{A}(t, V) = 0 \text{ for all } t.$$
Then $V \in Y$ and
$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}\left(U(t)\Big|V\right) \leq 0 \tag{3}$$
for any (weak) solution U of (1).
- **Gronwall inequality.** Let U be a (weak) solution of the system (1) ranging in the space X and V a more regular (strong) solution of the same problem ranging in the space Y. Then
$$\mathcal{E}\left(U(\tau)\Big|V(\tau)\right) \leq \mathcal{E}\left(U(0)\Big|V(0)\right) + c\int_0^\tau \mathcal{E}\left(U(t)\Big|V(t)\right)\ \mathrm{d}t \text{ for a.a. } \tau \geq 0. \tag{4}$$

Possessing a relative entropy provides a valuable piece of information concerning a given system of equations, in particular in the case when the latter is known to admit only global-in-time weak solutions - the situation typical for the Navier-Stokes system and related problems posed in the natural $3D$-topology, see Fefferman [13]. With a relative entropy at hand, it is possible to introduce the concept of *dissipative* solution and show the principle of *weak-strong uniqueness*. Specifically, the weak (dissipative) and strong solution coincide as long as the latter exists, meaning, the strong solutions are unique in the class of weak solutions - this is a direct consequence of (4). Another application of the relative entropy discussed in the present paper is the rigorous justification of several singular limits in fluid mechanics, in particular in the cases where viscosity becomes negligible.

The paper is organized as follows. In the first part, consisting in Sections 2 - 4 we introduce the concept of relative entropy and dissipative solutions to the Navier-Stokes-Fourier system describing the motion of a general compressible, viscous, and heat conducting fluid and compare it to the quantity introduced by Dafermos [7] in the context of hyperbolic conservation laws. Section 5 is devoted to the analysis of singular limits of the scaled problem by means of the method of relative entropies, in particular, the case of the inviscid incompressible limit. We present frequency localized Strichartz estimates for the acoustic equation and extend the result of [16] to more general physical domains.

2. **Thermostatics, relative entropies.**

To begin, we review some basic concepts of continuum fluid mechanics. We suppose that the *state* of a fluid in thermodynamic *equilibrium* is fully determined by its *mass density* ϱ and the *absolute temperature* ϑ. Alternatively, we may also replace ϱ by the *specific volume* $V = 1/\varrho$ and ϑ by the *internal energy* e. The internal energy e, the *pressure* p, and the *entropy* s satisfy *Gibbs' equation*:
$$\vartheta Ds = De + pDV,\ V = \frac{1}{\varrho}. \tag{5}$$

In this section, we discuss relative entropies $\eta(\varrho, \vartheta|\tilde{\varrho}, \tilde{\vartheta})$ relating the thermostatic variables ϱ, ϑ to some reference values $\tilde{\varrho}$, $\tilde{\vartheta}$.

2.1. **Thermodynamic stability.** The concept of relative entropy in hyperbolic systems of conservation laws was proposed by Dafermos [7] in order to study the stability issues. Following [7], we consider first the standard entropy $s = s(V, e)$ expressed as a function of the specific volume V and the internal energy e. Furthermore, we impose the hypothesis of *thermodynamic stability*:

$$\frac{\partial p(\varrho,\vartheta)}{\partial \varrho} > 0, \ \ \frac{\partial e(\varrho,\vartheta)}{\partial \vartheta} > 0, \tag{6}$$

where the former condition expresses positive *compressibility* of the fluid, while the latter enforces positivity of the *specific heat at constant volume*. Both conditions are rather natural and form one of the main building blocks of the theory developed below.

Expressing the thermodynamic functions p, s, as well as the absolute temperature ϑ in terms of ϱ, e we deduce from (5), (6) that the mapping

$$(V, e) \mapsto -s(V, e) \text{ is convex in } (V, e),$$

cf. Bechtel, Rooney, and Forest [2]. Consequently, a natural candidate for the relative entropy evaluated in terms of the thermostatic variables V, e is the quantity η,

$$\eta\left(V, e \middle| \tilde{V}, \tilde{e}\right) = -\Big(s(V,e) - \partial_V s(\tilde{V},\tilde{e})(V - \tilde{V}) - \partial_e s(\tilde{V},\tilde{e})(e - \tilde{e}) - s(\tilde{V},\tilde{e})\Big).$$

Going back to the independent variables ϱ, ϑ and using Gibbs' relation (5) we obtain

$$\eta\left(\varrho,\vartheta\middle|\tilde{\varrho},\tilde{\vartheta}\right) = -\left(s(\varrho,\vartheta) - \frac{p(\tilde{\varrho},\tilde{\vartheta})}{\tilde{\vartheta}}\left(\frac{1}{\varrho} - \frac{1}{\tilde{\varrho}}\right) - \frac{1}{\tilde{\vartheta}}\Big(e(\varrho,\vartheta) - e(\tilde{\varrho},\tilde{\vartheta})\Big) - s(\tilde{\varrho},\tilde{\vartheta})\right),$$

which may be viewed as a "specific" relative entropy related to unit mass. For applications to conservation laws, it is more convenient to replace $\eta \approx \varrho\eta$, specifically we take

$$\eta\left(\varrho,\vartheta\middle|\tilde{\varrho},\tilde{\vartheta}\right) = -\varrho\left(s(\varrho,\vartheta) - \frac{p(\tilde{\varrho},\tilde{\vartheta})}{\tilde{\vartheta}}\left(\frac{1}{\varrho} - \frac{1}{\tilde{\varrho}}\right) - \frac{1}{\tilde{\vartheta}}\Big(e(\varrho,\vartheta) - e(\tilde{\varrho},\tilde{\vartheta})\Big) - s(\tilde{\varrho},\tilde{\vartheta})\right) \tag{7}$$

2.2. **Ballistic free energy.** In applications to dissipative equations like the Navier-Stokes system, we further modify the functional η by introducing:

$$\begin{aligned} \xi\left(\varrho,\vartheta\middle|\tilde{\varrho},\tilde{\vartheta}\right) &= \tilde{\vartheta}\ \eta\left(\varrho,\vartheta\middle|\tilde{\varrho},\tilde{\vartheta}\right) \\ &= \Big(\varrho e(\varrho,\vartheta) - \tilde{\vartheta}\varrho s(\varrho,\vartheta)\Big) - \Big(\tilde{\varrho} e(\tilde{\varrho},\tilde{\vartheta}) - \tilde{\vartheta}\tilde{\varrho} s(\tilde{\varrho},\tilde{\vartheta})\Big) \\ &\quad + \left[\frac{p(\tilde{\varrho},\tilde{\vartheta})}{\tilde{\varrho}} + e(\tilde{\varrho},\tilde{\vartheta}) - \tilde{\vartheta} s(\tilde{\varrho},\tilde{\vartheta})\right](\tilde{\varrho} - \varrho). \end{aligned}$$

Consequently, using once more Gibbs'relation (5) we arrive at

$$\xi\left(\varrho,\vartheta\middle|\tilde{\varrho},\tilde{\vartheta}\right) = H_{\tilde{\vartheta}}(\varrho,\vartheta) - \frac{\partial H_{\tilde{\vartheta}}(\tilde{\varrho},\tilde{\vartheta})}{\partial \varrho}(\varrho - \tilde{\varrho}) - H_{\tilde{\vartheta}}(\tilde{\varrho},\tilde{\vartheta}), \tag{8}$$

where we have introduced another thermodynamic potential called *ballistic free energy* (cf. Ericksen [11]),

$$H_{\tilde{\vartheta}}(\varrho,\vartheta) = \varrho\Big(e(\varrho,\vartheta) - \overline{\vartheta} s(\varrho,\vartheta)\Big),$$

see [17]. Note that ξ has the physical dimension of *energy* rather than *entropy.*

3. **Fluids in motion.** Up to now, we have considered fluids in thermodynamic equilibrium characterized by the thermostatic variables ϱ, ϑ. Now, we suppose that the fluid moves with a *macroscopic* velocity $\mathbf{u} = \mathbf{u}(t,x)$, which is a function of the time t and the spatial position x. In accordance with the commonly accepted principles of continuum thermodynamics, we assume that the state of the fluid at each instant t is still described by the density $\varrho = \varrho(t,x)$ and the absolute temperature $\vartheta = \vartheta(t,x)$. Thus the trio $[\varrho, \vartheta, \mathbf{u}]$ provides a full description of the fluid at any time and any spatial position of a given physical domain $\Omega \subset R^3$.

3.1. **Navier-Stokes-Fourier system.** Given the initial state of the fluid

$$\varrho(0,\cdot) = \varrho_0, \ \mathbf{u}(0,\cdot) = \mathbf{u}_0, \ \vartheta(0,\cdot) = \vartheta_0, \tag{9}$$

the *time evolution* of the state variables is described by means of the following *Navier-Stokes-Fourier system* that expresses the fundamental physical principles:

MASS CONSERVATION

$$\partial_t \varrho + \mathrm{div}_x(\varrho \mathbf{u}) = 0; \tag{10}$$

MOMENTUM BALANCE

$$\partial_t(\varrho\mathbf{u}) + \mathrm{div}_x(\varrho\mathbf{u}\otimes\mathbf{u}) + \nabla_x p(\varrho,\vartheta) = \mathrm{div}_x \mathbb{S}(\vartheta, \nabla_x\mathbf{u}) + \varrho\mathbf{f}; \tag{11}$$

ENERGY BALANCE

$$\partial_t\left(\frac{1}{2}\varrho|\mathbf{u}|^2 + \varrho e(\varrho,\vartheta)\right) + \mathrm{div}_x\left[\left(\frac{1}{2}\varrho|\mathbf{u}|^2 + \varrho e(\varrho,\vartheta)\right)\mathbf{u} + p(\varrho,\vartheta)\mathbf{u} - \mathbb{S}(\vartheta,\nabla_x\mathbf{u})\cdot\mathbf{u}\right] \tag{12}$$
$$+\mathrm{div}_x\mathbf{q}(\vartheta,\nabla_x\vartheta) = \varrho\mathbf{f}\cdot\mathbf{u};$$

where $\mathbf{f}$ is an external force, $\mathbb{S}(\vartheta,\nabla_x\mathbf{u})$ is the *viscous stress tensor* here determined by

NEWTON'S LAW

$$\mathbb{S}(\vartheta,\nabla_x\mathbf{u}) = \mu(\vartheta)\left(\nabla_x\mathbf{u} + \nabla_x^t\mathbf{u} - \frac{2}{3}\mathrm{div}_x\mathbf{u}\mathbb{I}\right) + \eta(\vartheta)\mathrm{div}_x\mathbf{u}\mathbb{I}; \tag{13}$$

and $\mathbf{q}(\vartheta,\nabla_x\vartheta)$ is the heat flux given by

FOURIER'S LAW

$$\mathbf{q}(\vartheta,\nabla_x\vartheta) = -\kappa(\vartheta)\nabla_x\vartheta. \tag{14}$$

3.2. **Physical domains, boundary conditions.** In the case $\partial\Omega \neq \emptyset$, relevant boundary conditions must be prescribed. We focus on the domains with *impermeable* boundaries, both mechanically and thermally. Accordingly, we impose the boundary conditions

$$\mathbf{u}\cdot\mathbf{n}|_{\partial\Omega} = 0 \tag{15}$$

and

$$\mathbf{q}(\vartheta,\nabla_x\vartheta)\cdot\mathbf{n}|_{\partial\Omega} = 0. \tag{16}$$

In addition to (15), we suppose that the behavior of the fluid in the tangential direction to $\partial\Omega$ obeys

NAVIER'S SLIP BOUNDARY CONDITION

$$[\mathbb{S}(\vartheta, \nabla_x \mathbf{u}) \cdot \mathbf{n}]_{\rm tan} + \beta [\mathbf{u}]_{\rm tan}|_{\partial\Omega} = 0, \tag{17}$$

where $\beta \in [0, \infty]$ plays the role of a *friction* coefficient. We focus on the two extremal situations where either $\beta = 0$ and (17) reduces to the *complete slip* boundary condition

$$[\mathbb{S}(\vartheta, \nabla_x \mathbf{u}) \cdot \mathbf{n}] \times \mathbf{n}|_{\partial\Omega} = 0, \tag{18}$$

or $\beta = \infty$, for which (15), (17) give rise to the very common *no slip* condition

$$\mathbf{u}|_{\partial\Omega} = 0. \tag{19}$$

The boundary conditions (15), (16), supplemented with either (18) or (19), are *conservative* and give rise, by integrating (12), to

TOTAL ENERGY BALANCE

$$\frac{\rm d}{{\rm d}t} \int_\Omega \left(\frac{1}{2} \varrho |\mathbf{u}|^2 + \varrho e(\varrho, \vartheta) \right) \ {\rm d}x = \int_\Omega \varrho \mathbf{f} \cdot \mathbf{u} \ {\rm d}x \tag{20}$$

at least if $\Omega \subset R^3$ is a bounded domain.

If $\Omega \subset R^3$ is unbounded, the far-field behavior of the state variables must be prescribed, for instance,

$$\varrho \to \varrho_\infty, \ \vartheta \to \vartheta_\infty, \ \mathbf{u} \to \mathbf{u}_\infty \ \text{as} \ |x| \to \infty, \tag{21}$$

and the total energy balance (20) must be modified accordingly.

3.3. Equivalent formulation of the energy balance.

The energy balance equation (12) is very often replaced by another balance law that is equivalent to (12) at least in the framework of *classical solutions* to the Navier-Stokes-Fourier system.

3.3.1. *Thermal energy.* Introducing the *specific heat at constant volume* (cf. (6))

$$c_V(\varrho, \vartheta) = \frac{\partial e(\varrho, \vartheta)}{\partial \vartheta}$$

we may rewrite (12) in the form of

THERMAL ENERGY EQUATION

$$\varrho c_v(\varrho, \vartheta) \Big(\partial_t \vartheta + \mathbf{u} \cdot \nabla_x \vartheta \Big) - {\rm div}_x \Big(\kappa(\vartheta) \nabla_x \vartheta \Big) = \mathbb{S}(\vartheta, \nabla_x \mathbf{u}) : \nabla_x \mathbf{u} - \vartheta \frac{\partial p(\varrho, \vartheta)}{\partial \vartheta} {\rm div}_x \mathbf{u}, \tag{22}$$

where, of course, we have exploited several identities resulting from the remaining equations in the Navier-Stokes-Fourier system.

The formulation of the Navier-Stokes-Fourier system by means of the equations (10), (11), and (22) is frequently used in the literature, in particular, the nowadays standard existence theory in the framework of classical solutions developed by Matsumura and Nishida [28], [29], Tani [35], Valli [36], [37], Valli and Zajackowski [38] uses this setting.

3.3.2. *Entropy equation and the Second law of thermodynamics.* Unlike (12), the thermal energy equation (22) is not in a divergence form that is more convenient for the *weak formulation*, where the differential operators are typically transferred on suitable smooth *test functions*. To this end, it seems more convenient to use

ENTROPY PRODUCTION EQUATION

$$\partial_t(\varrho s(\varrho,\vartheta)) + \mathrm{div}_x(\varrho s(\varrho,\vartheta)\mathbf{u}) + \mathrm{div}_x\left(\frac{\mathbf{q}(\vartheta,\nabla_x\vartheta)}{\vartheta}\right) = \sigma, \tag{23}$$

with the *entropy production rate*

$$\sigma = \frac{1}{\vartheta}\left(\mathbb{S}(\vartheta,\nabla_x\mathbf{u}) : \nabla_x\mathbf{u} - \frac{\mathbf{q}(\vartheta,\nabla_x\vartheta)\cdot\nabla_x\vartheta}{\vartheta}\right), \tag{24}$$

which can be obtained dividing (22) on ϑ and using the continuity equation.

In accordance with the Second law of thermodynamics, the entropy production rate σ must be non-negative. On the other hand, it is difficult to establish (24) in the framework of *weak solutions* to the Navier-Stokes-Fourier system. The problem seems to be of the same origin as its counterpart in the theory of incompressible fluid flows discussed by Duchon and Robert [9], Eyink [12], Nagasawa [30], Shvydkoy [32], or, in the context of inviscid incompressible fluids by DeLellis and Székelyhidi [8]. In other words, the weak solutions may, *hypothetically*, dissipate more kinetic energy than expressed by the quantity on the right-hand side of (24), specifically

$$\sigma \geq \frac{1}{\vartheta}\left(\mathbb{S}(\vartheta,\nabla_x\mathbf{u}) : \nabla_x\mathbf{u} - \frac{\mathbf{q}(\vartheta,\nabla_x\vartheta)\cdot\nabla_x\vartheta}{\vartheta}\right). \tag{25}$$

On the other hand, under the conservative boundary conditions specified in Section 3.2, the balance of the total energy (20) remains valid. Consequently, we may use the equations (10), (11), together with the entropy production equation (23), where σ satisfies (25), *and* the total energy balance (20) as a new formulation of the Navier-Stokes-Fourier system. It can be shown (see [15, Chapter 2]) that this new formulation is perfectly equivalent to the original system of equations, in particular the entropy production rate is given by (24), as soon as the state variables $[\varrho,\vartheta,\mathbf{u}]$ are smooth. As we will see below, the new formulation can be suitably adapted in the context of weak (distributional) solutions to obtain a mathematically tractable object.

4. **Weak and dissipative solutions.**

In accordance with the previous discussion, one of possible *weak formulations* of the Navier-Stokes-Fourier system consists of the equation of continuity (10), the momentum equation (11), together with entropy production inequality (23), (25), supplemented with the total energy balance (20), where the derivatives as well as the boundary conditions are satisfied in the sense of distributions and their traces, see [15, Chapter 3] for details. Here, we introduce even more general class of the so-called *dissipative solutions* characterized by the satisfaction of the relative entropy inequality specified below.

4.1. **Relative entropy.**

Motivated by the discussion in Section 2.2, specifically by formula (8), we introduce a relative entropy

$$\mathcal{E}\left(\varrho,\vartheta,\mathbf{u}\;\middle|\; r,\Theta,\mathbf{U}\right) \tag{26}$$

$$= \int_\Omega \left[\frac{1}{2}\varrho|\mathbf{u}-\mathbf{U}|^2 + H_\Theta(\varrho,\vartheta) - \frac{\partial H_\Theta(r,\Theta)}{\partial\varrho}(\varrho - r) - H_\Theta(r,\Theta)\right]\,\mathrm{d}x.$$

If $[\varrho, \vartheta, \mathbf{u}]$ is a smooth solution of the Navier-Stokes-Fourier system, supplemented with the no-slip condition (19) or the complete slip condition (15), (18), and if $[r, \Theta, \mathbf{U}]$ is and arbitrary trio of smooth test functions satisfying

$$r > 0, \ \Theta > 0, \ \text{and } \mathbf{U}|_{\partial\Omega} = \ \text{ or } \mathbf{U} \cdot \mathbf{n}|_{\partial\Omega} = 0, \tag{27}$$

then it is a routine matter to check that the following *relative entropy inequality* holds:

$$\begin{aligned}
&\left[\mathcal{E} \left(\varrho, \vartheta, \mathbf{u} \middle| r, \Theta, \mathbf{U} \right) \right]_{t=0}^{t=\tau} + \int_0^\tau \int_\Omega \frac{\Theta}{\vartheta} \left(\mathbb{S}(\vartheta, \nabla_x \mathbf{U}) : \nabla_x \mathbf{u} - \frac{\mathbf{q}(\vartheta, \nabla_x \vartheta) \cdot \nabla_x \vartheta}{\vartheta} \right) \ \mathrm{d}x \ \mathrm{d}t \\
&\leq \int_0^\tau \int_\Omega \Big(\varrho(\mathbf{U} - \mathbf{u}) \cdot \partial_t \mathbf{U} + \varrho(\mathbf{U} - \mathbf{u}) \otimes \mathbf{u} : \nabla_x \mathbf{U} - p(\varrho, \vartheta) \mathrm{div}_x \mathbf{U} \Big) \ \mathrm{d}x \ \mathrm{d}t \\
&\quad + \int_0^\tau \int_\Omega \Big(\mathbb{S}(\vartheta, \nabla_x \mathbf{u}) : \nabla_x \mathbf{U} + \varrho \mathbf{f} \cdot (\mathbf{u} - \mathbf{U}) \Big) \ \mathrm{d}x \ \mathrm{d}t \\
&\quad - \int_0^\tau \int_\Omega \Big(\varrho \Big(s(\varrho, \vartheta) - s(r, \Theta) \Big) \partial_t \Theta + \varrho \Big(s(\varrho, \vartheta) - s(r, \Theta) \Big) \mathbf{u} \cdot \nabla_x \Theta \Big) \ \mathrm{d}x \ \mathrm{d}t \\
&\quad + \int_0^\tau \int_\Omega \frac{\mathbf{q}(\vartheta, \nabla_x \vartheta)}{\vartheta} \cdot \nabla_x \Theta \ \mathrm{d}x \ \mathrm{d}t \\
&\quad + \int_0^\tau \int_\Omega \left(\left(1 - \frac{\varrho}{r} \right) \partial_t p(r, \Theta) - \frac{\varrho}{r} \mathbf{u} \cdot \nabla_x p(r, \Theta) \right) \ \mathrm{d}x \ \mathrm{d}t
\end{aligned} \tag{28}$$

for a.a. $\tau \in [0, T]$.

Now, the crucial observation exploited in [17] is that the relative entropy inequality (28) remains valid also for any *weak solution* $[\varrho, \vartheta, \mathbf{U}]$ as long as the test functions $[r, \Theta, \mathbf{U}]$ are sufficiently smooth and satisfy the compatibility condition (27).

4.2. **Dissipative solutions.** Following the idea of DiPerna and Lions [25] we say that $[\varrho, \vartheta, \mathbf{u}]$ is a *dissipative solution* of the Navier-Stokes-Fourier system if

$$\varrho \in L^\infty(0, T; L^p(\Omega)) \text{ for a certain } p > 1, \ \varrho \geq 0 \text{ a.a. in } (0, T) \times \Omega,$$

$$\vartheta \in L^\infty(0, T; L^p(\Omega)) \cap L^r(0, T; W^{1,r}(\Omega)) \text{ for certain } q, r > 1, \ \vartheta > 0 \text{ a.a. in } (0, T) \times \Omega,$$

$$\mathbf{u} \in L^s(0, T; W^{1,s}(\Omega; R^3)) \text{ for a certain } s > 1, \ \mathbf{u}|_{\partial\Omega} = 0 \text{ or } \mathbf{u} \cdot \mathbf{n}|_{\partial\Omega} = 0,$$

and the relative entropy inequality (28) holds for any trio of smooth test functions $[r, \Theta, \mathbf{U}]$ satisfying (27). Of course, the exponents p, q, r, and s are not arbitrary and must be adjusted so that all integrals appearing in (28) make sense. This issue will be discussed in detail in the following part of the paper.

4.3. **Existence theory.** The main advantage of the weak formulation of the Navier-Stokes-Fourier system based on the entropy production balance discussed in Section 3.3.2 is that the resulting problem is mathematically tractable, specifically, we can establish an existence theory of global-in-time solutions in the spirit of Leray's seminal paper [24].

4.3.1. *Hypotheses.* In order to present the main existence result in the framework of weak solutions, the class of thermodynamic functions p, e, and s as well as the transport coefficient μ, η and κ must be restricted.

To begin, we assume that the pressure p obeys a state equation in the form

$$p(\varrho, \vartheta) = \vartheta^{5/2} P\left(\frac{\varrho}{\vartheta^{3/2}}\right) + \frac{a}{3}\vartheta^4, \ a > 0, \tag{29}$$

with $P \in C^1[0, \infty)$. The first expression on the right-hand side is a general pressure of a monoatomic gas, while the second one accounts for the effect of radiation, see Eliezer, Ghatak, and Hora [10]. The reader may consult [15, Chapter 1] for details concerning the physical background of (29) as well as the other hypotheses introduced below.

The specific internal energy will be taken in the form

$$e(\varrho, \vartheta) = \frac{3}{2}\frac{\vartheta^{5/2}}{\varrho} P\left(\frac{\varrho}{\vartheta^{3/2}}\right) + \frac{a}{\varrho}\vartheta^4, \tag{30}$$

and

$$s(\varrho, \vartheta) = S\left(\frac{\varrho}{\vartheta^{3/2}}\right) + \frac{4a}{3}\frac{\vartheta^3}{\varrho}, \tag{31}$$

where

$$S'(Z) = -\frac{3}{2}\frac{\frac{5}{3}P(Z) - P'(Z)Z}{Z^2}. \tag{32}$$

In accordance with the hypothesis of thermodynamic stability, we further suppose that

$$P'(Z) > 0 \text{ for any } Z \geq 0, \ \frac{\frac{5}{3}P(Z) - P'(Z)Z}{Z} > 0 \text{ for any } Z > 0, \tag{33}$$

and

$$\lim_{Z \to \infty} \frac{P(Z)}{Z^{5/3}} = p_\infty > 0. \tag{34}$$

Finally, we impose technical but physically grounded hypotheses (cf. [15, Chapter 1])

$$P(0) = 0, \ \frac{\frac{5}{3}P(Z) - P'(Z)Z}{Z} < c \text{ for all } Z > 0. \tag{35}$$

The transport coefficients μ, η, and κ are continuously differentiable for $\vartheta \in [0, \infty)$ satisfying

$$\underline{\mu}(1 + \vartheta^\Lambda) \leq \mu(\vartheta) \leq \overline{\mu}(1 + \vartheta^\Lambda), \ |\mu'(\vartheta)| < c \text{ for all } \vartheta \in [0, \infty) \text{ for some } \frac{2}{5} < \Lambda \leq 1, \tag{36}$$

$$0 \leq \eta(\vartheta) \leq \overline{\eta}(1 + \vartheta^\Lambda) \text{ for all } \vartheta \in [0, \infty), \tag{37}$$

$$\underline{\kappa}(1 + \vartheta^3) \leq \kappa(\vartheta) \leq \overline{\kappa}(1 + \vartheta^3) \text{ for all } \vartheta \in [0, \infty). \tag{38}$$

4.3.2. *Global-in-time existence.* Having specified the basic hypotheses, we are ready to state the following global-in-time existence result for the Navier-Stokes-Fourier system in the framework of weak solutions, see [15, Theorem 3.1].

Theorem 4.1. *Let $\Omega \subset R^3$ be a bounded domain of class $C^{2+\nu}$, $\nu > 0$. Assume that the initial data satisfy*

$$\varrho_0 \in L^\infty(\Omega), \ \vartheta_0 \in L^\infty(\Omega), (\varrho \mathbf{u})_0 \in L^\infty(\Omega; R^3), \ \varrho_0 > 0, \ \vartheta_0 > 0 \ a.a. \ in \ \Omega,$$

and let $\mathbf{f} \in L^\infty((0, T) \times \Omega; R^3)$ be given. Let the functions p, e, s and the transport coefficients μ, η, and κ satisfy the hypotheses (29 - 38).

Then the Navier-Stokes-Fourier system admits a weak solution $[\varrho, \vartheta, \mathbf{u}]$ *in the set* $(0,T) \times \Omega$ *for any* $T > 0$.

4.3.3. *Weak-strong uniqueness and regularity criterion.* As observed in [17], any weak solution satisfies the relative entropy inequality (28). This fact can be used for deriving a version of the Gronwall inequality (4), in particular, the weak and strong solutions emanating from the same initial data coincide as long as the latter exists. This is the weak-strong uniqueness property shown in [17, Theorem 2.1]:

Theorem 4.2. *In addition to the hypotheses of Theorem 4.1 suppose that the initial data belong to the class:*

$$\varrho_0,\ \vartheta_0 \in W^{3,2}(\Omega),\ \mathbf{u}_0 \in W^{3,2}(\Omega; R^3). \tag{39}$$

Let $[\varrho, \vartheta, \mathbf{u}]$ *be the weak solution of the Navier-Stokes-Fourier system, the existence of which is guaranteed by Theorem 4.1, and let* $[\tilde{\varrho}, \tilde{\vartheta}, \tilde{\mathbf{u}}]$ *be a strong solution of the same problem belonging to the class*

$$\begin{gathered}\tilde{\varrho},\ \tilde{\vartheta} \in C([0,T]; W^{3,2}(\Omega)),\ \tilde{\mathbf{u}} \in C([0,T]; W^{3,2}(\Omega; R^3)),\\ \tilde{\vartheta} \in L^2(0,T; W^{4,2}(\Omega)),\ \partial_t \tilde{\vartheta} \in L^2(0,T; W^{2,2}(\Omega)),\\ \tilde{\mathbf{u}} \in L^2(0,T; W^{4,2}(\Omega; R^3)),\ \partial_t \tilde{\mathbf{u}} \in L^2(0,T; W^{2,2}(\Omega; R^3)),\end{gathered}$$

and emanating from the same initial data.

Then $\varrho = \tilde{\varrho}$, $\vartheta = \tilde{\vartheta}$, *and* $\mathbf{u} = \tilde{\mathbf{u}}$ *in* $[0,T]$.

Note that local-in-time strong solutions in the afore-mentioned class were constructed by Valli [36], [37], Valli and Zajackowski [38]. Since the proof uses only the relative entropy inequality, the same result is valid in the class of dissipative solutions.

Finally, we report a conditional regularity result in the spirit of Beale, Kato, and Majda [1], see [18, Theorem 2.1]:

Theorem 4.3. *In addition to the hypotheses of Theorem 4.1 suppose that the initial data belong to the regularity class (39) and satisfy the compatibility conditions:*

$$\nabla_x \vartheta_0 \cdot \mathbf{n}|_{\partial\Omega} = \mathbf{u}_0|_{\partial\Omega} = 0,\ \nabla_x p(\varrho_0, \vartheta_0)|_{\partial\Omega} = \mathrm{div}_x \mathbb{S}(\vartheta_0, \nabla_x \mathbf{u}_0) + \varrho_0 f|_{\partial\Omega}. \tag{40}$$

Let $[\varrho, \vartheta, \mathbf{u}]$ *be a weak (dissipative) solution of the Navier-Stokes-Fourier system satisfying*

$$\operatorname*{ess\ sup}_{(t,x)\in(0,T)\times\Omega} |\nabla_x \mathbf{u}(t,x)| < \infty.$$

Then $[\varrho, \vartheta, \mathbf{u}]$ *is a classical solution in the open space-time cylinder* $(0,T) \times \Omega$.

The reader will have noticed that the compatibility conditions (40) reflex the no-slip boundary condition for the velocity. The same result, with an obvious modification, applies to a general Navier slip boundary condition.

5. **Singular limits.** Singular limits are closely related to scale analysis of differential equations - an efficient tool used both theoretically and in numerical experiments to reduce the undesirable and mostly unnecessary complexity of the underlying physical system. The Navier-Stokes-Fourier system, in the entropy formulation, can be written in the dimensionless form:

$$\mathrm{Sr}\ \partial_t \varrho + \mathrm{div}_x(\varrho \mathbf{u}) = 0, \tag{41}$$

$$\mathrm{Sr}\ \partial_t(\varrho \mathbf{u}) + \mathrm{div}_x(\varrho \mathbf{u} \otimes \mathbf{u}) + \frac{1}{\mathrm{Ma}^2} \nabla_x p = \frac{1}{\mathrm{Re}} \mathrm{div}_x \mathbb{S}(\vartheta, \nabla_x \mathbf{u}) + \frac{1}{\mathrm{Fr}^2} \varrho \nabla_x F, \tag{42}$$

$$\mathrm{Sr}\ \partial_t(\varrho s) + \mathrm{div}_x(\varrho s \mathbf{u}) + \frac{1}{\mathrm{Pe}} \mathrm{div}_x \Big(\frac{\mathbf{q}(\vartheta, \nabla_x \vartheta)}{\vartheta} \Big) = \sigma, \tag{43}$$

$$\mathrm{Sr}\ \frac{\mathrm{d}}{\mathrm{d}t} \int_\Omega \Big(\frac{\mathrm{Ma}^2}{2} \varrho |\mathbf{u}|^2 + \varrho e - \frac{\mathrm{Ma}^2}{\mathrm{Fr}^2} \varrho F \Big) \mathrm{d}x = 0, \tag{44}$$

with the scaled entropy production rate

$$\sigma \geq \frac{1}{\vartheta} \Big(\frac{\mathrm{Ma}^2}{\mathrm{Re}} \mathbb{S} : \nabla_x \mathbf{u} - \frac{1}{\mathrm{Pe}} \frac{\mathbf{q} \cdot \nabla_x \vartheta}{\vartheta} \Big), \tag{45}$$

where we have taken the potential driving force $\mathbf{f} = \nabla_x F(x)$.

The dimensionless *characteristic numbers* appearing in the preceding system are defined as follows, see Klein et al. [23]:

Symbol	Definition	Name
Sr	$L_{\mathrm{ref}}/(T_{\mathrm{ref}} U_{\mathrm{ref}})$	Strouhal number
Ma	$U_{\mathrm{ref}}/\sqrt{p_{\mathrm{ref}}/\varrho_{\mathrm{ref}}}$	Mach number
Re	$\varrho_{\mathrm{ref}} U_{\mathrm{ref}} L_{\mathrm{ref}}/\mu_{\mathrm{ref}}$	Reynolds number
Fr	$U_{\mathrm{ref}}/\sqrt{L_{\mathrm{ref}} f_{\mathrm{ref}}}$	Froude number
Pe	$p_{\mathrm{ref}} L_{\mathrm{ref}} U_{\mathrm{ref}}/(\vartheta_{\mathrm{ref}} \kappa_{\mathrm{ref}})$	Péclet number

Here L_{ref} stands for the characteristic length, T_{ref} is the characteristic time, and U_{ref} is the characteristic velocity.

5.1. **Inviscid incompressible limits.** In many real world applications, in particular in meteorology, the fluid motion is rather slow, and, at the same time, the transport coefficients are small. This the situation corresponding to the choice:

$$\mathrm{Sr} = 1,\ \mathrm{Ma} = \varepsilon,\ \mathrm{Re} = \varepsilon^{-a},\ \mathrm{Pe} = \varepsilon^{-b},\ a, b > 0,$$

where $\varepsilon \to 0$ is a small parameter. Moreover, for the sake of simplicity, we set $F = 0$.

The initial data are *ill-prepared*, specifically,

$$\varrho(0, \cdot) = \varrho_{0,\varepsilon} = \overline{\varrho} + \varepsilon \varrho^{(1)}_{0,\varepsilon},\ \vartheta(0, \cdot) = \vartheta_{0,\varepsilon} = \overline{\vartheta} + \varepsilon \vartheta^{(1)}_{0,\varepsilon},\ \mathbf{u}(0, \cdot) = \mathbf{u}_{0,\varepsilon}, \tag{46}$$

where $\overline{\varrho}$, $\overline{\vartheta}$ are positive constants, and the perturbations $\varrho^{(1)}_{0,\varepsilon}$, $\vartheta^{(1)}_{0,\varepsilon}$ are allowed to be large.

For $[\varrho_\varepsilon, \vartheta_\varepsilon, \mathbf{u}_\varepsilon]$ a family of solutions to the scaled Navier-Stokes-Fourier system, we may anticipate that

$$\varrho_\varepsilon \to \overline{\varrho},\ \vartheta_\varepsilon \to \overline{\vartheta},\ \mathbf{u}_\varepsilon \to \mathbf{v},\ \frac{\vartheta_\varepsilon - \overline{\vartheta}}{\varepsilon} \to T, \tag{47}$$

where the limit velocity $\mathbf{v}$ and the temperature deviation T satisfy

$$\mathrm{div}_x \mathbf{v} = 0, \tag{48}$$

$$\partial_t \mathbf{v} + \mathbf{v} \cdot \nabla_x \mathbf{v} + \nabla_x \Pi = 0, \tag{49}$$

$$\partial_t T + \mathbf{v} \cdot \nabla_x T = 0, \tag{50}$$

cf. [16]. The system (48), (49) is nothing other than the incompressible Euler system known to possess a local in time strong solution for any regular initial data. The equation (50) represents pure transport of the temperature deviation.

5.2. **Mathematical analysis.** A rigorous justification of the limit (47), carried over in [16], is rather technical and demonstrates the strength of the method of relative entropies. Results of this type for a simpler compressible Navier-Stokes system (without temperature) were obtained by Masmoudi [26], [27].

The leading idea of the analysis is rather simple, namely, take the trio

$$\mathbf{U} = \nabla_x \Phi_\varepsilon + \mathbf{v},\ r = \overline{\varrho} + \varepsilon R_\varepsilon,\ \Theta = \overline{\vartheta} + \varepsilon T_\varepsilon$$

as test functions in the relative entropy inequality (28). The function $\mathbf{v}$ is the solution of the Euler system (48), (49), while R_ε, T_ε, and Φ_ε solve the *acoustic-transport system*:

$$\varepsilon \partial_t (\alpha R_\varepsilon + \beta T_\varepsilon) + \omega \Delta \Phi_\varepsilon = 0, \tag{51}$$

$$\varepsilon \partial_t \nabla_x \Phi_\varepsilon + \nabla_x (\alpha R_\varepsilon + \beta T_\varepsilon) = 0, \tag{52}$$

$$\partial_t (\delta T_\varepsilon - \beta R_\varepsilon) + \mathbf{U}_\varepsilon \cdot \nabla_x (\delta T_\varepsilon - \beta R_\varepsilon) + (\delta T_\varepsilon - \beta R_\varepsilon) \mathrm{div}_x \mathbf{U}_\varepsilon = 0, \tag{53}$$

with the constants

$$\alpha = \frac{1}{\overline{\varrho}} \frac{\partial p(\overline{\varrho}, \overline{\vartheta})}{\partial \varrho},\ \beta = \frac{1}{\overline{\varrho}} \frac{\partial p(\overline{\varrho}, \overline{\vartheta})}{\partial \vartheta},\ \delta = \overline{\varrho} \frac{\partial s(\overline{\varrho}, \overline{\vartheta})}{\partial \vartheta},\ \omega = \overline{\varrho} \left(\alpha + \frac{\beta^2}{\delta} \right).$$

For $Z_\varepsilon = \alpha R_\varepsilon + \beta T_\varepsilon$, the system (51), (52) can be written in the form of

ACOUSTIC EQUATION

$$\varepsilon \partial_t Z_\varepsilon + \omega \Delta \Phi_\varepsilon = 0,\ \varepsilon \partial_t \Phi_\varepsilon + Z_\varepsilon = 0. \tag{54}$$

The system (54) governs the propagation of acoustic waves supposed to "disappear" in the incompressible limit. The principal idea of the analysis is therefore to show that

$$\Phi_\varepsilon \to 0,\ Z_\varepsilon \to 0 \text{ in some sense,} \tag{55}$$

and to recover the limit equation (50) from (53). In order to show (55), we use the dispersive (Strichartz type) estimates discussed in the next section.

5.3. **Propagation of acoustic waves.** We consider a fluid flow confined to a general (unbounded) domain $\Omega \subset R^3$, where the velocity $\mathbf{u}_\varepsilon$ satisfies the complete slip boundary conditions (15), (18). Accordingly, the *acoustic potential* Φ_ε appearing in (54) satisfies the homogeneous *Neumann boundary condition*

$$\nabla_x \Phi_\varepsilon \cdot \mathbf{n}|_{\partial\Omega} = 0. \tag{56}$$

Note that the complete slip boundary conditions are also necessary in order to avoid the up to now unsurmountable difficulties connected with the presence of a boundary layer in the inviscid limit, see e.g. Kato [21].

5.3.1. *Frequency localized Strichartz estimates.* A short inspection of the solution formula associated to the acoustic problem (55), (56) reveals that solutions may be expressed by means of the wave propagator

$$h \mapsto \exp\left(\pm \mathrm{i} \frac{t}{\varepsilon} \sqrt{-\Delta_N} \right) [h],$$

where Δ_N denotes the L^2-realization of the Neumann Laplacean on Ω. Our goal will be to show

$$\int_{-\infty}^{\infty} \left\| G(-\Delta_N) \exp\left(\pm \mathrm{i} \sqrt{-\Delta_N} t \right) [h] \right\|^p_{L^q(\Omega)} \leq c(G) \|h\|^p_{H^{1,2}(\Omega)},\ \frac{1}{2} = \frac{1}{p} + \frac{3}{q},\ q < \infty, \tag{57}$$

where $G \in C_c^\infty(0,\infty)$, and where $H^{1,2}$ denotes the homogeneous Sobolev space. The estimate (57) can be viewed as frequency localized Strichartz estimates, cf. [34]. They provide the necessary piece of information in order to show the (local) decay of acoustic waves claimed in (55), cf. [16]. In the remaining part of this section, we show (57) by means of the arguments of developed by Burq [4], Smith and Sogge [33]. To this end, we suppose that $\Omega = R^3 \setminus K$ is a regular *exterior* domain, K a compact set in R^3 with a smooth boundary.

5.3.2. *Dispersive estimates for the free Laplacean.* We recall the standard *Strichartz estimates* for the free Laplacean Δ in R^3,

$$\int_{-\infty}^{\infty} \left\| \exp\left(\pm \mathrm{i}\sqrt{-\Delta} t \right) [h] \right\|_{L^q(R^3)}^p \ \mathrm{d}t \leq \|h\|_{H^{1,2}(R^3)}^p,\ \frac{1}{2} = \frac{1}{p} + \frac{3}{q},\ q < \infty, \tag{58}$$

see Keel and Tao [22], Strichartz [34].

In addition, the free Laplacean satisfies the local energy decay in the form

$$\int_{-\infty}^{\infty} \left\| \varphi \exp\left(\pm \mathrm{i}\sqrt{-\Delta} t \right) [h] \right\|_{H^{\alpha,2}(R^3)}^2 \ \mathrm{d}t \leq c(\varphi) \|h\|_{H^{\alpha,2}(R^3)}^2,\ \alpha \leq \frac{3}{2}, \tag{59}$$

see Smith and Sogge [33, Lemma 2.2].

5.3.3. *Frequency localized estimates.* To show (57), we decompose the function

$$U(t,\cdot) = G(-\Delta_N) \exp\left(\pm \mathrm{i}\sqrt{-\Delta_N} t \right) [h] = \exp\left(\pm \mathrm{i}\sqrt{-\Delta_N} t \right) G(-\Delta_N)[h]$$

as

$$U = v + w,\ v = \chi U,\ w = (1-\chi)U,$$

where

$$\chi \in C_c^\infty(R^3),\ 0 \leq \chi \leq 1,\ \chi(x) = 1 \text{ for } |x| \leq R.$$

Here R is chosen so large that the complement K of Ω is contained in the ball of the radius R.

Thus we write

$$w = w^1 + w^2,$$

where w^1 solves the homogeneous wave equation

$$\partial^2_{t,t} w^1 - \Delta w^1 = 0 \text{ in } R^3,$$

supplemented with the initial conditions

$$w^1(0) = (1-\chi)G(-\Delta_N)[h],\ \partial_t w^1(0) = \pm \mathrm{i}(1-\chi)\sqrt{-\Delta_N} G(-\Delta_N)[h],$$

while

$$\partial^2_{t,t} w^2 - \Delta w^2 = F \text{ in } R^3,$$
$$w^2(0) = \partial_t w^2(0) = 0,$$

with

$$F = -\nabla_x \chi \nabla_x U - U \Delta \chi.$$

As a direct consequence of the standard Strichartz estimates (58), we obtain

$$\int_{-\infty}^{\infty} \|w^1\|_{L^q(R^3)}^p \ \mathrm{d}t \leq c(G)\|h\|_{H^{1,2}(R^3)}^p,\ \frac{1}{2} = \frac{1}{p} + \frac{3}{q},\ q < \infty. \tag{60}$$

As the next step, we use Duhamel's formula to deduce

$$w^2(\tau,\cdot) = \frac{1}{2\sqrt{-\Delta}} \left[\exp\left(\mathrm{i}\sqrt{-\Delta}\tau \right) \int_0^\tau \exp\left(-\mathrm{i}\sqrt{-\Delta} s \right) [\eta^2 F(s)] \ \mathrm{d}s \right]$$

$$-\frac{1}{2\sqrt{-\Delta}}\left[\exp\left(-\mathrm{i}\sqrt{-\Delta}\tau\right)\int_0^\tau \exp\left(\mathrm{i}\sqrt{-\Delta}s\right)[\eta^2 F(s)]\ \mathrm{d}s\right],$$

with

$$\eta\in C_c^\infty(R^3),\ 0\le\eta\le 1,\ \eta=1 \text{ on } \mathrm{supp}[F].$$

Now, similarly to Burq [4], we use the following result of Christ and Kiselev [6]:

Lemma 5.1. *Let X and Y be Banach spaces and assume that $K(t,s)$ is a continuous function taking its values in the space of bounded linear operators from X to Y. Set*

$$\mathcal{T}[f](t)=\int_a^b K(t,s)f(s)\ \mathrm{d}s,\ \mathcal{W}[f](t)=\int_a^t K(t,s)f(s)\ \mathrm{d}s,$$

where

$$0\le a\le b\le\infty.$$

Suppose that

$$\|\mathcal{T}[f]\|_{L^p(a,b;Y)}\le c_1\|f\|_{L^r(a,b;X)}$$

for certain

$$1\le r<p\le\infty.$$

Then

$$\|\mathcal{W}[f]\|_{L^p(a,b;Y)}\le c_2\|f\|_{L^r(a,b;X)},$$

where c_2 depends only on c_1, p, and r.

We apply Lemma 5.1 to

$$X=L^2(R^3),\ Y=L^q(R^3),\ q<\infty,\ \frac{1}{2}=\frac{1}{p}+\frac{3}{q},\ r=2,$$

and

$$f=F,\ K(t,s)[F]=\frac{1}{\sqrt{-\Delta}}\exp\left(\pm\mathrm{i}\sqrt{-\Delta}(t-s)\right)[\eta^2F].$$

Writing

$$\int_0^\infty K(t,s)F(s)\ \mathrm{d}s=\exp\left(\pm\mathrm{i}\sqrt{-\Delta}t\right)\frac{1}{\sqrt{-\Delta}}\int_0^\infty \exp\left(\mp\mathrm{i}\sqrt{-\Delta}s\right)[\chi^2F(s)]\ \mathrm{d}s,$$

we have to show, in accordance with the Strichartz estimates (58), that

$$\left\|\int_0^\infty \exp\left(\pm\mathrm{i}\sqrt{-\Delta}s\right)[\eta^2F(s)]\ \mathrm{d}s\right\|_{L^2(R^3)}\le c\|F\|_{L^2(0,\infty;L^2(R^3))}. \tag{61}$$

On the other hand, however,

$$\left\|\int_0^\infty \exp\left(\pm\mathrm{i}\sqrt{-\Delta}s\right)[\chi^2F(s)]\ \mathrm{d}s\right\|_{L^2(R^3)}$$

$$=\sup_{\|v\|_{L^2(R^3)}\le 1}\int_0^\infty\left\langle \exp\left(\pm\mathrm{i}\sqrt{-\Delta}s\right)[\chi^2F(s)];v\right\rangle\ \mathrm{d}s$$

$$=\sup_{\|v\|_{L^2(R^3)}\le 1}\int_0^\infty\left\langle \chi F(s);\chi\exp\left(-\mathrm{i}\sqrt{-\Delta}s\right)[v]\right\rangle\ \mathrm{d}s;$$

whence the desired conclusion (61) follows from the local energy decay estimates (59). As the norm of F is bounded, we may infer that

$$\int_{-\infty}^\infty \|w^2\|^p_{L^q(R^3)}\ \mathrm{d}t\le c(G)\|h\|^p_{H^{1,2}(R^3)},\ \frac{1}{2}=\frac{1}{p}+\frac{3}{q},\ q<\infty. \tag{62}$$

Finally, since $v = \chi U$ is compactly supported, we deduce form the standard elliptic regularity for $-\Delta_N$ that

$$\int_0^\infty \|v\|^2_{L^q(\Omega)} \ {\rm d}t \leq c(G) \|h\|^2_{H^{1,2}(\Omega)}; \tag{63}$$

while, by virtue of the standard energy estimates,

$$\sup_{t>0} \|v(t,\cdot)\|_{L^q(\Omega)} \leq c(G)\|h\|_{H^{1,2}(\Omega)}. \tag{64}$$

where $q < \infty$ is the same as in (58). Interpolating (63), (64), we get the desired conclusion (57).

To conclude this section, we note that similar estimates on *exterior* domain can be obtained by the method of Isozaki [19]. On the other hand, the present method seems more versatile and applicable to a larger class of unbounded domains, for instance to a perturbed half-space or wave operators with non-constant coefficients arising in the stratified limits, cf. [14].

5.4. **Singular limit - main result.**

In order to formulate our main result, several remarks are in order. In agreement with the previous section, we consider the fluid confined to an unbounded domain $\Omega \subset R^3$ with a compact and regular boundary $\partial\Omega$, on which the velocity field $\mathbf{u}_\varepsilon$ satisfies the complete slip boundary conditions (15), (18). Moreover, the initial data are taken in the form (46), where

$$\varrho^{(1)}_{0,\varepsilon} \to \varrho^{(1)}_0 \text{ in } L^2(\Omega),\ \vartheta^{(1)}_{0,\varepsilon} \to \vartheta^{(1)}_0 \text{ in } L^2(\Omega),\ \|\varrho^{(1)}_{0,\varepsilon}\|_{L^\infty(\Omega)},\ \|\varrho^{(1)}_{0,\varepsilon}\|_{L^\infty(\Omega)} \leq c, \tag{65}$$

and

$$\mathbf{u}_{0,\varepsilon} \to \mathbf{u}_0 \text{ in } L^2(\Omega;R^3). \tag{66}$$

Since the spatial domain is un bounded, the far field conditions must be prescribed. In agreement with (65), (66), we take

$$\varrho_\varepsilon \to \overline{\varrho},\ \vartheta_\varepsilon \to \overline{\vartheta},\ \mathbf{u}_\varepsilon \to 0 \text{ as } |x| \to \infty. \tag{67}$$

Accordingly, the natural function spaces the solution is sought in read

$$\frac{\varrho_\varepsilon - \overline{\varrho}}{\varepsilon} \in L^\infty(0,T;L^{5/3} + L^2(\Omega)),\ \frac{\vartheta_\varepsilon - \overline{\vartheta}}{\varepsilon} \in L^\infty(0,T;L^4 + L^2(\Omega)), \tag{68}$$

and, if we fix $\Lambda = 1$ in the hypotheses (36 - 38),

$$\vartheta_\varepsilon \in L^2(0,T;W^{1,2}(\Omega)),\ \mathbf{u}_\varepsilon \in L^2(0,T;W^{1,2}(\Omega;R^3)). \tag{69}$$

Finally, we denote

$$\mathbf{v}_0 = \mathbf{H}[\mathbf{u}_0],\ \text{where } \mathbf{H} \text{ denotes the standard Helmholtz projection,}$$

and suppose that

$$\mathbf{v}_0 \in W^{k,2}(\Omega;R^3),\ k > \frac{5}{2}.$$

Our result concerning the inviscid, incompressible limit of the Navier-Stokes-Fourier system will be formulated directly in terms of the dissipative solutions, meaning the functions $[\varrho_\varepsilon, \vartheta_\varepsilon, \mathbf{u}_\varepsilon]$ satisfying the relative entropy inequality (28). Since the domain Ω is unbounded, we have to modify the space of test functions accordingly, namely

$$r > 0,\ \Theta > 0,\ \text{and } \mathbf{U}\cdot\mathbf{n}|_{\partial\Omega} = 0,\ r - \overline{\varrho}, \Theta - \overline{\vartheta},\ \mathbf{U} \text{ in } C^\infty_c([0,T] \times \overline{\Omega}).$$

Combining the dispersive estimates obtained in Section 5.3 with the method of [16] we obtain the following generalization of [16, Theorem 3.1]:

Theorem 5.2. *Let $\Omega \subset R^3$ be an unbounded domain with a compact boundary of class $C^{2+\nu}$. Suppose that the thermodynamic functions p, e, and s and the transport coefficients μ, η, κ satisfy the hypotheses (29 - 38), with $\Lambda = 1$. Let*

$$b > 0, \ \frac{10}{3} > a > 0.$$

Furthermore, suppose that the initial data (46) are chosen in such a way that

$\{\varrho_{0,\varepsilon}^{(1)}\}_{\varepsilon>0}$, $\{\vartheta_{0,\varepsilon}^{(1)}\}_{\varepsilon>0}$ *are bounded in* $L^2 \cap L^\infty(\Omega)$, $\varrho_{0,\varepsilon}^{(1)} \to \varrho_0^{(1)}$, $\vartheta_{0,\varepsilon}^{(1)} \to \vartheta_0^{(1)}$ *in* $L^2(\Omega)$, *and*

$$\{\mathbf{u}_{0,\varepsilon}\}_{\varepsilon>0} \text{ is bounded in } L^2(\Omega; R^3), \ \mathbf{u}_{0,\varepsilon} \to \mathbf{u}_0 \text{ in } L^2(\Omega; R^3),$$

where

$$\varrho_0^{(1)}, \ \vartheta_0^{(1)} \in W^{1,2} \cap W^{1,\infty}(\Omega), \ \mathbf{H}[\mathbf{u}_0] = \mathbf{v}_0 \in W^{k,2}(\Omega; R^3) \text{ for a certain } k > \frac{5}{2}.$$

Let $T_{\max} \in (0, \infty]$ denote the maximal life-span of the regular solution $\mathbf{v}$ to the Euler system (48), (49) satisfying $\mathbf{v}(0, \cdot) = \mathbf{v}_0$. Finally, let $\{\varrho_\varepsilon, \vartheta_\varepsilon, \mathbf{u}_\varepsilon\}$ be a dissipative solution of the scaled Navier-Stokes-Fourier system in $(0, T) \times \Omega$, $T < T_{\max}$, with

$$\mathrm{Sr} = 1, \ \mathrm{Ma} = \varepsilon, \ \mathrm{Re} = \varepsilon^{-a}, \ \mathrm{Pe} = \varepsilon^{-b}.$$

Then

$$\operatorname*{ess\,sup}_{t \in (0,T)} \| \varrho_\varepsilon(t, \cdot) - \overline{\varrho} \|_{L^2 + L^{5/3}(\Omega)} \leq \varepsilon c,$$

$\sqrt{\varrho_\varepsilon}\mathbf{u}_\varepsilon \to \sqrt{\overline{\varrho}}\, \mathbf{v}$ *in* $L^\infty_{\rm loc}((0, T]; L^2_{\rm loc}(\Omega; R^3))$ *and weakly-(*) in* $L^\infty(0, T; L^2(\Omega; R^3))$, *and*

$$\frac{\vartheta_\varepsilon - \overline{\vartheta}}{\varepsilon} \to T \text{ in } L^\infty_{\rm loc}((0, T]; L^q_{\rm loc}(\Omega; R^3)), \ 1 \leq q < 2,$$

$$\text{and weakly-(*) in } L^\infty(0, T; L^2(\Omega)),$$

where $\mathbf{v}$, T is the unique solution of the Euler-Boussinesq system (48 - 50), with the initial data

$$\mathbf{v}_0 = \mathbf{H}[\mathbf{u}_0], \ T_0 = \overline{\varrho}\frac{\partial s(\overline{\varrho}, \overline{\vartheta})}{\partial \vartheta}\vartheta_0^{(1)} - \frac{1}{\overline{\varrho}}\frac{\partial p(\overline{\varrho}, \overline{\vartheta})}{\partial \vartheta}\varrho_0^{(1)}.$$

Finally, we note that *existence* of the dissipative solutions for the Navier-Stokes-Fourier system in general (unbounded) domains was shown by Jesslé, Jin, and Novotný [20].

REFERENCES

[1] J. T. Beale, T. Kato, and A. Majda. Remarks on the breakdown of smooth solutions for the 3-D Euler equations. *Comm. Math. Phys.*, 94(1):61–66, 1984.

[2] S. E. Bechtel, F.J. Rooney, and M.G. Forest. Connection between stability, convexity of internal energy, and the second law for compressible Newtonian fuids. *J. Appl. Mech.*, **72**:299–300, 2005.

[3] F. Berthelin and A. Vasseur. From kinetic equations to multidimensional isentropic gas dynamics before shocks. *SIAM J. Math. Anal.*, **36**:1807–1835, 2005.

[4] N. Burq. Global Strichartz estimates for nontrapping geometries: about an article by H. F. Smith and C. D. Sogge: "Global Strichartz estimates for nontrapping perturbations of the Laplacian". *Comm. Partial Differential Equations*, **28**(9-10):1675–1683, 2003.

[5] J. Carrillo. Entropy solutions for nonlinear degenerate problems. *Arch. Rational Mech. Anal.*, **147**:269–361, 1999.

[6] M. Christ and A. Kiselev. Maximal functions associated to filtrations. *J. Funct. Anal.*, **179**(2):409–425, 2001.

[7] C.M. Dafermos. The second law of thermodynamics and stability. *Arch. Rational Mech. Anal.*, **70**:167–179, 1979.

[8] C. DeLellis and L. Székelyhidi. Dissipative Euler flows and Onsagers conjecture. 2012. Preprint.

[9] J. Duchon and R. Robert. Inertial energy dissipation for weak solutions of incompressible Euler and Navier-Stokes equations. *Nonlinearity*, **13**:249–255, 2000.

[10] S. Eliezer, A. Ghatak, and H. Hora. *An introduction to equations of states, theory and applications*. Cambridge University Press, Cambridge, 1986.

[11] J.L. Ericksen. *Introduction to the thermodynamics of solids, revised ed.* Applied Mathematical Sciences, vol. 131, Springer-Verlag, New York, 1998.

[12] G. L. Eyink. Local 4/5 law and energy dissipation anomaly in turbulence. *Nonlinearity*, **16**:137–145, 2003.

[13] C. L. Fefferman. Existence and smoothness of the Navier-Stokes equation. In *The millennium prize problems*, pages 57–67. Clay Math. Inst., Cambridge, MA, 2006.

[14] E. Feireisl, Bum Ja Jin, and A. Novotný. Inviscid incompressible limits of strongly stratified fluids. *Ann. Inst. H. Poincaré*, 2012. submitted.

[15] E. Feireisl and A. Novotný. *Singular limits in thermodynamics of viscous fluids*. Birkhäuser-Verlag, Basel, 2009.

[16] E. Feireisl and A. Novotný. Inviscid incompressible limits of the full Navier-Stokes-Fourier system. *Commun. Math. Phys.*, 2012. to appear.

[17] E. Feireisl and A. Novotný. Weak-strong uniqueness property for the full Navier-Stokes-Fourier system. *Arch. Rational Mech. Anal.*, **204**:683–706, 2012.

[18] E. Feireisl, A. Novotný, and Y. Sun. A regularity criterion for the weak solutions to the Navier-Stokes-Fourier system. *Arch. Rational Mech. Anal.*, 2012. Submitted.

[19] H. Isozaki. Singular limits for the compressible Euler equation in an exterior domain. *J. Reine Angew. Math.*, 381:1–36, 1987.

[20] D. Jesslé, B.J. Jin, and A. Novotný. Navier-Stokes-Fourier system on unbounded domains: weak solutions, relative entropies, weak-strong uniqueness. 2012. Preprint.

[21] T. Kato. Remarks on the zero viscosity limit for nonstationary Navier–Stokes flows with boundary. *In Seminar on PDE's, S.S. Chern (ed.), Springer, New York*, 1984.

[22] M. Keel and T. Tao. Endpoint Strichartz estimates. *Amer. J. Math.*, **120**(5):955–980, 1998.

[23] R. Klein, N. Botta, T. Schneider, C.D. Munz, S. Roller, A. Meister, L. Hoffmann, and T. Sonar. Asymptotic adaptive methods for multi-scale problems in fluid mechanics. *J. Engrg. Math.*, **39**:261–343, 2001.

[24] J. Leray. Sur le mouvement d'un liquide visqueux emplissant l'espace. *Acta Math.*, **63**:193–248, 1934.

[25] P.-L. Lions. *Mathematical topics in fluid dynamics, Vol.1, Incompressible models*. Oxford Science Publication, Oxford, 1996.

[26] N. Masmoudi. Incompressible inviscid limit of the compressible Navier–Stokes system. *Ann. Inst. H. Poincaré, Anal. non linéaire*, **18**:199–224, 2001.

[27] N. Masmoudi. Examples of singular limits in hydrodynamics. *In Handbook of Differential Equations, III, C. Dafermos, E. Feireisl Eds., Elsevier, Amsterdam*, 2006.

[28] A. Matsumura and T. Nishida. The initial value problem for the equations of motion of viscous and heat-conductive gases. *J. Math. Kyoto Univ.*, **20**:67–104, 1980.

[29] A. Matsumura and T. Nishida. The initial value problem for the equations of motion of compressible and heat conductive fluids. *Comm. Math. Phys.*, **89**:445–464, 1983.

[30] T. Nagasawa. A new energy inequality and partial regularity for weak solutions of Navier-Stokes equations. *J. Math. Fluid Mech.*, **3**:40–56, 2001.

[31] L. Saint-Raymond. Hydrodynamic limits: some improvements of the relative entropy method. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, **26**(3):705–744, 2009.

[32] R. Shvydkoy. Lectures on the Onsager conjecture. *Discrete Contin. Dyn. Syst. Ser. S*, **3**(3):473–496, 2010.

[33] H. F. Smith and C. D. Sogge. Global Strichartz estimates for nontrapping perturbations of the Laplacian. *Comm. Partial Differential Equations*, 25(11-12):2171–2183, 2000.

[34] R. S. Strichartz. A priori estimates for the wave equation and some applications. *J. Functional Analysis*, **5**:218–235, 1970.

[35] A. Tani. On the first initial-boundary value problem of compressible viscous fluid motion. *Publ. RIMS Kyoto Univ.*, **13**:193–253, 1977.

[36] A. Valli. A correction to the paper: “An existence theorem for compressible viscous fluids” [Ann. Mat. Pura Appl. (4) **130** (1982), 197–213; MR 83h:35112]. *Ann. Mat. Pura Appl. (4)*, **132**:399–400 (1983), 1982.

[37] A. Valli. An existence theorem for compressible viscous fluids. *Ann. Mat. Pura Appl. (4)*, **130**:197–213, 1982.

[38] A. Valli and M. Zajaczkowski. Navier-Stokes equations for compressible fluids: Global existence and qualitative properties of the solutions in the general case. *Commun. Math. Phys.*, **103**:259–296, 1986.

E-mail address: feireisl@math.cas.cz

RECENT PROGRESS IN THE THEORY OF HOMOGENIZATION WITH OSCILLATING DIRICHLET DATA

David Gérard-Varet
Institut de Mathématiques de Jussieu and University Paris 7
175 rue du Chevaleret, 75013
Paris, France

Nader Masmoudi
Courant Institute of Mathematical Sciences
251 Mercer Street
New York, NY 10012, USA

Abstract. In this paper we study the homogenization of elliptic systems with Dirichlet boundary condition, when both the coefficients and the boundary datum are oscillating, namely ε-periodic. In particular, in the paper [9], we showed that, as $\varepsilon \to 0$, the solutions converge in L^2 with a power rate in ε, and we identified the homogenized limit system and the homogenized boundary data. Due to a boundary layer phenomenon, this homogenized system depends in a non trivial way on the boundary. The analysis in [9] answers a longstanding open problem, raised for instance in [4].

1. **Introduction.** Homogenization of elliptic systems arises in several physical problems where a mixture is present. Some of the main applications of the theory are the diffusion of heat or electricity in a non-homogeneous media, the theory of elasticity of mixtures, ... Physically, the main goal of the theory is to try to compute accurate and effective properties of these mixtures. Mathematically, we have to find a limit system towards which the solutions of homogenization problem converge. This passage from "microscopic" to "macroscopic" description is called in the literature "homogenization".

When both the coefficients of the system and the boundary datum are oscillating (ε-periodic) and due to a boundary layer phenomenon, this homogenized system depends in a non trivial way on the boundary. In this talk, we answer a longstanding open problem, raised for instance by Bensoussan, Lions and Papanicolaou in their book "Asymptotic analysis for periodic structures" [4, page xiii]:

> *Of particular importance is the analysis of the behavior of solutions near boundaries and, possibly, any associated boundary layers. Relatively little seems to be known about this problem.*

In particular this result extends substantially previous works obtained for polygonal domains with sides of rational slopes as well as our previous paper [8] where the case of irrational slopes was considered. We hope that these notes give a better understanding of the proof of the result in [9].

2000 *Mathematics Subject Classification.* Primary: 35B27; Secondary: 35J25.

The first author is partially supported by the project Instabilities in Hydrodynamics. The second author is partially supported by NSF Grant DMS-1211806.

2. **The homogenization problem.** We consider the homogenization of elliptic systems in divergence form

$$-\nabla \cdot \left(A\left(\cdot/\varepsilon\right)\nabla u\right)(x) = f, \quad x \in \Omega, \tag{1}$$

set in a bounded domain Ω of $\mathbb{R}^d$, $d \geq 2$, with an oscillating Dirichlet data

$$u(x) = \varphi(x, x/\varepsilon), \quad x \in \partial\Omega. \tag{2}$$

As is customary, $\varepsilon > 0$ is a small parameter, and $A = A(y)$ takes values in $M_d\left(M_N(\mathbb{R})\right)$, namely $A^{\alpha\beta}(y) \in M_N(\mathbb{R})$ is a family of functions of $y \in \mathbb{R}^d$, indexed by $1 \leq \alpha, \beta \leq d$, with values in the set of $N \times N$ matrices. Here, $u = u(x)$ and $\varphi = \varphi(x,y)$ take their values in $\mathbb{R}^N$. We recall, using Einstein convention for summation, that for each $1 \leq i \leq N$,

$$\left(\nabla \cdot A\left(\cdot/\varepsilon\right)\nabla u\right)_i(x) \;:=\; \partial_{x_\alpha}\left[A^{\alpha\beta}_{ij}\left(\cdot/\varepsilon\right)\,\partial_{x_\beta} u_j\right](x).$$

In the sequel, Greek letters $\alpha, \beta, ...$ will range between 1 and d and Latin letters $i, j, k, ...$ will range between 1 and N.

In the context of thermics, $d = 2$ or 3, $N = 1$, u is the temperature, and $\sigma = A(\cdot/\varepsilon)\nabla u$ is the heat flux given by Fourier law. The parameter ε models heterogeneity, that is short-length variations of the material conducting properties. The boundary term φ in (2) corresponds to a prescribed temperature at the surface of the body and f is a source term. In the context of linear elasticity, $d = 2$ or 3, $N = d$, u is the unknown displacement, f is the external load and A is a fourth order tensor that models Hooke's law.

We make three hypotheses:

- **i):** Ellipticity: For some $\lambda > 0$, for all family of vectors $\xi = \xi^\alpha_i \in \mathbb{R}^{Nd}$

$$\lambda \sum_\alpha \xi^\alpha \cdot \xi^\alpha \;\leq\; \sum_{\alpha,\beta,i,j} A^{\alpha,\beta}_{ij}\,\xi^\beta_j\,\xi^\alpha_i \;\leq\; \lambda^{-1}\sum_\alpha \xi^\alpha \cdot \xi^\alpha.$$

- **ii):** Periodicity: $\forall y \in \mathbb{R}^d$, $\forall h \in \mathbb{Z}^d$, $\forall x \in \partial\Omega$, $A(y+h) = A(y)$, $\varphi(x,y) = \varphi(x, y+h)$.
- **iii):** Smoothness: The functions A, f and φ, as well as the domain Ω are smooth. It is actually enough to assume that φ and Ω are in some H^s for s big enough, but we will not try to compute the optimal regularity.

The main question we are trying to answer is the following:

Question: What is the limit behavior of the solutions u^ε as $\varepsilon \to 0$? Can we go beyond the limit and compute a full expansion of u^ε ?

This question goes back at least to the 1970's, and a classical approach consists in trying a two-scale expansion:

Classical approach: *Two-scale asymptotic expansion*:

$$\boxed{u^\varepsilon_{app} \;=\; u^0(x) \;+\; \varepsilon u^1(x, x/\varepsilon) \;+\; \ldots \;+\; \varepsilon^n u^n(x, x/\varepsilon)} \tag{3}$$

with $u^i = u^i(x,y)$ periodic in y.

3. **Case without boundary.** The two-scale approach works well in the case without boundary, namely in the whole space case or in the case of a periodic domain (say of period 1 and ε is taken to be equal to $1/n$ with n an integer). In particular one can construct inductively all the terms in the expansions. Let us recall few classical facts (see for instance [23, 20, 13, 6]) :

i): The construction of the u^i's involves the famous *cell problem*

$$\boxed{-\nabla\cdot\,(A\nabla\chi^\gamma)\,(y)\;=\;\nabla_\alpha\cdot A^{\alpha\gamma}\,(y),\quad y\text{ in }\mathbb{T}^d} \tag{4}$$

with solution $\chi^\gamma\in M_N(\mathbb{R})$.

ii): The solvability condition for u^2 yields the equation satisfied by u^0, namely u^0 (which does not depend on y) satisfies

$$\nabla\cdot A^0\nabla u^0\;=\;f \tag{5}$$

where the constant homogenized matrix is given by

$$\boxed{A^{0,\alpha\beta}\;=\;\int_{\mathbb{T}^d}A^{\alpha\beta}(y)\,dy\;+\;\int_{\mathbb{T}^d}A^{\alpha\gamma}(y)\partial_{y_\gamma}\chi^\beta(y)\,dy.}$$

The second term in the expansion (3) reads

$$u^1(x,y)\;:=\;\tilde u^1(x,y)+\bar u^1(x)\;:=\;-\chi^\alpha(y)\partial_{x_\alpha}u^0(x)+\bar u^1(x), \tag{6}$$

where χ is again the solution of (4).

To find an equation for the average part $\bar u^1(x)$, one needs to introduce another family of 1-periodic matrices

$$\Upsilon^{\alpha\beta}=\Upsilon^{\alpha\beta}(y)\in M_n(\mathbb{R}),\;\alpha,\beta=1,...,d,$$

satisfying

$$-\nabla_y\cdot A\,\nabla_y\Upsilon^{\alpha\beta}\;=\;B^{\alpha\beta}-\int_y B^{\alpha\beta},\quad\int_y\Upsilon^{\alpha\beta}=0, \tag{7}$$

where

$$B^{\alpha\beta}\;:=\;A^{\alpha\beta}-A^{\alpha\gamma}\frac{\partial\chi^\beta}{\partial y_\gamma}-\frac{\partial}{\partial y_\gamma}\left(A^{\gamma\alpha}\chi^\beta\right).$$

Formal considerations yield

$$u^2(x,y)\;:=\;\Upsilon^{\alpha,\beta}\frac{\partial^2u^0}{\partial x_\alpha\partial x_\beta}\;-\;\chi^\alpha\partial_\alpha\bar u^1+\bar u^2 \tag{8}$$

and that the average term $\bar u^1=\bar u^1(x)$ formally satisfies the equation

$$-\nabla\cdot A^0\nabla\bar u^1=c^{\alpha\beta\gamma}\frac{\partial^3u^0}{\partial x_\alpha\partial x_\beta\partial x_\gamma},\quad c^{\alpha\beta\gamma}\;:=\;\int_y A^{\gamma\eta}\frac{\partial\Upsilon^{\alpha\beta}}{\partial y_\eta}-A^{\alpha\beta}\chi^\gamma. \tag{9}$$

We refer to [2] for more details.

Inductively, one can keep constructing all the terms of the expansion by introducing new corrector families as in (7) and solving homogenized systems to determine $\bar u^k$ as in (9). Note that in this case, we do not need an extra boundary condition to solve (9).

4. **Case with boundary.** Two boundary conditions have been widely studied and are by now well understood as long as we are only interested in the first term of the expansion:

1. *The non-oscillating Dirichlet problem*, that is (1) and (2) with $\varphi = \varphi(x)$.
2. *The oscillating Neumann problem*, that is (1) and

$$n(x) \cdot \left(A(\cdot/\varepsilon)\nabla u\right)(x) \;=\; \varphi(x, x/\varepsilon), \quad x \in \partial\Omega, \tag{10}$$

   where $n(x)$ is the normal vector and with a standard compatibility condition on φ. Note that in thermics, this boundary condition corresponds to a given heat flux at the solid surface.

Notice that in both problems, the usual energy estimate provides a uniform bound on the solution u^ε in $H^1(\Omega)$.

For the non-oscillating Dirichlet problem, one shows that u^ε weakly converges in $H^1(\Omega)$ to the solution u^0 of the homogenized system

$$\begin{cases} -\nabla \cdot \left(A^0 \nabla u^0\right)(x) = f, & x \in \Omega, \\ \qquad\qquad u^0(x) = \varphi(x), & x \in \partial\Omega. \end{cases} \tag{11}$$

It is also proved in [4] that

$$u^\varepsilon(x) \;=\; u^0(x) \;+\; \varepsilon u^1(x, x/\varepsilon) \;+\; O(\sqrt{\varepsilon}), \;\text{ in }\; H^1(\Omega). \tag{12}$$

Actually, an open problem in this area was to compute the next term in the expansion in the presence of a boundary, namely to compute $u^1(x, x/\varepsilon)$. Indeed, it is not difficult to see that

$$u^1(x,y) \;=\; -\chi^\alpha(y)\partial_{x_\alpha} u^0(x) + \bar{u}^1(x), \tag{13}$$

where $\bar{u}^1(x)$ solves the homogenized equation (9). However, the main difficulty is to find the boundary data for $\bar{u}^1(x)$. The new analysis of [9] gives an answer to this problem (see also next section).

For the oscillating Neumann problem, two cases must be distinguished. On one hand, if $\partial\Omega$ does not contain flat pieces, or if it contains finitely many flat pieces whose normal vectors do not belong to $\mathbb{R}\mathbb{Z}^n$, then

$$\varphi(\cdot, \cdot/\varepsilon) \to \overline{\varphi} := \int_{[0,1]^d} \varphi \;\text{ weakly in } L^2(\partial\Omega)$$

and u^ε converges weakly to the solution u^0 of

$$\begin{cases} \qquad -\nabla \cdot \left(A^0 \nabla u^0\right)(x) = 0, & x \in \Omega, \\ n(x) \cdot \left(A^0 \nabla u^0\right)(x) = \overline{\varphi}(x), & x \in \partial\Omega. \end{cases} \tag{14}$$

On the other hand, if $\partial\Omega$ does contain a flat piece whose normal vector belongs to $\mathbb{R}\,\mathbb{Q}^d$, then the family $\varphi(\cdot, \cdot/\varepsilon)$ may have a continuum of accumulation points as $\varepsilon \to 0$. Hence, u^ε may have a continuum of accumulation points in H^1 weak, corresponding to different Neumann boundary data. We refer to [4] for all details.

5. **Case of an oscillating Dirichlet data.** Here we study (1) with the boundary data (2). One of the motivation to study this case is actually to understand the boundary condition for $\bar{u}^1(x)$ which appears in (6).

Let us explain the two main sources of difficulties in studying (1)-(2):

i): One has uniform L^p bounds on the solutions u^ε of (1)-(2), but no uniform H^1 bound *a priori*. This is due to the fact that

$$\|x \mapsto \varphi(x, x/\varepsilon)\|_{H^{1/2}(\partial\Omega)} = O(\varepsilon^{-1/2}), \text{ resp. } \|x \mapsto \varphi(x, x/\varepsilon)\|_{L^p(\partial\Omega)} = O(1),\ p > 1.$$

The usual energy inequality, resp. the estimates in article [3, page 8, Thm 3] yields

$$\|u^\varepsilon\|_{H^1(\Omega)} = O(\varepsilon^{-1/2}), \text{ resp. } \|u^\varepsilon\|_{L^p(\Omega)} = O(1),\ p > 1.$$

This indicates that singularities of u^ε are *a priori* stronger than in the usual situations. It is rigorously established in the core of the paper [9].

ii): Furthermore, one can not expect these stronger singularities to be periodic oscillations. Indeed, the oscillations of φ are at the boundary, along which they do not have any periodicity property. Hence, it is reasonable that u^ε should exhibit concentration near $\partial\Omega$, with no periodic character, as $\varepsilon \to 0$. This is a so-called *boundary layer phenomenon*. The key point is to describe this boundary layer, and its effect on the possible weak limits of u^ε.

It is important to note that there is also a boundary layer in the non-oscillating Dirichlet problem, although it has in this case a lower amplitude (it is only necessary to compute the boundary data of $\bar{u}^1$ to solve (9)). More precisely, it is responsible for the $O(\sqrt{\varepsilon})$ loss in the error estimate (12). If either the L^2 norm, or the H^1 norm in a relatively compact subset $\omega \subset \Omega$ is considered, one may avoid this loss as strong gradients near the boundary are filtered out. Following Allaire and Amar (see [2, Theorem 2.3]), we can give a more precise description than (12):

$$u^\varepsilon = u^0(x) + O(\varepsilon) \text{ in } L^2(\Omega), \quad u^\varepsilon(x) = u^0(x) + \varepsilon u^1(x, x/\varepsilon) + O(\varepsilon) \text{ in } H^1(\omega). \tag{15}$$

Still following [2], another way to put the emphasis on the boundary layer is to introduce the solution $u_{bl}^{1,\varepsilon}(x)$ of

$$\begin{cases} -\nabla \cdot A\left(\dfrac{x}{\varepsilon}\right) \nabla u_{bl}^{1,\varepsilon} = 0, & x \in \Omega \subset \mathbb{R}^d, \\ u_{bl}^{1,\varepsilon} = -u^1(x, x/\varepsilon), & x \in \partial\Omega. \end{cases} \tag{16}$$

Actually, understanding this system and requiring that $u_{bl}^{1,\varepsilon}$ goes to zero inside the domain Ω allows to determine the right boundary condition for $\bar{u}^1$. Hence, one can show that

$$u^\varepsilon(x) \ = \ u^0(x) + \varepsilon u^1(x, x/\varepsilon) + \varepsilon u_{bl}^{1,\varepsilon}(x) \ + \ O(\varepsilon), \ \text{ in } H^1(\Omega). \tag{17}$$

or

$$u^\varepsilon(x) \ = \ u^0(x) + \varepsilon u^1(x, x/\varepsilon) + \varepsilon u_{bl}^{1,\varepsilon}(x) \ + \ O(\varepsilon^2), \ \text{ in } L^2(\Omega). \tag{18}$$

Note that system (16) is a special case of (1)-(2). Thus, the homogenization of the oscillating Dirichlet problem may give a refined description of the non-oscillating one.

6. **Prior results.** Until recently, *results were all limited to convex polygons with rational normals.* This means that

$$\Omega \;:=\; \cap_{k=1}^{K} \left\{x, \quad n^k \cdot x > c^k\right\}$$

is bounded by K hyperplanes, *whose unit normal vectors n^k belong to $\mathbb{R}\,\mathbb{Q}^d$.* Under this assumption, the study of (1)-(2) can be carried out. The keypoint is the addition of boundary layer correctors to the formal two-scale expansion:

$$u^\varepsilon(x) \;\sim\; u^0(x) \;+\; \varepsilon u^1(x, x/\varepsilon) \;+\; \sum_k v_{bl}^k \left(x, \frac{x}{\varepsilon}\right), \tag{19}$$

where $v_{bl}^k = v_{bl}^k(x,y) \in \mathbb{R}^n$ is defined for $x \in \Omega$, and y in the half-space

$$\Omega^{\varepsilon,k} \;=\; \left\{y, \quad n^k \cdot y > c^k/\varepsilon\right\}.$$

These correctors satisfy

$$\begin{cases} -\nabla_y \cdot \, A(y)\nabla_y \, v_{bl}^k = 0, & y \in \Omega^{\varepsilon,k}, \\ \; v_{bl}^k = \varphi(x,y) - u^0(x), & y \in \partial\Omega^{\varepsilon,k}. \end{cases} \tag{20}$$

We refer to the papers by Moskow and Vogelius [19], and Allaire and Amar [2] for more details. These papers deal with the special case (16), but the results adapt to more general oscillating data. Note that x is just a parameter in (20) and that the assumption $n^k \in \mathbb{R}\,\mathbb{Z}^d$ yields periodicity of the function $A(y)$ tangentially to the hyperplanes. The periodicity property is used in a crucial way in the aforementioned references. First, it yields easily well-posedness of the boundary layer systems (20). Second, as was shown by Tartar in [18, Lemma 10.1] (see also subsection 7.2), the solution $v_{bl}^k(x,y)$ converges exponentially fast to some $v_{bl,*}^k(x) = \varphi_*^k(x) - u^0(x)$, when y goes to infinity transversely to the k-th hyperplane. In order for the boundary layer correctors to vanish at infinity (and to be $o(1)$ in L^2), one must have $v_{bl,*}^k = 0$, which provides the boundary condition for u^0. Hence, u^0 should satisfy a system of the type

$$\begin{cases} -\nabla \cdot \left(A^0 \nabla u^0\right)(x) = f, & x \in \Omega, \\ \qquad\qquad u^0(x) = \varphi_*(x), & x \in \partial\Omega. \end{cases} \tag{21}$$

where $\varphi_*(x) \;:=\; \varphi_*^k(x)$ on the k-th side of Ω. Nevertheless, this picture is not completely correct. Indeed, there is still *a priori* a dependence of φ_*^k on ε, through the domain $\Omega^{\varepsilon,k}$. In fact, Moskow and Vogelius exhibit examples for which there is an infinity of accumulation points for the φ_*^k's, as $\varepsilon \to 0$. Eventually, they show that the accumulation points of u^ε in L^2 are the solutions u^0 of systems like (21), in which the φ_*^k's are replaced by their accumulation points. See [19] for rigorous statements and proofs. We stress that their analysis relies heavily on the special shape of Ω, especially the rationality assumption.

A step towards more generality has been made in our recent paper [8] (see also [7]), in which generic convex polygonal domains are considered. Indeed, we assume in [8] that *the normals $n = n^k$ satisfy the Diophantine condition:*

$$\text{For all } \; \xi \in \mathbb{Z}^d \setminus \{0\} \quad |P_{n^\perp}(\xi)| > \kappa \, |\xi|^{-l}, \quad \text{for some } \; \kappa, \, l > 0, \tag{22}$$

where $P_{n^\perp}$ is the projector orthogonal to n. Note that for dimension $d = 2$ this condition amounts to:

$$\text{For all } \; \xi \in \mathbb{Z}^d \setminus \{0\} \quad |n^\perp \cdot \xi| \;:=\; |-n_2\xi_1 + n_1\xi_2| > \kappa \, |\xi|^{-l}, \quad \text{for some } \; \kappa, \, l > 0,$$

whereas for $d = 3$, it is equivalent to:

$$\text{For all } \xi \in \mathbb{Z}^d \setminus \{0\} \quad |n \times \xi| > \kappa\, |\xi|^{-l}, \quad \text{for some } \kappa,\, l > 0.$$

Condition (22) is generic in the sense that it holds for almost every $n \in S^{d-1}$.

Under this Diophantine assumption, one can perform the homogenization of problem (1)-(2). *Stricto sensu*, only the case (16), $d = 2, 3$ is treated in [8], but our analysis extends straightforwardly to the general setting. Despite a loss of periodicity in the tangential variable, we manage to solve the boundary layer equations, and prove convergence of v^k_{bl} away from the boundary. The main idea is to work with quasi-periodic functions instead of periodic ones (see also subsection 7.3). Interestingly, and contrary to the "rational case", the field φ^k_* does not depend on ε. As a result, we establish convergence of the whole sequence u^ε to the single solution u^0 of (21). We stress that, even in this polygonal setting, the boundary datum φ_* depends in a non trivial way on the boundary. In particular, it is not simply the average of φ with respect to y, contrary to what happens in the Neumann case.

7. Main new result and sketch of proof.

The main new result of [9] is to treat the case of a smooth domain:

Theorem 7.1. (Homogenization in smooth domains)

Let Ω be a smooth bounded domain of $\mathbb{R}^d$, $d \geq 2$. We assume that it is uniformly convex (all the principal curvatures are bounded from below).

Let u^ε be the solution of system (1)-(2), under the ellipticity, periodicity and smoothness conditions i)-iii).

There exists a boundary term φ_* (depending on φ, A and Ω), with $\varphi_* \in L^p(\partial\Omega)$ for all finite p, and a solution u^0 of (21), with $u^0 \in L^p(\Omega)$ for all finite p, such that:

$$\|u^\varepsilon - u^0\|_{L^2(\Omega)} \leq C_\alpha\, \varepsilon^\alpha, \qquad \textit{for all } 0 < \alpha < \frac{d-1}{3d+5}. \tag{23}$$

We will present a sketch of the proof of theorem 7.1:

From the two difficulties explain in section 5, we know that the first term in the expansion (3) should be independent of y and should solve (5). The main question is :

Question: What is the boundary value φ^0 of u^0 ?

Solution: We need a boundary layer corrector

Difficulty: There is no clear structure for the boundary layer.

Guess: The boundary layer has typical scale ε and there are no curvature effect:

- Near a point $x_0 \in \partial\Omega$, we replace $\partial\Omega$ by the tangent plane at x_0:

$$T_0(\partial\Omega) \;:=\; \{x,\; x \cdot n_0 = x_0 \cdot n_0\}$$

- We dilate by a factor ε^{-1}.

Formally, for $x \approx x_0$, one looks for

$$\boxed{u^{\varepsilon,bl}(x) \;\approx\; U_0(x/\varepsilon)}$$

where the profile $U_0 = U_0(y)$ is defined in the half plane

$$H_0^\varepsilon \;=\; \{y,\; y \cdot n_0 > \varepsilon^{-1} x_0 \cdot n_0\}.$$

It satisfies the system:

$$\boxed{\left\{\begin{array}{c} \nabla_y \cdot (A \nabla_y U_0) = 0 \quad \text{in } H_0^\varepsilon, \\ U_0|_{\partial H_0^\varepsilon} = \varphi - \varphi^0(x_0). \end{array}\right.} \tag{24}$$

Notice that in this system, x_0 is just a parameter.

7.1. **Study of an auxiliary boundary layer system.** The previous heuristic justifies the study of

$$\boxed{\left\{\begin{array}{c} \nabla_y \cdot (A \nabla_y U) = 0 \quad \text{in } H, \\ U|_{\partial H} = \phi. \end{array}\right.} \tag{BL}$$

where $H := \{y, \quad y \cdot n > a\}$ and ϕ is 1-periodic in y.

We expect that the solution U of (BL) satisfies:

$$U \to U_\infty(\phi), \qquad \text{as } y \cdot n \to +\infty,$$

for some constant $U_\infty = U_\infty(\phi)$ that depends linearly on ϕ.

If we go back to U_0 which solves (24), one can derive the homogenized boundary data φ^0. Indeed:

- On one hand, one wants $U_0 \to 0$ (localization property) when $y \cdot n \to +\infty$.
- On the other hand,

  $$U_0 \to U_\infty(\varphi - \varphi^0(x_0)) \;=\; U_\infty(\varphi) - \varphi^0(x_0)$$

  so that we need to take:

  $$\varphi^0(x_0) \;:=\; U_\infty(\varphi).$$

This formal reasoning raises many problems :

1. *The well-posedness of* (BL) *is unclear:*

   - No natural functional setting (no decay along the boundary).
   - No Poincaré inequality.
   - No maximum principle.

2. *The existence of a limit U_∞ for* (BL) *is unclear:*
   There is an underlying problem of ergodicity.

3. *U_∞ depends also on H, that is on n and a:*

   - There is no obvious regularity of U_∞ with respect to n.

   - Back to the original problem, our definition of $\varphi^0(x_0)$ depends on x_0, but also on the subsequence ε. Indeed, there is possibly many accumulation points as $\varepsilon \to 0$ (see [19]).

7.2. **Polygons with sides of rational slopes.** In this cases, the boundary layer systems of type (BL) can be fully understood (see [19, 2]). For simplicity, we only concentrate on the case $d = 2$.

1. Well-posedness: *The coefficients of the systems are periodic tangentially to the boundary.* After rotation, they turn into systems of the type

$$\boxed{\left\{\begin{array}{c} \nabla_z \cdot (B \nabla_z V) = 0, \quad z_2 > a, \\ V|_{z_2 = a} = \psi, \end{array}\right.} \tag{BL1}$$

with coefficients and boundary data that are periodic in z_1 which yields a natural variational formulation.

2. Existence of the limit : *Saint-Venant estimates* on (BL1).

   One shows that $F(t) := \int_{z_2>t} |\nabla_z V|^2 \, dz$ satisfies the differential inequality.
$$F(t) \; \leq \; -CF'(t).$$
   From there, one gets exponential decay of all derivatives, and the fact that:
$$V \to V_\infty, \;\; \text{exponentially fast, as} \;\; z_2 \to +\infty$$
   and hence going back to (BL), we get
$$U \to U_\infty, \;\; \text{exponentially fast, as} \;\; y \cdot n \to +\infty.$$
   A Key ingredient in this case is the Poincaré inequality for functions periodic in z_1 with zero mean.

3. In polygonal domains, the regularity of U_∞ with respect to n does not matter. However, for rational slopes, the limit U^∞ does depend on a. This means that if we go back to our original problem (in polygons with rational slopes), The analogue of our theorem is only available up to subsequences in ε. Moreover, the boundary data of the homogenized system may depend on the subsequence. Indeed, there are examples with a continuum of accumulation points (see [19]).

7.3. **More general treatment of** (BL)**.** It is worth pointing out that one can not be fully general: The existence of U_∞ requires some ergodicity property. A simple example is :
$$\boxed{\Delta U = 0 \quad \text{in } \{y_2 > 0\}, \quad U|_{y_2=0} = \phi}.$$

- If ϕ 1-periodic, then $U(0, y_2) \to \int_0^1 \phi$ exponentially fast.
- *But there exists $\phi \in L^\infty$ such that $U(0, y_2)$ has no limit.*

Indeed, we have an explicit formula: $U(0, y_2) \; = \; \dfrac{1}{\pi} \displaystyle\int_{\mathbb{R}} \frac{y_2}{y_2^2 + t^2} \phi(t) \, dt$. For ϕ with values in $\{+1, -1\}$, the asymptotics relates to *coin tossing*. Hence, we need some extra structure (or ergodicity) to solve the problem.

In our case, we have some ergodicity property ! For general half planes, the coefficients of (BL) or (BL1) are not periodic, *but they are quasiperiodic in the tangential variable.* We recall that a function $F = F(z_1)$ is quasiperiodic if it reads
$$\boxed{F(z_1) = \mathcal{F}(\lambda z_1)},$$
where $\lambda \in \mathbb{R}^D$ and $\mathcal{F} = \mathcal{F}(\theta)$ is periodic over $\mathbb{R}^D$ ($D \geq 1$). As an example: For (BL1), $D = 2$, and $\lambda = n^\perp$ (the tangent vector).

Notice that the previous results (subsection 7.2) correspond to the case: $n \in \mathbb{R}\,\mathbb{Q}^2$. Now, we replace this by the small divisor assumption:
$$\boxed{\text{(H)} \quad \exists \kappa > 0, \, |n \cdot \xi| \geq \kappa |\xi|^{-2}, \quad \forall \xi \in \mathbb{Z}^2 \setminus \{0\}.}$$

Note that the assumption (H) is generic in the normal n: It is satisfied by a set of full measure in $\mathbb{S}^1$. But it does not include the previous result of subsection 7.2.

Proposition 1. *If n satisfies (H), the system* (BL) *is "well-posed", with a smooth solution U that converges fast to some constant U_∞. Moreover, U_∞ does not depend on a.*

Proof of the proposition:

1. Well-posedness: involves quasiperiodicity. One has:

$$\boxed{\begin{cases} \nabla_z \cdot (B\nabla_z V) = 0, \quad z_2 > a, \\ V|_{z_2=a} = \psi, \end{cases}}$$

where $B(z) = \mathcal{B}(\lambda z_1, z_2)$, $\psi(z) = \mathcal{P}(\lambda z_1, z_2)$.

Notice that the functions $\mathcal{B} = \mathcal{B}(\theta, t)$ and $\mathcal{P} = \mathcal{P}(\theta, t)$ are periodic in $\theta \in \mathbb{T}^2$.

The idea is to consider an enlarged system in θ, t, of unknown $\mathcal{V} = \mathcal{V}(\theta, t)$:

$$\boxed{\begin{cases} D \cdot (\mathcal{B}D\mathcal{V}) = 0, \quad t > a, \\ \mathcal{V}|_{t=a} = \mathcal{P} \end{cases}} \tag{BL2}$$

where D is the "degenerate gradient" given by $D = (\lambda \cdot \nabla_\theta, \partial_t)$

Advantage: Back to a periodic setting ($\theta \in \mathbb{T}^2$).

Drawback: We have a degenerate elliptic equation. However, we are still able to prove the following :

- Variational formulation with a unique weak solution $\mathcal{V}$.
- One can prove through energy estimates than $\mathcal{V}$ is smooth.
- Allows to recover V through the formula $V(z) = \mathcal{V}(\lambda z_1, z_2)$.

2. To prove the convergence to a constant at infinity, we rely again on Saint-Venant type estimates, adapted to (BL2). Thanks to (H), we prove that $F(t) := \int_{t'>t} |D\mathcal{V}|^2 \, d\theta \, dt'$ satisfies

$$\boxed{F(t) \le C(-F'(t))^\alpha, \quad \forall \alpha < 1.}$$

But, we have only polynomial convergence towards a constant.

We point out that this better understanding of the auxiliary boundary layer systems allows to handle the generic polygonal domains in the next subsection.

7.4. **Extension to smooth domains.** The are at least three main difficulties to extend the previous analysis to smooth domain :

1. The none smoothness of U_∞ with respect to n. Indeed, U_∞ is only defined almost everywhere (diophantine assumption).

   Idea: *For any $\kappa > 0$, we can prove that U_∞ is Lipschitz when it is restricted to*

$$A_\kappa := \left\{ n \in \mathbb{S}^1, \, |n \cdot \xi| \ge \frac{\kappa}{|\xi|^2}, \, \forall \xi \in \mathbb{Z}^2 \setminus \{0\} \right\}.$$

   Moreover, we have that $|A_\kappa^c| = O(\kappa)$.

   In the course of the proof, the construction of the boundary layer corrector can be performed in the vicinity of points x such that $n(x) \in A_\kappa$. In some sense, the contribution of the remaining part of the boundary is negligible when $\kappa \ll 1$. More precisely,

2. We have to approximate the smooth domains by some polygons with sides having normal vectors in the set A_κ. In doing so, we will introduce another small parameter ε^α.

3. We have to construct a more accurate approximation due to the many errors made in the previous two points.

Broadly, optimizing in κ, α and ε yields a rate of convergence. We refer to [9] for the details.

8. **Conclusions.** We would like to conclude by mentioning a few related results. Recently there was many activity in the theory of homogenization and many new problems were addressed. We would like to mention some of them since we think they may give a better understand of our result or/and may be combined with our result:

- Our results on the boundary data problem were recently extended to the eigenvalue problem, see [21]. Also, the behavior of the reduced boundary layer system (BL) was recently investigated by C. Prange in [22], without any diophantine assumption.
- The Avellaneda-Lin type estimates were extended to the case of Neumann boundary conditions by Kenig, Lin and Shen [14, 15, 16] (see also [5] for a related work). These estimates should be helpful to study the next order approximation for the Neumann boundary condition case
- Many new probabilistic results were proved when an interface is present (see [12]) or in the trying to compute the accurate value of the homogenized matrix (see [10, 11]).
- Some different method was used to compute homogenized boundary data for none oscillating coefficient ([17, 1]).

REFERENCES

[1] Aleksanyan, H. Shahgholian, H., Sjölin, P.: Applications of Fourier analysis in homogenization and boundary layer. Available at arXiv:1205.5210v2 (2012)

[2] G. Allaire and M. Amar. Boundary layer tails in periodic homogenization. *ESAIM Control Optim. Calc. Var.*, 4:209–243 (electronic), 1999.

[3] M. Avellaneda and F.-H. Lin. Compactness methods in the theory of homogenization. *Comm. Pure Appl. Math.*, 40(6):803–847, 1987.

[4] A. Bensoussan, J.-L. Lions, and G. Papanicolaou. *Asymptotic analysis for periodic structures.* North-Holland Publishing Co., Amsterdam, 1978.

[5] X. Blanc, F. Legoll, and A. Anantharaman. Asymptotic behavior of green functions of divergence form operators with periodic coefficients. *Applied Mathematics Research eXpress*, 2012.

[6] D. Cioranescu and P. Donato. *An introduction to homogenization*, volume 17 of *Oxford Lecture Series in Mathematics and its Applications.* The Clarendon Press Oxford University Press, New York, 1999.

[7] D. Gérard-Varet and N. Masmoudi. Relevance of the slip condition for fluid flows near an irregular boundary. *Comm. Math. Phys.*, 295(1):99–137, 2010.

[8] D. Gérard-Varet and N. Masmoudi. Homogenization in polygonal domains. *J. Eur. Math. Soc. (JEMS)*, 13(5):1477–1503, 2011.

[9] D. Gérard-Varet and N. Masmoudi. Homogenization and boundary layers. *Acta Math.*, 209(1):133–178, 2012.

[10] A. Gloria and F. Otto. An optimal variance estimate in stochastic homogenization of discrete elliptic equations. *Ann. Probab.*, 39(3):779–856, 2011.

[11] A. Gloria and F. Otto. An optimal error estimate in stochastic homogenization of discrete elliptic equations. *Ann. Appl. Probab.*, 22(1):1–28, 2012.

[12] M. Hairer and C. Manson. Periodic homogenization with an interface: the multi-dimensional case. *The Annals of Probability*, 39(2):648–682, 2011.

[13] V. V. Jikov, S. M. Kozlov, and O. A. Oleinik. *Homogenization of differential operators and integral functionals.* Springer-Verlag, Berlin, 1994. Translated from the Russian by G. A. Yosifian.

[14] Kenig, C. , Lin, F., Shen, Zh.: Periodic Homogenization of Green and Neumann Functions. Available at arXiv:1201.1440v1 (2012)

[15] C. E. Kenig, F. Lin, and Z. Shen. Convergence rates in L^2 for elliptic homogenization problems. *Arch. Ration. Mech. Anal.*, 203(3):1009–1036, 2012.

[16] Kenig, C. , Lin, F., Shen, Zh.: Homogenization of Elliptic Systems with Neumann Boundary Conditions. Available at arXiv:1010.6114v1 (2010)

[17] Lee, K., Shahgholian, H.: Homogenization of the boundary value for the Dirichlet problem. Avaliable at arXiv:1201.6683v1 (2012)

[18] J.-L. Lions. *Some methods in the mathematical analysis of systems and their control.* Kexue Chubanshe (Science Press), Beijing, 1981.

[19] S. Moskow and M. Vogelius. First-order corrections to the homogenised eigenvalues of a periodic composite medium. A convergence proof. *Proc. Roy. Soc. Edinburgh Sect. A*, 127(6):1263–1299, 1997.

[20] F. Murat and L. Tartar. Calculus of variations and homogenization [MR0844873 (87i:73059)]. In *Topics in the mathematical modelling of composite materials*, volume 31 of *Progr. Nonlinear Differential Equations Appl.*, pages 139–173. Birkhäuser Boston, Boston, MA, 1997.

[21] C. Prange First-order expansion for the Dirichlet eigenvalues of an elliptic system with oscillating coefficients A paratre dans *Asymptotic Analysis*

[22] C. Prange Asymptotic analysis of boundary layer correctors in periodic homogenization A paratre dans *SIAM Journal on Mathematical Analysis*

[23] E. Sánchez-Palencia. *Nonhomogeneous media and vibration theory.* Springer-Verlag, Berlin, 1980.

E-mail address: gerard-varet@math.jussieu.fr
E-mail address: masmoudi@cims.nyu.edu

EXPLICIT AND IMPLICIT FINITE VOLUME SCHEMES FOR RADIATION MHD AND THE EFFECTS OF RADIATION ON WAVE PROPAGATION IN STRATIFIED ATMOSPHERES

Franz Fuchs

SINTEF Applied Mathematics
Oslo, Norway

Andrew McMurry and Nils Henrik Risebro

Center of Mathematics for Applications
University of Oslo
Oslo-0316, Norway

Siddhartha Mishra

Seminar for Applied Mathematics (SAM)
ETH Zurich
HG G 57.2, Rämistrasse 101, Zurich-8092, Switzerland

Abstract. We present a radiation MHD model based on moments (M1) of the radiative transport equation. This M1 model is approximated numerically by robust finite volume schemes. We compare explicit and semi-implicit schemes and show how radiation affects wave propagation in stratified atmospheres.

1. **The model.** Radiation and plasma dynamics play significant roles in energy transfer by wave propagation in stratified magneto-atmospheres. Such a configuration is modeled [1] by the equations of *stratified radiation magnetohydrodynamics (stratified RMHD)* given by,

$$
\begin{aligned}
\rho_t + \operatorname{div}(\rho\mathbf{u}) &= 0,\\
(\rho\mathbf{u})_t + \operatorname{div}\left(\rho\mathbf{u}\otimes\mathbf{u} + \left(p + \frac{1}{2}\left|\bar{\mathbf{B}}\right|^2\right)I - \bar{\mathbf{B}}\otimes\bar{\mathbf{B}}\right) &= -\rho g\mathbf{e}_3,\\
\bar{\mathbf{B}}_t + \operatorname{div}\left(\mathbf{u}\otimes\bar{\mathbf{B}} - \bar{\mathbf{B}}\otimes\mathbf{u}\right) &= 0, \qquad (1.1a)\\
E_t + \operatorname{div}\left(\left(E + p + \frac{1}{2}\left|\bar{\mathbf{B}}\right|^2\right)\mathbf{u} - \left(\mathbf{u}\cdot\bar{\mathbf{B}}\right)\bar{\mathbf{B}}\right) &= -\rho g\left(\mathbf{u}\cdot\mathbf{e}_3\right) + \mathcal{Q}^{rad},\\
\operatorname{div}(\bar{\mathbf{B}}) &= 0,
\end{aligned}
$$

where ρ is the density, $\mathbf{u} = \{u_1, u_2, u_3\}$ and $\bar{\mathbf{B}} = \{\bar{B}_1, \bar{B}_2, \bar{B}_3\}$ are the velocity and magnetic fields respectively, p is the thermal pressure, g is the constant acceleration due to gravity , $\mathbf{e}_3$ represents the unit vector in the vertical (z-) direction. E is the total energy, for simplicity determined by the ideal gas equation of state:

$$E = \frac{p}{\gamma - 1} + \frac{1}{2}\rho\left|\mathbf{u}\right|^2 + \frac{1}{2}\left|\bar{\mathbf{B}}\right|^2,$$

where $\gamma > 1$ is the adiabatic gas constant.

2000 *Mathematics Subject Classification.* Primary: 65M06; Secondary: 35C65.
Key words and phrases. MHD, radiation, explicit and implicit schemes.

The term $\mathcal{Q}^{rad}$ in (1.1a) represents energy transfer due to *radiation* and depends on the radiative intensity $I = I(\mathbf{x}, t, \Omega, \nu)$ which is a function of space $\mathbf{x} \in \mathbb{R}^3$, time $t \in \mathbb{R}$, the angle $\Omega \in S^2$ and the frequency $\nu \in \mathbb{R}$. The radiative intensity evolves in accordance with the *time-dependent* radiative transport equation,

$$\frac{1}{c}I_t + \Omega \cdot \nabla_{\mathbf{x}} I = S - \sigma^{\mathrm{ext}} I + \frac{\sigma^{\mathrm{sc}}}{4\pi} \int_{S^2} K(\Omega, \Omega') I(\Omega') d\Omega' d\nu, \tag{1.1b}$$

where $\nabla_{\mathbf{x}}$ denotes the spatial gradient, c is the speed of light, $\sigma^{\mathrm{ext}} = \sigma + \sigma^{\mathrm{sc}}$ is the extinction opacity, σ is the absorption opacity and σ^{sc} is the scattering opacity. Furthermore, $S = S(T)$ is the emission term with $T = \frac{p}{\rho g H}$ being the local temperature. For simplicity, we can assume *locally thermodynamic equilibrium* (LTE) which implies that

$$S = \sigma B(T), \tag{1.2}$$

with $B = aT^4$ being the Planck function. The scattering term in (1.1b) is given in terms of the kernel K. More details regarding the derivation of (1.1) can be checked from [4].

The term $\mathcal{Q}^{rad}$ in (1.1a) is determined by minus the integral over all frequencies ν and angles Ω of the right hand side (sources) of equation (1.1b). Since the integral over the scattering term equals $\sigma^{\mathrm{sc}} I$, this $\mathcal{Q}^{rad}$ is given by the integral of $S - \sigma I$.

The main difficulty associated with the numerical simulation of (1.1) lies in the fact that the radiative intensity is a seven dimensional function as it depends on space (3), time (1), angle (2) and frequency (1) variables. None of the currently available methods are able to resolve such a high-dimensional problem efficiently. Hence, we need to simplify the radiative transport model by reducing dimensions. Notice that the radiative energy flux $\mathcal{Q}^{rad}$ involves integrating over angle and frequency, so a detailed approximation of the radiative intensity may not be necessary in order to account for the role of radiation in MHD.

1.1. **A moments based model (M1) for radiative transfer.** As we have already pointed out, a direct simulation of the equation for radiative transfer (1.1b) is too costly. An analogous situation prevails in fluid mechanics as the Boltzmann equation modeling mesoscopic scales of the flow is high-dimensional. Suitable macroscopic scale approximations are obtained by taking moments of the Boltzmann equation that yield the Navier-Stokes equations of fluid dynamics, including closure models for completing the system of equations. Similarly, (see [5, 6] and references therein) we consider the first three angular moments of the radiative intensity I in equation (1.1b),

$$\begin{aligned} \mathcal{E} &= \frac{1}{c}\int_{S^2} I(\Omega) d\Omega, \\ \boldsymbol{\mathcal{F}} &= \frac{1}{c}\int_{S^2} \Omega I(\Omega) d\Omega, \\ \mathcal{P} &= \frac{1}{c}\int_{S^2} \Omega \otimes \Omega I(\Omega) d\Omega. \end{aligned} \tag{1.3}$$

Here, $\mathcal{E}(\nu)$, $c\boldsymbol{\mathcal{F}}(\nu)$ and $\mathcal{P}(\nu)$ are the spectral radiative energy, spectral radiative flux and spectral radiative pressure, respectively. All the quantities in the above expressions depend on the frequency variable. For simplicity, we neglect the frequency dependence and therefore work with a uni-group model. A multi-group model with explicit frequency dependence will be considered in a later paper.

Taking the zeroth and first angular moments of (1.1b) and neglecting scattering terms ($\sigma^{\mathrm{sc}} = 0$), we obtain the *M1 radiation model* ([5]),

$$\begin{aligned} \mathcal{E}_t + c\,\mathrm{div}\,\boldsymbol{\mathcal{F}} &= c\sigma(aT^4 - \mathcal{E}), \\ \boldsymbol{\mathcal{F}}_t + c\,\mathrm{div}\,\mathcal{P} &= -c\sigma\boldsymbol{\mathcal{F}}. \end{aligned} \tag{1.4}$$

In our units the speed of light is $c = 3 \cdot 10^4$. For wave propagation in stratified atmospheres, the absorption opacity $\sigma(\mathbf{x}, t)$ scales as the density ρ.

The equations have to be closed by specifying the radiative pressure $\mathcal{P}$ in terms of the lower moments. An entropy maximization method was employed in [5] to obtain the following moment closure,

$$\mathcal{P} = D\mathcal{E}, \quad D = \frac{1-\chi}{2}\mathbf{Id} + \frac{3\chi - 1}{2}\frac{\boldsymbol{\mathcal{F}} \otimes \boldsymbol{\mathcal{F}}}{\|\boldsymbol{\mathcal{F}}\|^2}, \quad \chi = \frac{3 + 4f^2}{5 + 2\sqrt{4 - 3f^2}}, \quad f = \left\|\frac{\boldsymbol{\mathcal{F}}}{\mathcal{E}}\right\|. \tag{1.5}$$

Here, $\mathbf{Id}$ is the 3×3 identity matrix, D is referred to as the Eddington tensor and χ as the Eddington factor.

Remark 1.1. Although, the M1 model is derived by integrating over all angles Ω and frequencies ν, directional information is partly recovered through the radiative flux $c\boldsymbol{\mathcal{F}}$. In addition, frequency dependence can be recovered by using a multi-group model.

Combining the M1 model (1.4), (1.5) with stratified MHD, based on the Godunov-Powell form and embedded steady states [3], we obtain the following reduced model, henceforth called *reduced stratified radiation MHD equations*,

$$\begin{aligned} &\rho_t + \mathrm{div}(\rho\mathbf{u}) = 0, \\ &(\rho\mathbf{u})_t + \mathrm{div}\left(\rho\mathbf{u} \otimes \mathbf{u} + \left(p + \frac{1}{2}|\mathbf{B}|^2 + \tilde{\mathbf{B}} \cdot \mathbf{B}\right) I - \mathbf{B} \otimes \mathbf{B} - \tilde{\mathbf{B}} \otimes \mathbf{B} - \mathbf{B} \otimes \tilde{\mathbf{B}}\right) \\ &\qquad\qquad = -\left(\mathbf{B} + \tilde{\mathbf{B}}\right)(\mathrm{div}\,\mathbf{B}) - \rho g \mathbf{e}_3, \\ &\mathbf{B}_t + \mathrm{div}\left(\mathbf{u} \otimes \mathbf{B} - \mathbf{B} \otimes \mathbf{u} + \mathbf{u} \otimes \tilde{\mathbf{B}} - \tilde{\mathbf{B}} \otimes \mathbf{u}\right) = -\mathbf{u}(\mathrm{div}\mathbf{B}), \\ &E_t + \mathrm{div}\left(\left(E + p + \frac{1}{2}|\mathbf{B}|^2 + \mathbf{B} \cdot \tilde{\mathbf{B}}\right)\mathbf{u} - (\mathbf{u} \cdot \mathbf{B})\mathbf{B} - \left(\mathbf{u} \cdot \tilde{\mathbf{B}}\right)\tilde{\mathbf{B}}\right) \\ &\qquad\qquad = -(\mathbf{u} \cdot \mathbf{B})(\mathrm{div}\,\mathbf{B}) - \rho g\left(\mathbf{u} \cdot \mathbf{e}_3\right) - c\sigma(aT^4 - \mathcal{E}), \\ &\mathcal{E}_t + c\,\mathrm{div}\,\boldsymbol{\mathcal{F}} = +c\sigma(aT^4 - \mathcal{E}), \\ &\boldsymbol{\mathcal{F}}_t + c\,\mathrm{div}\,\mathcal{P} = -c\sigma\boldsymbol{\mathcal{F}}. \end{aligned} \tag{1.6}$$

Here, the radiative pressure $\mathcal{P}$ is determined by the moment closure (1.5) and $\tilde{\mathbf{B}}$ is any *potential* background magnetic field with

$$\mathrm{div}(\tilde{\mathbf{B}}) \equiv 0, \quad \mathrm{curl}(\tilde{\mathbf{B}}) \equiv 0.$$

The stratified radiation MHD model (1.6) can be written in the following balance law form,

$$\begin{aligned} \mathbf{U}_t^{mhd} + \mathrm{div}\left(\mathbf{F}^{mhd}(\mathbf{U}^{mhd})\right) &= \mathbf{S}^{GP} + \mathbf{S}^g + \mathbf{S}^{rad}(T, \mathcal{E}), \\ \mathbf{U}_t^{rad} + \mathrm{div}\left(\mathbf{F}^{rad}(\mathbf{U}^{rad})\right) &= \tilde{\mathbf{S}}^{rad}(T, \mathcal{E}), \end{aligned} \tag{1.7}$$

where $\mathbf{U}^{mhd} = \{\rho, \mathbf{u}, \mathbf{B}, E\}$ denote the plasma variables and $\mathbf{U}^{rad} = \{\mathcal{E}, \boldsymbol{\mathcal{F}}\}$ are the radiation variables, and $T = \frac{p}{\rho g H}$. Note that (1.7) brings out the split structure of

(1.6) quite clearly as the flux $\mathbf{F}^{mhd}$ is independent of $\mathbf{U}^{rad}$, and $\mathbf{F}^{rad}$ is independent of $\mathbf{U}^{mhd}$. The only coupling between the radiative and plasma variables in (1.7) is through the source term $c\sigma(aT^4 - \mathcal{E})$. This split structure can be employed in designing suitable numerical schemes for (1.6) by combining efficient schemes for the ideal MHD equations together with schemes for the M1 model.

It is essential to consider some theoretical properties of (1.6) in order to design robust numerical schemes to approximate the solutions of the stratified RMHD equations.

1.2. **Theoretical properties of the stratified RMHD equations.** The stratified RMHD equations (1.6) have many desirable physical properties. It is important for the design of numerical schemes to inherit those. We summarize some properties below, starting with the hyperbolicity of RMHD.

- **Hyperbolicity.** Consider (1.6) in the x-direction and evaluate the flux Jacobian A of the flux $\mathbf{f} = \left(\mathbf{F}^{mhd}, \mathbf{F}^{rad}\right)$. The split structure of (1.7) is reflected in the block diagonal form of the Jacobian,

$$A = \begin{pmatrix} A^{mhd} & \mathbf{0} \\ \mathbf{0} & A^{rad} \end{pmatrix},$$

  where A^{mhd}, A^{rad} are the Jacobians corresponding to $\mathbf{F}^{mhd}$ and $\mathbf{F}^{rad}$, respectively. The eigenvalues of A^{mhd} are well-known ([2]), and the eigenvalues of A^{rad} can be explicitly calculated ([6]), and are listed below. The strict hyperbolicity of the M1 model is a consequence of its derivation by an entropy principle.
- **Positivity.** The standard positivity requirement is that

$$\rho \geq 0, \quad p \geq 0, \quad \mathcal{E} \geq 0. \tag{1.8}$$

  A solution of the stratified RMHD equations (1.6) with initial conditions satisfying (1.8) remains positive for all time, see [6].
- **Flux limitation.** For the eigenvalues to be less than the speed of light, the normalized radiative flux needs to be limited, i.e.

$$f = \left\| \frac{\boldsymbol{\mathcal{F}}}{\mathcal{E}} \right\| \leq 1. \tag{1.9}$$

  Again, a solution of the stratified RMHD equations (1.6) with initial conditions satisfying (1.9) has a limited flux for all time, see [6].
- **Energy balance.** The variation of the total energy

$$\tilde{E} = E + \mathcal{E}$$

  is only due to the Godunov-Powell source term and the gravity term in (1.6). In particular, if $\operatorname{div}\mathbf{B} \equiv 0$ and $g = 0$, then the total energy is conserved.
- **Steady states.** The steady state of interest for (1.6) is given by,

$$\begin{aligned} &\mathbf{u} \equiv 0, \quad \mathbf{B} \equiv 0, \quad p = p_0 e^{\frac{-z}{H}}, \quad \rho = \rho_0 e^{\frac{-z}{H}}, \\ &\mathcal{E} = aT_0^4, \quad \boldsymbol{\mathcal{F}} = 0, \end{aligned} \tag{1.10}$$

  where $T_0 = \frac{p_0}{gH\rho_0}$ is the constant model temperature and p_0, $\rho_0 = \frac{p_0}{gHT_0}$ are the pressure and density at the bottom $z = 0$. The embedded magnetic field $\tilde{\mathbf{B}}$ can be ANY divergence- and curl-free field.

- **Asymptotic behavior.** As pointed out in [6], the M1 system recovers the equilibrum diffusion regime for large absorption coefficients σ. That is, as $\sigma \to \infty$ (the limit for an opaque medium), the M1 model recovers the correct equlibrium diffusion equation for the temperature.

Remark 1.2. Eigensystem. The eigenvalues of the radiation part of the flux in (1.7), i.e. $\mathbf{F}^{rad}$, are scaled with the *speed of light* c. Both the *diffusion* and *transport* limit are captured by the equations. At equilibrium, when the flux is zero, i.e. $f = \|\mathbf{f}\| = 0$, the correct *diffusion limit* is recovered. That is, $\mathcal{P} = \mathcal{E}/3$ and the largest eigenvalues are $\lambda^{\pm} = \pm\frac{c}{\sqrt{3}}$.On the other hand, in the case of extreme non-equilibrium, i.e. $\|\mathbf{f}\| = 1$, the proper *transport limit* is recovered, i.e. $\mathcal{P} = \mathcal{E}$. Regarding the eigenvalues in this case, we have that the largest eigenvalues are $\lambda_{\pm} = \pm c$.

In 2 dimensions the eigenvalues of the Jacobian A^{rad} can be explicitly calculated to be (see [6])

$$\begin{aligned} \lambda^{\pm} &= c\left(\frac{f_1}{\xi} \pm \frac{\sqrt{2(\xi-1)(\xi+2)(2(\xi-1)(\xi+2)+3f_3^2)}}{\sqrt{3}\xi(\xi+2)}\right), \\ \lambda^0 &= \frac{c(2-\xi)f_1}{f^2}. \end{aligned} \tag{1.11}$$

Here, $\xi = \sqrt{4-3f^2}$ and $f = (f_1, f_3) = (\frac{\mathcal{F}^1}{\mathcal{E}}, \frac{\mathcal{F}^3}{\mathcal{E}})$. The eigenvalues of the Jacobian in the z-direction can be calculated by replacing f_1 with f_3 in the above expression. The eigenvalues of the Jacobian in 2d are depicted in figure 1. It is easy to see that the eigenvalues coincide if $\|f\| = 1$, i.e. we have that

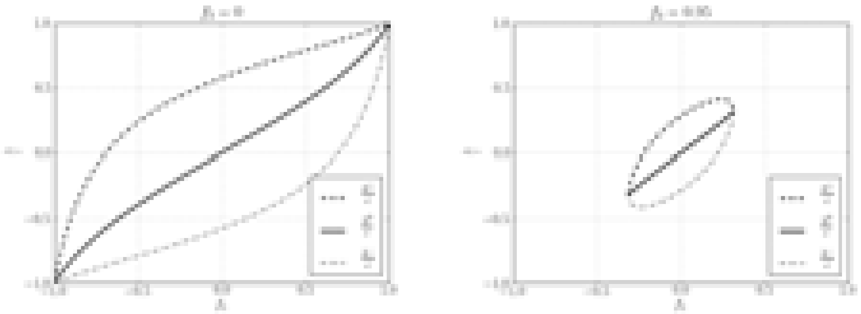


FIGURE 1. The dimensionless eigenvalues of M1 in 2d.

$$\lambda^+ = \lambda^0 = \lambda^- = cf_1, \quad \text{if } \|f\| = 1.$$

Moreover, the characteristic fields associated with $\lambda^{\pm}$ are genuinely nonlinear, whereas the field associated with λ^0 is linearly degenerate, see [6].

Remark 1.3. Diffusion limit and boundary conditions. The diffusion limit $(f = 0)$ is of importance for simulations concerning the solar atmosphere, as the bottom boundary conditions can be derived by using this limit. At the bottom of the photosphere the Sun becomes opaque to visible light, i.e. $\sigma \to \infty$. The asymptotic behavior as $\sigma \to \infty$ is captured by the diffusion limit where $f = 0$. Therefore, the bottom boundary condition for the radiative flux should be

$$\boldsymbol{\mathcal{F}}(z = 0) = \mathbf{0}, \tag{1.12}$$

and for the radiative energy we can use Neumann boundary conditions or set $\mathcal{E}(z = 0) = aE^4(z = 0)$.

Due to the steady state structure we can model waves in stratified atmospheres (at least in the chromosphere) as perturbations of an equilibrium given in equation (1.10). Waves introduced at the bottom boundary will perturb the radiative equilibrium as they move up the domain and numerical simulations will be focused on the dynamical behavior. However, the main difficulty in numerical computations is, that the fastest wave speeds for equation (1.6) are of the order of the speed of light. Hence, they are much faster than the fast magneto-sonic waves in stratified MHD. According to the CFL condition, an explicit numerical scheme will suffer from a severe restriction of the time step of the order of $\Delta t = \mathcal{O}(\Delta x/c)$. In the following, we will describe a semi-implicit scheme that circumvents this problem and compare it to the explicit scheme.

2. **Finite volume schemes.** We need to design a robust and efficient finite volume scheme that preserves at least some of the properties outlined above. For simplicity, we approximate (1.6) in a Cartesian domain $\mathbf{x} = (x, y, z) \in [X_l, X_r] \times [Y_l, Y_r] \times [Z_b, Z_t]$ and discretize it by a uniform grid in all directions with the grid spacing $\Delta x, \Delta y$ and Δz. We set $x_i = X_l + i\Delta x$, $y_j = Y_l + j\Delta y$ and $z_k = Z_b + k\Delta z$. The indices are $0 \le i \le N_x$, $0 \le j \le N_y$ and $0 \le k \le N_z$. Set $x_{i+1/2} = x_i + \Delta x/2$, $y_{j+1/2} = y_j + \Delta y/2$ and $z_{k+1/2} = z_k + \Delta z/2$, and let $\mathcal{C}_{i,j,k} = [x_{i-1/2}, x_{i+1/2}) \times [y_{j-1/2}, y_{j+1/2}) \times [z_{k-1/2}, z_{k+1/2})$ denote a typical cell. The cell average of the unknown state vector $\mathbf{U}$ over $\mathcal{C}_{i,j,k}$ at time t^n is denoted $\mathbf{U}_{i,j,k}$. Given the decoupled structure of the RMHD equations (1.7), we use the following finite volume scheme (in semi-discrete form),

$$\begin{aligned}
\frac{d}{dt}\mathbf{U}^{mhd}_{i,j,k} = &-\frac{1}{\Delta x}(\tilde{\mathbf{F}}^{1,mhd}_{i+1/2,j,k} - \tilde{\mathbf{F}}^{1,mhd}_{i-1/2,j,k}) - \frac{1}{\Delta y}(\tilde{\mathbf{F}}^{2,mhd}_{i,j+1/2,k} - \tilde{\mathbf{F}}^{2,mhd}_{i,j-1/2,k}) \\
&- \frac{1}{\Delta z}(\tilde{\mathbf{F}}^{3,mhd}_{i,j,k+1/2} - \tilde{\mathbf{F}}^{3,mhd}_{i,j,k-1/2}) + \tilde{\mathbf{S}}^{1}_{i,j,k} + \tilde{\mathbf{S}}^{2}_{i,j,k} + \tilde{\mathbf{S}}^{3}_{i,j,k} + \mathbf{S}^{g}_{i,j,k} + \mathbf{S}^{rad}_{i,j,k}, \\
\frac{d}{dt}\mathbf{U}^{rad}_{i,j,k} = &-\frac{1}{\Delta x}(\tilde{\mathbf{F}}^{1,rad}_{i+1/2,j,k} - \tilde{\mathbf{F}}^{1,rad}_{i-1/2,j,k}) - \frac{1}{\Delta y}(\tilde{\mathbf{F}}^{2,rad}_{i,j+1/2,k} - \tilde{\mathbf{F}}^{2,rad}_{i,j-1/2,k}) \\
&- \frac{1}{\Delta z}(\tilde{\mathbf{F}}^{3,rad}_{i,j,k+1/2} - \tilde{\mathbf{F}}^{3,rad}_{i,j,k-1/2}) + \tilde{\mathbf{S}}^{rad}_{i,j,k}.
\end{aligned} \tag{2.1}$$

Here

$$\mathbf{U}^{mhd}_{i,j,k} = \{\rho_{i,j,k}, \mathbf{u}_{i,j,k}, \mathbf{B}_{i,j,k}, E_{i,j,k}\}, \quad \mathbf{U}^{rad}_{i,j,k} = \{\mathcal{E}_{i,j,k}, \boldsymbol{\mathcal{F}}_{i,j,k}\},$$

are cell averages of the unknowns in the cell $\mathcal{C}_{i,j,k}$. The numerical fluxes $\tilde{\mathbf{F}}^{1,mhd}_{i+1/2,j,k}$, $\tilde{\mathbf{F}}^{2,mhd}_{i,j+1/2,k}$ and $\tilde{\mathbf{F}}^{3,mhd}_{i,j,k+1/2}$, the Godunov Powell sources $\tilde{\mathbf{S}}^{1,2,3}_{i,j,k}$, and the gravity source $\mathbf{S}^{g}_{i,j,k}$ are all independent of $\mathbf{U}^{rad}_{i,j,k}$. We can therefore directly use any numerical scheme devised for approximating the solutions of stratified MHD equations. In particular, the *well-balanced* three-wave HLLC fluxes with the *upwind* discretization of the Godunov-Powell source terms $\mathbf{S}^{1,2,3}$ of (1.7) described in [3] is a suitable choice. In this case, the gravity term $\mathbf{S}^g$ is discretized in a well balanced manner (see [3]) that differs from the standard pointwise evaluation of the source. We are left with having to design suitable discretizations of $\tilde{\mathbf{F}}^{rad}$ and the source terms $S^{rad}, \tilde{S}^{rad}$. We start by describing an explicit discretization. Due to the large restriction on the time step for an explicit scheme in this case, we propose a semi-implicit scheme that allows for much larger time steps.

2.1. **Explicit discretization of the radiative terms.** The key step in completing (2.1) is to define the radiation flux $\tilde{\mathbf{F}}^{rad}$. For simplicity, we choose an HLL two-wave flux. We start with the description of the numerical flux $\tilde{\mathbf{F}}^{1,rad}_{i+1/2.j,k}$ which is defined in terms of an approximate solution to the following Riemann problem,

$$\begin{aligned} \mathcal{E}_t + c\boldsymbol{\mathcal{F}}^1_x &= 0, \\ \boldsymbol{\mathcal{F}}^1_t + c(\mathcal{E}D^1)_x &= 0, \\ \boldsymbol{\mathcal{F}}^2_t + c(\mathcal{E}D^2)_x &= 0, \\ \boldsymbol{\mathcal{F}}^3_t + c(\mathcal{E}D^3)_x &= 0, \\ \mathbf{U}^{rad}(x,t^n) &= \begin{cases} \mathbf{U}^{rad}_L = \mathbf{U}^{rad}_{i,j,k}, & \text{if} \quad x \le x_{i+1/2}, \\ \mathbf{U}^{rad}_R = \mathbf{U}^{rad}_{i+1,j,k}, & \text{if} \quad x > x_{i+1/2}, \end{cases} \end{aligned} \tag{2.2}$$

Here,

$$D^1 = \frac{1-\chi}{2} + \frac{3\chi-1}{2}\frac{(\boldsymbol{\mathcal{F}}^1)^2}{\|\boldsymbol{\mathcal{F}}\|^2}, \quad D^2 = \frac{3\chi-1}{2}\frac{\boldsymbol{\mathcal{F}}^1\boldsymbol{\mathcal{F}}^2}{\|\boldsymbol{\mathcal{F}}\|^2}, \quad D^3 = \frac{3\chi-1}{2}\frac{\boldsymbol{\mathcal{F}}^1\boldsymbol{\mathcal{F}}^3}{\|\boldsymbol{\mathcal{F}}\|^2}, \tag{2.3}$$

and χ is the Eddington factor defined in (1.5).

We approximate the solution of (2.2) with the following wave structure,

$$\mathbf{U}^{rad}_{H2} = \begin{cases} \mathbf{U}^{rad}_L & \text{if } \frac{x}{t} \le s_L, \\ \mathbf{U}^{rad}_* & \text{if } s_L < \frac{x}{t} < s_R, \\ \mathbf{U}^{rad}_R & \text{if } s_R \le \frac{x}{t}, \end{cases} \qquad \mathbf{F}^{1,rad}_{H2} = \begin{cases} \mathbf{F}^{1,rad}_L & \text{if } \frac{x}{t} \le s_L, \\ \mathbf{F}^{1,rad}_* & \text{if } s_L < \frac{x}{t} < s_R, \\ \mathbf{F}^{1,rad}_R & \text{if } s_R \le \frac{x}{t}. \end{cases} \tag{2.4}$$

Note that we have used the standard HLL two-wave solver in (2.4). A straightforward calculation using the Rankine-Hugoniot condition leads to the following middle flux,

$$\mathbf{F}^{1,rad}_* = \frac{s_R\mathbf{F}^{1,rad}_L - s_L\mathbf{F}^{1,rad}_R + s_L s_R(\mathbf{U}^{rad}_R - \mathbf{U}^{rad}_L)}{s_R - s_L}. \tag{2.5}$$

The resulting numerical flux is

$$\tilde{\mathbf{F}}^{1,rad}_{i+1/2,j,k} = \begin{cases} \mathbf{F}^{1,rad}_{i,j,k} & \text{, if } (s_L)_{i+1/2,j,k} > 0, \\ \mathbf{F}^{1,rad,*}_{i,j,k} & \text{, if } (s_L)_{i+1/2,j,k} \le 0 \wedge (s_R)_{i+1/2,j,k} \ge 0, \\ \mathbf{F}^{1,rad}_{i+1,j,k} & \text{, if } (s_R)_{i+1/2,j,k} < 0. \end{cases} \tag{2.6}$$

Choice of wave speeds. The wave speeds $s_{L,R}$ in (2.4) need to be chosen suitably. As they approximate the fastest waves in the M1-system (1.4), the simplest stable choice of wave speeds is,

$$s_L = -c, \quad s_R = +c, \tag{2.7}$$

where c is the constant speed of light. Note that this choice leads to a Rusanov type scheme (with a global $s_L = -s_R$ rather than a local one) for the M1 model. That means we always use the middle state and the numerical flux is given by

$$\tilde{\mathbf{F}}^{1,rad}_{i+1/2,j,k} = \frac{1}{2}(\mathbf{F}^{1,rad}_{i,j,k} + \mathbf{F}^{1,rad}_{i+1,j,k}) - \frac{c}{2}(\mathbf{U}^{rad}_{i+1,j,k} - \mathbf{U}^{rad}_{i,j,k}). \tag{2.8}$$

This choice will be dissipative (particularly at first order) but some accuracy is recovered with a second-order approximation.

A more accurate choice follows [6] and leads to,

$$\begin{aligned} s_L &= \min\{0, cf_L^1, c\frac{f_L^1 - \chi_L}{1 - f_l^1}, c\frac{f_L^1 - \chi_L}{1 + f_l^1}, \lambda^-(\mathbf{U}_L^{rad})\}, \\ s_R &= \max\{0, cf_R^1, c\frac{f_R^1 - \chi_R}{1 - f_l^1}, c\frac{f_R^1 - \chi_R}{1 + f_l^1}, \lambda^+(\mathbf{U}_R^{rad})\}, \end{aligned} \tag{2.9}$$

where $f^1 = \frac{\boldsymbol{\mathcal{F}}^1}{\mathcal{E}}$, and $\chi^{L,R} = \chi(f_{L,R}^1)$ is the Eddington factor (1.5). Similarly, $\lambda^{\pm}$ are the eigenvalues for a given state defined in (1.11). The results of [6] show that the above choice is robust.

The fluxes in the y- and z-directions are obtained by replacing the appropriate quantities in (2.5) and (2.6). This completes the description of the scheme (2.1). Second-order accuracy can be obtained by the reconstruction routines of [3].

Radiative source terms. We are left with describing the discretization of the radiative source terms. The simplest way would be a point-wise evolution of the source at the current time step n (explicit). But, depending on the constants the source term is potentially stiff. Hence, we use an implicit discretization of the radiative source term.

$$\begin{aligned} \mathbf{S}_{i,j,k}^{rad,n+1} &= \{0, \mathbf{0}, \mathbf{0}, -c\sigma_{i,j,k}^{n+1}(a(T_{i,j,k}^{n+1})^4 - \mathcal{E}_{i,j,k}^{n+1})\}, \\ \tilde{\mathbf{S}}_{i,j,k}^{rad,n+1} &= \{c\sigma_{i,j,k}^{n+1}(a(T_{i,j,k}^{n+1})^4 - \mathcal{E}_{i,j,k}^{n+1}), c\sigma_{i,j,k}^{n+1}\boldsymbol{\mathcal{F}}_{i,j,k}^{n+1}\}. \end{aligned} \tag{2.10}$$

Observe, that we still call our scheme explicit, although we discretize the radiation sources implicitly. This is because it is only locally implicit and the resulting nonlinear systems can be solved for each cell independently.

The main difficulty associated with (2.1) is the presence of time scales dominated by the speed of light c, dictating a very small time step for explicit finite volume schemes. However, we are only interested in the effect of radiation on the plasma (stratified MHD equation) and do not need to approximate the solutions of the M1 model accurately. Therefore, we devise semi-implicit schemes in the next section that remove this limitation on the time step.

2.2. **Semi-implicit solvers for Radiation-MHD.** In typical applications, the fast magneto-acoustic wave speeds are at least two to three orders of magnitude smaller than the speed of light c. Hence, due to the CFL condition any explicit scheme desigend to resolve time scales of the MHD model will be extremely slow. In order to be able to do realistic simulations, we need to devise semi-implicit schemes that allow us to increase the time step by some orders of magnitude. In this section we will describe semi-implicit solvers for approximating the solutions of RMHD (1.6). Since the only coupling of the MHD part to the M1 model for radiative transport is given by the source term, we can devide our numerical scheme into two parts.

We discretize the MHD part, i.e. the fluxes $\mathbf{F}^{mhd}$, the Powell source $\mathbf{S}^{GP}$ and the gravity source $\mathbf{S}^g$ explicitly. The M1 model along with the radiative source $\mathbf{S}^{rad}$ in the energy equation on the other hand are discretized in an implicit manner. We present the resulting nonlinear system that has to be solved in order to get an approximation to the solution at the next time level. For the sake of clarity we restrict ourself to one space dimension in the description. It is straightforward to extend this approach to several space dimensions. In one dimension we set

$(\mathcal{F}^2, \mathcal{F}^3) = (0,0)$ and the system of equations to be discretized implicitly simplifies to

$$\begin{aligned} E_t + H &= -c\sigma(aT^4 - \mathcal{E}),\\ \mathcal{E}_t + c\mathcal{F}^1_x &= +c\sigma(aT^4 - \mathcal{E}),\\ \mathcal{F}^1_t + c(\mathcal{E}\chi)_x &= -c\sigma\mathcal{F}^1. \end{aligned} \tag{2.11}$$

Here, H contains the energy-flux, the gravity source and the Powell source of the energy equation. The semi-implicit numerical solver has the following form

$$\begin{aligned} E_i^{n+1} - E_i^n + \Delta t H^n &= -\Delta t c\sigma_i^{n+1}(a(T_i^{n+1})^4 - \mathcal{E}_i^{n+1}),\\ \mathcal{E}_i^{n+1} - \mathcal{E}_i^n + \frac{\Delta t}{\Delta x}(A^{n+1}_{i+1/2} - A^{n+1}_{i-1/2}) &= +\Delta t c\sigma_i^{n+1}(a(T_i^{n+1})^4 - \mathcal{E}_i^{n+1}),\\ \mathcal{F}_i^{1,n+1} - \mathcal{F}_i^{1,n} + \frac{\Delta t}{\Delta x}(B^{n+1}_{i+1/2} - B^{n+1}_{i-1/2}) &= -\Delta t c\sigma_i^{n+1}\mathcal{F}_i^{1,n+1}. \end{aligned} \tag{2.12}$$

Note that the discretization of H in the energy equation is explicit and is readily provided by the solver for the MHD part of equation (1.6). Using the first order HLL flux (2.8) with $s_R = -s_L = c$, the flux differences simplify to

$$\begin{aligned} A^{n+1}_{i+1/2} - A^{n+1}_{i-1/2} &= \frac{c}{2}(\mathcal{F}^{1,n+1}_{i+1} - \mathcal{F}^{1,n+1}_{i-1} - \mathcal{E}^{n+1}_{i+1} + 2\mathcal{E}^{n+1}_i - \mathcal{E}^{n+1}_{i-1}),\\ B^{n+1}_{i+1/2} - B^{n+1}_{i-1/2} &= \frac{c}{2}(\mathcal{E}^{n+1}_{i+1}\chi^{n+1}_{i+1} - \mathcal{E}^{n+1}_{i-1}\chi^{n+1}_{i-1} - \mathcal{F}^{1,n+1}_{i+1} + 2\mathcal{F}^{1,n+1}_i - \mathcal{F}^{1,n+1}_{i-1}). \end{aligned} \tag{2.13}$$

The update at the new time level $\xi_i = (\vec{\alpha}, \vec{\beta}, \vec{\delta})_i = (E_i^{n+1}, \mathcal{E}_i^{n+1}, \mathcal{F}_i^{n+1})$ is given by finding a root of the following function

$$F_i(\vec{\alpha}, \vec{\beta}, \vec{\delta}) = \begin{pmatrix} \alpha_i - E_i^n + \Delta t H^n + \Delta t c\sigma_i^{n+1}(w_i^{n+1}(\alpha_i - v_i^{n+1})^4 - \beta_i)\\ \beta_i - \mathcal{E}_i^n + \frac{\Delta t}{\Delta x}A(\xi_{i-1}, \xi_i, \xi_{i+1}) - \Delta t c\sigma_i^{n+1}(w_i^{n+1}(\alpha_i - v_i^{n+1})^4 - \beta_i)\\ \delta_i - \mathcal{F}_i^n + \frac{\Delta t}{\Delta x}B(\xi_{i-1}, \xi_i, \xi_{i+1}) + \Delta t c\sigma_i^{n+1}\delta_i \end{pmatrix}, \tag{2.14}$$

where we have defined $w_i^{n+1} = a\left(\frac{\gamma-1}{\rho_i^{n+1}}\right)^4$ and $v_i^{n+1} = \frac{1}{2}\rho_i^{n+1}(\mathbf{u}_i^{n+1})^2 + \frac{1}{2}(\bar{\mathbf{B}}_i^{n+1})^2$. The values of ρ_i^{n+1}, $\mathbf{u}_i^{n+1}$, $\bar{\mathbf{B}}_i^{n+1}$ and H^n are directly available from the explicit solver for the MHD equations. The expressions for the fluxes A and B depend on the flux discretization. For the flux (2.8) we have that

$$\begin{aligned} A(\xi_{i-1}, \xi_i, \xi_{i+1}) &= \frac{c}{2}(\delta_{i+1} - \delta_{i-1} - \beta_{i+1} + 2\beta_i - \beta_{i-1}),\\ B(\xi_{i-1}, \xi_i, \xi_{i+1}) &= \frac{c}{2}\left(\beta_{i+1}\chi\left(\frac{|\delta_{i+1}|}{\beta_{i+1}}\right) - \beta_{i-1}\chi\left(\frac{|\delta_{i-1}|}{\beta_{i-1}}\right) - \delta_{i+1} + 2\delta_i - \delta_{i-1}\right). \end{aligned} \tag{2.15}$$

A solution ξ' with $F(\xi') = 0$ provides the values at the new time level $n+1$, namely $(E_i^{n+1}, \mathcal{E}_i^{n+1}, \mathcal{F}_i^{n+1}) = \xi'_i$. In order to find a root of $F(\xi)$ we use Newton iteration

$$DF(\xi_j)(\xi_{j+1} - \xi_j) = -F(\xi_j), \tag{2.16}$$

with the start value $(\xi_0)_i = (E_i^n, \mathcal{E}_i^n, \mathcal{F}_i^n)$. We stop the iteration, if $\|F(\xi_j)\| \le \text{tol}$ or after a certain number of iterations. The Jacobian has the following structure

$$DF = \begin{pmatrix} DF_{1,1} & DF_{1,2} & 0\\ DF_{2,1} & DF_{2,2} & DF_{2,3}\\ 0 & DF_{3,2} & DF_{3,3} \end{pmatrix}, \tag{2.17}$$

where $DF_{1,1}$, $DF_{1,2}$ and $DF_{2,1}$ are diagonal matrices with the diagonal entries given by $df_{1,1} = 1 + 4\Delta t c\sigma_i^{n+1} w_i^{n+1}(x_i^1 - v_i^{n+1})^3$, $df_{1,2} = -\Delta t c\sigma_i^{n+1}$ and $df_{2,1} = -\Delta t c\sigma_i^{n+1} w_i^{n+1}(x_i^1 - v_i^{n+1})^3$, respectively.

For $p, q \in \{2, 3\}$ we have the following tridiagonal structure

$$DF_{p,q} = \begin{pmatrix} \ddots & \ddots & 0 & 0 & 0 \\ \ddots & \ddots & \ddots & 0 & 0 \\ 0 & dl_{p,q} & d_{p,q} & dr_{p,q} & 0 \\ 0 & 0 & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & \ddots & \ddots \end{pmatrix}. \tag{2.18}$$

The components of $DF_{2,2}$ and $DF_{2,3}$ can be calculated as

$$\begin{aligned} dl_{2,2} = dr_{2,2} &= -\frac{c\Delta t}{2\Delta x}, d_{2,2} = 1 + \Delta t c\sigma_i^{n+1} + \frac{c\Delta t}{\Delta x}, \\ dl_{2,3} &= -\frac{c\Delta t}{2\Delta x}, dr_{2,3} = +\frac{c\Delta t}{2\Delta x}, d_{2,3} = 0. \end{aligned} \tag{2.19}$$

The derivatives of the third component of F are more complicated, due to the non-linearity of the flux function. However, for the Newton iteration to work we only need an approximation of the Jacobian of F. The function χ is a monotone function with $\frac{1}{3} \le \chi \le 1$. It is therefore reasonable to assume that χ is independent of $\frac{\mathcal{F}}{\mathcal{E}}$ in order to approximate the Jacobian. That means we have

$$\frac{\partial \beta_i \chi\left(\frac{|\delta_i|}{\beta_i}\right)}{\partial \beta_i} \approx \chi\left(\frac{|\delta_i|}{\beta_i}\right), \quad \frac{\partial \beta_i \chi\left(\frac{|\delta_i|}{\beta_i}\right)}{\partial \delta_i} \approx 0. \tag{2.20}$$

Using this assumption, it follows that the compontents of $DF_{3,2}$ and $DF_{3,3}$ are given by

$$\begin{aligned} dl_{3,2} &= -\frac{c\Delta t}{2\Delta x}\chi_{i-1}, dr_{3,2} = +\frac{c\Delta t}{2\Delta x}\chi_{i+1}, d_{3,2} = 0, \\ dl_{3,3} = dr_{3,3} &= -\frac{c\Delta t}{2\Delta x}, d_{3,3} = 1 + \Delta t c\sigma_i^{n+1} + \frac{c\Delta t}{\Delta x}. \end{aligned} \tag{2.21}$$

This concludes the description of the implicit solver for the interior points. For the Newton (2.16) iteration to be complete, we need to implement boundary conditions for both the right hand side F and the Jacobian DF. Using Neumann boundary conditions for the M_1 model in the explicit solver, translates to the following boundary conditions for the implicit solver. For $F(x)$ in equation (2.14) we need to define values for the spatial derivatives at the boundaries, namely

$$\begin{aligned} A(\xi_0, \xi_1, \xi_2) &= \frac{c}{2}(\delta_2 - \delta_1 - \beta_2 + \beta_1), \\ A(\xi_{N_x-1}, \xi_{N_x}, \xi_{N_x+1}) &= \frac{c}{2}(\delta_{N_x} - \delta_{N_x-1} + \beta_{N_x} - \beta_{N_x-1}), \\ B(\xi_0, \xi_1, \xi_2) &= \frac{c}{2}\left(\beta_2 \chi\left(\frac{|\delta_2|}{\beta_2}\right) - \beta_1 \chi\left(\frac{|\delta_1|}{\beta_1}\right) - \delta_2 + \delta_1\right), \\ B(\xi_{N_x-1}, \xi_{N_x}, \xi_{N_x+1}) &= \frac{c}{2}\left(\beta_{N_x} \chi\left(\frac{|\delta_{N_x}|}{\beta_{N_x}}\right) - \beta_{N_x-1} \chi\left(\frac{|\delta_{N_x-1}|}{\beta_{N_x-1}}\right) + \delta_{N_x} - \delta_{N_x-1}\right). \end{aligned} \tag{2.22}$$

Furthermore, for the Jacobian DF in (2.17) we define

$$\chi\left(\frac{|\delta_{N_x+1}|}{\beta_{N_x+1}}\right) = \chi\left(\frac{|\delta_{N_x}|}{\beta_{N_x}}\right), \quad \chi\left(\frac{|\delta_0|}{\beta_0}\right) = \chi\left(\frac{|\delta_1|}{\beta_1}\right) \tag{2.23}$$

at the boundaries. For different types of boundary conditions those definitions will change.

This semi-implicit approach is easily extended to multi space dimensions. We omit the description in this article, and continue with describing the properties of the various schemes.

2.3. **Properties of the explicit and semi-implicit schemes.** We consider finite volume schemes of type (2.1) approximating the solutions of the stratified RMHD equations (1.6). In the numerical experiments of this paper we choose to test and compare the following two numerical schemes. Since it is essential to preserve discrete versions of the steady states, we use the *well-balanced* three-wave HLLC solver of [3] (with the corresponding discretiziations of the Powell source and the gravitational source term as well as the balanced Neumann type boundary conditions). For the radiative part of (1.6) we choose the HLL solver described in sections 2.1 and 2.2.

Therefore, we have the following two possiblities.

- Explicit solver $M_{\text{HLLC}}R_{\text{HLL}_e}$: explicit well-balanced HLLC solver for MHD combined with the explicit HLL scheme for the radiation part, and
- semi-implicit solver $M_{\text{HLLC}}R_{\text{HLL}_i}$: explicit well-balanced HLLC solver for MHD combined with the implicit HLL scheme for the radiation part.

We want to remark that the fully explicit solver $M_{\text{HLLC}}R_{\text{HLL}_e}$ uses a locally implicit time discretization for the sources radiative sources. In one dimension the general numerical scheme has the form

$$\begin{aligned}
\mathbf{U}^{mhd,n+1}_{i,j,k} - \mathbf{U}^{mhd,n}_{i,j,k} + \frac{\Delta t}{\Delta x}(\tilde{\mathbf{F}}^{1,mhd,n}_{i+1/2,j,k} - \tilde{\mathbf{F}}^{1,mhd,n}_{i-1/2,j,k}) &= \Delta t \tilde{\mathbf{S}}^{1,n}_{i,j,k} + \Delta t \mathbf{S}^{g,n}_{i,j,k} + \\
&\qquad + \Delta t \mathbf{S}^{rad,n+1}_{i,j,k}, \\
\mathcal{E}^{n+1}_{i,j,k}(1+\Delta t c\sigma^{n+1}_{i,j,k}) - \mathcal{E}^{n}_{i,j,k} + \frac{\Delta t}{\Delta x}(A^p_{i+1/2,j,k} - A^p_{i-1/2,j,k}) &= +\Delta t c \sigma^{n+1}_{i,j,k} a (T^{n+1}_{i,j,k})^4, \\
\mathcal{F}^{1,n+1}_{i,j,k}(1+\Delta t c\sigma^{n+1}_{i,j,k}) - \mathcal{F}^{1,n}_{i,j,k} + \frac{\Delta t}{\Delta x}(B^{1,p}_{i+1/2,j,k} - B^{1,p}_{i-1/2,j,k}) &= 0, \\
\mathcal{F}^{2,n+1}_{i,j,k}(1+\Delta t c\sigma^{n+1}_{i,j,k}) - \mathcal{F}^{2,n}_{i,j,k} + \frac{\Delta t}{\Delta x}(B^{2,p}_{i+1/2,j,k} - B^{2,p}_{i-1/2,j,k}) &= 0, \\
\mathcal{F}^{3,n+1}_{i,j,k}(1+\Delta t c\sigma^{n+1}_{i,j,k}) - \mathcal{F}^{3,n}_{i,j,k} + \frac{\Delta t}{\Delta x}(B^{3,p}_{i+1/2,j,k} - B^{3,p}_{i-1/2,j,k}) &= 0,
\end{aligned} \tag{2.24}$$

where we get the explicit scheme $M_{\text{HLLC}}R_{\text{HLL}_e}$ for $p = n$, and the semi-implicit scheme $M_{\text{HLLC}}R_{\text{HLL}_i}$ for $p = n+1$. Moreover, the flux differences for the Rusanov type scheme are given by

$$\begin{aligned}
A^p_{i+1/2} - A^p_{i-1/2} &= \frac{c}{2}(\mathcal{F}^{1,p}_{i+1} - \mathcal{F}^{1,p}_{i-1} - \mathcal{E}^p_{i+1} + 2\mathcal{E}^p_i - \mathcal{E}^p_{i-1}), \\
B^{k,p}_{i+1/2} - B^{k,p}_{i-1/2} &= \frac{c}{2}(\mathcal{E}^p_{i+1}D^{k,p}_{i+1} - \mathcal{E}^p_{i-1}D^{k,p}_{i-1} - \mathcal{F}^{k,p}_{i+1} + 2\mathcal{F}^{k,p}_i - \mathcal{F}^{k,p}_{i-1}),
\end{aligned} \tag{2.25}$$

with D^k defined in (2.3).

In the following theorem we summarize the properties of the finite volume schemes $M_{\text{HLLC}}R_{\text{HLL}_e}$, $M_{\text{HLLC}}R_{\text{HLL}_i}$ approximating the solutions of the stratified RMHD equations (1.6).

Theorem 2.1. *Consider the finite volume schemes $M_{HLLC}R_{HLL_e}$, $M_{HLLC}R_{HLL_i}$ approximating the solutions of equation (1.6). Both schemes*

- *are consistent with (1.6) and first order accurate in both space and time (for smooth solutions),*
- *preserve $\mathcal{E} \geq 0$ and $\|\boldsymbol{\mathcal{F}}/\mathcal{E}\| \leq 1$ discretely provided $c, a, \sigma \geq 0$,*
- *are well-balanced, i.e. preserve discrete versions of the steady states (1.10) for any background magnetic field $\tilde{\mathbf{B}}$.*

Proof. The schemes are first order accurate by construction

Next, we show positivity and flux limitation for the explicit scheme. In 1d we have $\mathcal{F}_i^{2,n} = \mathcal{F}_i^{3,n} = 0$. Let us assume that for a fixed n, the two inequalities

$$\mathcal{E}_i^n \geq 0, |\mathcal{F}_i^{1,n}| \leq \mathcal{E}_i^n \tag{2.26}$$

hold true for all i. In that case we can write $\pm\mathcal{F}_i^{1,n} - \mathcal{E}_i^n \leq 0$, which will be used several times in the sequel. Then from the discrete equation (2.24) for radiative energy

$$\mathcal{E}_i^{n+1}(1+\Delta t c\sigma^{n+1}) = \mathcal{E}_i^n - \frac{c\Delta t}{2\Delta x}(\mathcal{F}_{i+1}^{1,n} - \mathcal{F}_{i-1}^{1,n} - \mathcal{E}_{i+1}^n + 2\mathcal{E}_i^n - \mathcal{E}_{i-1}^n) + \Delta t c\sigma^{n+1} a(T_i^{n+1})^4,$$

we can conclude that

$$\mathcal{E}_i^{n+1}(1+\Delta t c\sigma^{n+1}) \geq \mathcal{E}_i^n - \frac{\Delta t}{\Delta x}c\mathcal{E}_i^n + \Delta t c\sigma^{n+1} a(T_i^{n+1})^4 \geq \mathcal{E}_i^n(1 - \frac{\Delta t}{\Delta x}c),$$

by using that $\pm\mathcal{F}_i^{1,n} - \mathcal{E}_i^n \leq 0$. So we have that the radiative energy remains positive, if the CFL condition $\Delta t \leq \frac{\Delta x}{c}$ is fulfilled.

In order to show that the flux is limited, we proceed as follows. Assume that (2.26) holds true for a fixed n and all i. Then we need the prove the following two inequalities $\pm\mathcal{F}_i^{1,n+1} - \mathcal{E}_i^{n+1} \leq 0$, using that we have already shown that $\mathcal{E}_i^n \geq 0$ for all i and n. For the scheme (2.24), we have

$$\begin{aligned}(\pm\mathcal{F}_i^{1,n+1} - \mathcal{E}_i^{n+1})(1+\Delta t c\sigma^{n+1}) =& (\pm\mathcal{F}_i^{1,n} - \mathcal{E}_i^n)(1 - \frac{\Delta t}{\Delta x}c) + \Delta t c\sigma_i^{n+1} a(T_i^{n+1})^4 \\ &- \frac{c\Delta t}{2\Delta x}(\pm\mathcal{E}_{i+1}^n\chi_{i+1}^n \mp \mathcal{E}_{i-1}^n\chi_{i-1}^n \mp \mathcal{F}_{i+1}^{1,n} \mp \mathcal{F}_{i-1}^{1,n} \\ &- \mathcal{F}_{i+1}^{1,n} + \mathcal{F}_{i-1}^{1,n} + \mathcal{E}_{i+1}^n + \mathcal{E}_{i-1}^n).\end{aligned}$$

So the flux stays limited if the above expression on the right hand side is always negative. This is the case, if the CFL condition $\Delta t \leq \frac{\Delta x}{c}$ is fulfilled and the following holds for all i

$$\begin{aligned}(-1 \mp 1)\mathcal{F}_i^{1,n} + (1 \pm \chi_i^n)\mathcal{E}_i^n \geq 0, \\ (1 \mp 1)\mathcal{F}_i^{1,n} + (1 \mp \chi_i^n)\mathcal{E}_i^n \geq 0,\end{aligned}$$

All of them are fulfilled. First of all we have $(1 - \chi_i^n)\mathcal{E}_i^n \geq \frac{1}{3}\mathcal{E}_i^n \geq 0$, under condition (2.26). Second, we see that the expression $\pm 2\mathcal{F}_i^{1,n} + (1 + \chi_i^n)\mathcal{E}_i^n$ is always positive under condition (2.26), by looking at the function (divide by $\mathcal{E}_i^n > 0$. For $\mathcal{E}_i^n = 0 \Rightarrow \mathcal{F}_i^n = 0$ and the above inequalities hold trivially)

$$h^{\pm}(z) = \pm 2z + 1 + \frac{3+4z^2}{5+2\sqrt{4-3z^2}}, \quad -1 \leq z \leq 1$$

In the interval $[-1, 1]$ the function $h^{\pm}$ is continuous and therefore has a minimum. The derivative of this function is nonzero for all $z \in (-1, 1)$. Furthermore, we have that $h^{\pm}(\mp 1) = 0$ and $h^{\pm}(\pm 1) = 4$. From this we conclude that $h^{\pm}(z) \geq 0$ for

$|z| \leq 1$. Summarizing, we have shown that $\pm\mathcal{F}_i^{1,n+1} - \mathcal{E}_i^{n+1} \leq 0$ and therefore that the flux stays bounded.

The explicit scheme preserves the steady states up to machine precision. First of all, the radiation fluxes are all equal to zero at the steady state. Second, the radiation sources are equal to zero and stable points of the remaining ordinary differential equations. In combination with the fact that the MHD part is well-balanced (see [3]) we have that steady states are preserved up to machine precision.

The proof of flux limitation and steady state preservation for the implicit scheme is very similar to the above proof for the explicit scheme and we omit it here.

□

The numerical experiments for (1.6) fall into two different categories. First we test and compare the explicit and semi-implicit scheme described above for wave propagation. It turns out that the semi-implicit solver is several orders of magnitude more efficient in comparison with the explicit solver for radiation hydrodynamics/MHD. The second category consists of a suit of numerical experiments showing the effects of radiation on wave propagation in stratified magneto-atmospheres.

3. **Numerical experiments in 1 dimension/Efficiency study.** In this section we present numerical experiments for testing how robustly and efficiently our schemes of type (2.1) work, comparing the explicit $M_{\mathrm{HLLC}}R_{\mathrm{HLL}_e}$ with the semi-implicit $M_{\mathrm{HLLC}}R_{\mathrm{HLL}_i}$ solver. We show that for the radiation hydrodynamic/MHD equations (1.6) the semi-implicit solver is several orders of magnitude more efficient than the explicit solver.

To begin with, we compare our numerical schemes for radiation hydrodynamics, given by (1.6) with a zero magnetic field $\tilde{\mathbf{B}} = \mathbf{0}$. The initial conditions are a discrete version of the steady state background (1.10) with $\gamma = 5/3, a = 1, p_0 = 1.13, H = 0.158$ and a gravity constant of $g = 2.74$. Furthermore, we choose $\sigma(z,t) = \frac{\rho(z,t)}{\rho(0,0)}$, so that the medium is opaque at the bottom $z = 0$ at time $t = 0$. The domain is given by $z \in [0,1]$.

For numerical simulations concerned with wave propagation in stratified atmospheres it is desirable, if not necessary, to use a well-balanced finite volume scheme, see [3]. As shown in theorem 2.1, both the explicit $M_{\mathrm{HLLC}}R_{\mathrm{HLL}_e}$ and the semi-implicit $M_{\mathrm{HLLC}}R_{\mathrm{HLL}_i}$ preserve the steady state discretely, and are therefore suitable for our task. On top of this steady state background we model the waves by introducing a local sinusoidal (in time) driving of the velocity field perpendicular to the boundary, given by the following boundary condition for the normal velocity at the bottom

$$u_3(0,t) = 0.3\sin(6\pi t). \tag{3.1}$$

As time evolves, those waves move up through the domain and are modified by the stratified RMHD equations. In figure 2 we present the results for the well-balanced explicit $M_{\mathrm{HLLC}}R_{\mathrm{HLL}_e}$ and semi-implicit $M_{\mathrm{HLLC}}R_{\mathrm{HLL}_i}$ schemes at time $t = 0.8$ for different meshes. We can see that the waves are resolved very well and the semi-implicit solver seems to have slightly more accuracy compared to the explicit solver on the same grid, at least in the plasma variables ρ, m_3 and E, despite the fact that it uses much less computational time.

To quantify the efficiency of the explicit $M_{\mathrm{HLLC}}R_{\mathrm{HLL}_e}$ compared with the semi-implicit $M_{\mathrm{HLLC}}R_{\mathrm{HLL}_i}$ scheme, we first compute a reference solution with the explicit scheme on a very fine grid. Then we plot the computational time over the relative

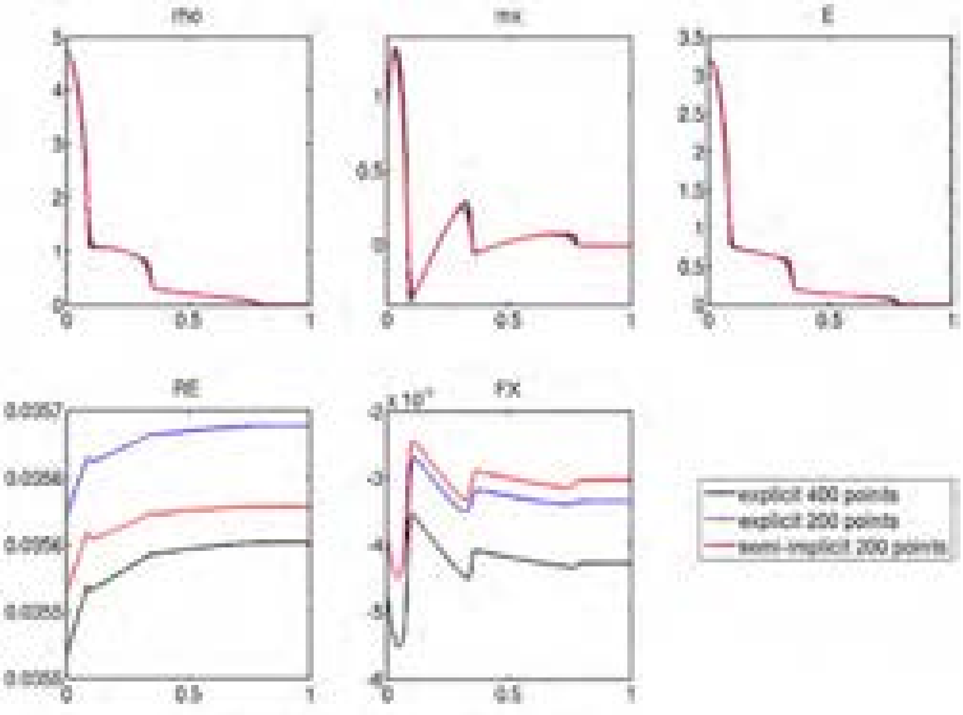


FIGURE 2. Comparison of explicit solver $M_{\mathrm{HLLC}}R_{\mathrm{HLL}_e}$ with semi-implicit solver $M_{\mathrm{HLLC}}R_{\mathrm{HLL}_i}$ for different CFL numbers at time $t = 0.8$. From top to bottom and left to right: density ρ, momentum m_3, energy E, radiative energy $\mathcal{E}$ and radiative flux $\boldsymbol{\mathcal{F}}$.

errors with respect to this reference solution for both the explicit and semi-implicit scheme on different mesh resolutions. We can learn from figure 3 that for the same relative error the explicit $M_{\mathrm{HLLC}}R_{\mathrm{HLL}_e}$ solver is $\mathcal{O}(10^4)$ slower compared with the semi-implicit $M_{\mathrm{HLLC}}R_{\mathrm{HLL}_i}$ solver. The reason for this tremendous difference in efficiency is the following. The maximum eigenvalue of the hydrodynamic equations compared to the speed of light is $c/\max(\lambda_{\mathrm{hydro}}(t)) = \mathcal{O}(10^4)$. That means we do $\mathcal{O}(10^4)$ more time steps with the explicit solver. There is no loss of accuracy due to the large time steps because the waves in stratified RMHD (1.6) are induced by a forcing of the hydrodynamic variable u_3. Therefore, the waves in the radiation variables $\mathbf{U}^{rad}$ are rather weak and without strong shocks, and the radiation part can be well approximated by a rather diffusive implicit solver. We can conclude that the higher cost of solving a system of nonlinear equations at each time step is more than compensated for by allowing much larger time steps. So in this case the semi-implicit scheme $M_{\mathrm{HLLC}}R_{\mathrm{HLL}_i}$ clearly wins over the explicit version in terms of efficiency.

In the next section we focus on the influence of radiative transfer on wave propagation in stratified atmospheres.

4. **Numerical experiments in 2 dimensions/Comparison of MHD with radiation MHD.** In this section we compare the solution of the MHD equations with the solution of the RMHD equations on a suit of numerical experiments using

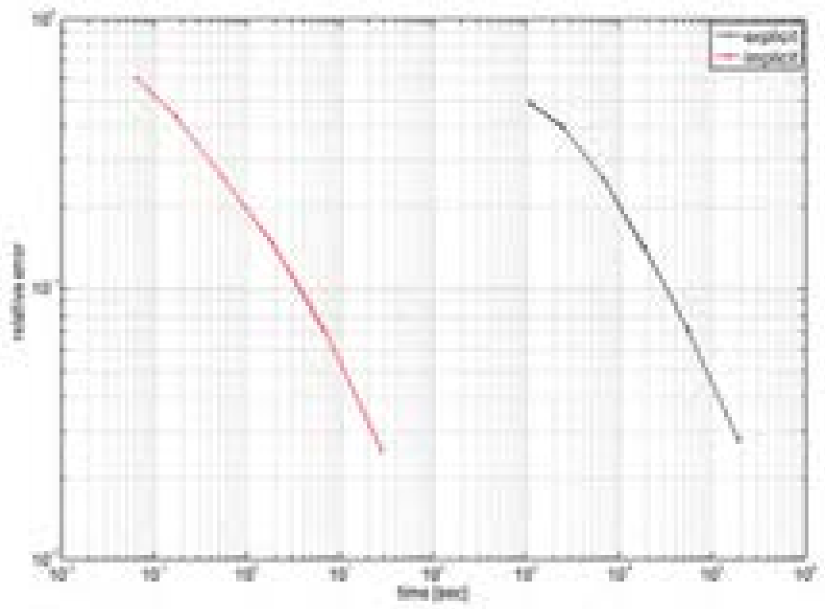


FIGURE 3. Comparison of the efficiency of the explicit $M_{\text{HLLC}}R_{\text{HLL}_e}$ and semi-implicit $M_{\text{HLLC}}R_{\text{HLL}_i}$ solver for the solution at time $t = 0.8$. The relative error is the sum of the relative errors of all variables (divided by number of variables).

the explicit solver $M_{\text{HLLC}}R_{\text{HLL}_e}$. The first order scheme is run with a CFL number of 0.45 and for the constants in the model we choose for the acceleration due to gravity, $g = 2.74$, constant $H = 0.158$, gas constant $\gamma = 5/3$ and initial pressure $p_0 = 1.13$. All subsequent two-dimensional experiments are performed on the domain $[x, z] \in [0, 4] \times [0, 1]$. Again, we choose $\sigma(z, t) = \frac{\rho(z,t)}{\rho(0,0)}$, so that the medium is opaque at the bottom $z = 0$ at time $t = 0$.

We want to compare the MHD equations with the radiation MHD equations (1.6). Since we are concerned with wave propagation in startified atmospheres we follow the setup described in articles [1, 2, 3]. We choose a discrete version of the steady state (1.10) with different background magnetic fields $\tilde{\mathbf{B}}$. A small part of the bottom boundary acts a piston and sends in temporally sinusoidal waves, perturbing the steady state. Those boundary conditions are given by a forcing in the normal velocity field, namely

$$u^3(x, 0) = 0.3e^{-100(x-1.9)^2} \sin(6\pi t)\, \mathbf{1}_{\{[1.65,2.15]\}}. \tag{4.1}$$

We start with the simplest case, i.e. in absence of a magnetic field.

4.1. **Hydrodynamics vs radiation hydrodynamics.** The setup in the pure hydrodynamic case is given by chosing the embedded magnetic field $\tilde{\mathbf{B}}$ to be zero. The results are shown in figure 4. The top row depicts the solution at time $t = 1$ for standard hydrodynamics and the bottom row the one for hydrodynamics combined with the M1 model. In order to better compare the two cases, the solution for each variable uses the same scaling for MHD and RMHD in the figure. By comparing

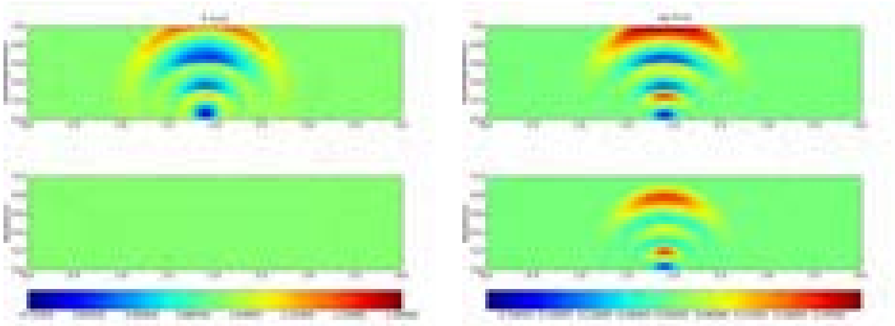


FIGURE 4. Comprison of hydrodynamics with radiation hydrodynamics. Solution from the explicit solver $R_{\mathrm{HLL}_e S_i}$ at $t = 1$ for 400×100 mesh points. Left: relative temperature change. Right: vertical velocity u_3. Each variable has the same scaling for MHD and RMHD.

the plots for temperature in figure 4 we can immediatly see that the temperature variations introduced by the boundary conditions (4.1) of radiation hydrodynamics are too small to be seen compared to the results of hydrodynamics. This means that radiative transfer compensates the (by the boundary conditions) introduced temperature variations. This has a profound effect on the velocity of the wave fronts propagating through the medium. We can see on the right of figure 4 that the first wave front for hydrodynamics without radiation is about to exit the domain on the top at time $t = 1$. In contrast, the velocity plot for radiation hydrodynamics at the same time $t = 1$ reveals that the leading wave front is still a distance away from the top boundary and is therefor clearly propagating with a slower speeed due to the action of radiative transfer. Furthermore, we can see that the amplitude of the velocity disturbances is reduced if we account for radiative transport. The overall dynamics are, however, comparable for both cases.

We continue by studying the effects of radiative transport in the case of complicated nontrivial magnetic fields.

4.2. **Comparison of MHD with Radiation MHD.** The above experiment was the comparison of hydrodyamics with radiation hydrodynamics since the magnetic field stayed zero during the whole computation. In order to see the effect of radiative transfer on the solution of the MHD equations in two dimensions, we choose a *realistic* two-dimensional background magnetic field in the following way. We let $\widetilde{B}_3(x,0,0)$ approximate

$$\widetilde{B}_3(x,0,0) = 2.7e^{-7.2r^2} - 1.3e^{-40(r-0.6)^2}, \; r = |x-2|, x \in [0,4] \tag{4.2}$$

by using a Fourier expansion of vector harmonic functions (see also [1, 2, 3]), i.e.

$$\widetilde{B}_1(x,y,z) = \sum_{l=0}^{L} f_l \sin(2\pi l x)\, e^{-2\pi l z}, \quad \widetilde{B}_3(x,y,z) = \sum_{l=0}^{L} f_l \cos(2\pi l x)\, e^{-2\pi l z}, \tag{4.3}$$
$$B_2(x,y,z) \equiv 0,$$

where the f_l's are Fourier coefficients corresponding to the background magnetic field at the bottom of the domain and L is the total number of Fourier modes.

The computations presented here use the first fourteen terms in the Fourier series. The full magnetic field $\widetilde{\mathbf{B}}(x,y,z)$ follows then from the potential field assumption, i.e. (4.3). The resulting potential field consists of a large unipolar magnetic flux concentration surrounded on each side by two smaller concentrations of opposite polarity field (see[1], figure 1 for an illustration). The rest of the constants and σ is chosen as in the case above.

The numerical results are presented in figure 4.2. Again, each variable is scaled in the same way, in order to allow for a good comparison of the solutions of MHD and RMHD. As is expected (see [1, 2, 3]), the waves are more focused compared to the hydrodynamics case due to the presence of the nonzero magentic field. Looking at the temperature plot, we again see that the radiative transport results in an almost constant temperature distribution compared to the solution of the standard MHD equations. Looking at the velocity in the direction of the magentic field, we again see that the wave fronts are propagating with a smaller speed in the case of RMHD. This difference is less prominent if we consider the velocity perpendicular to the magentic field.

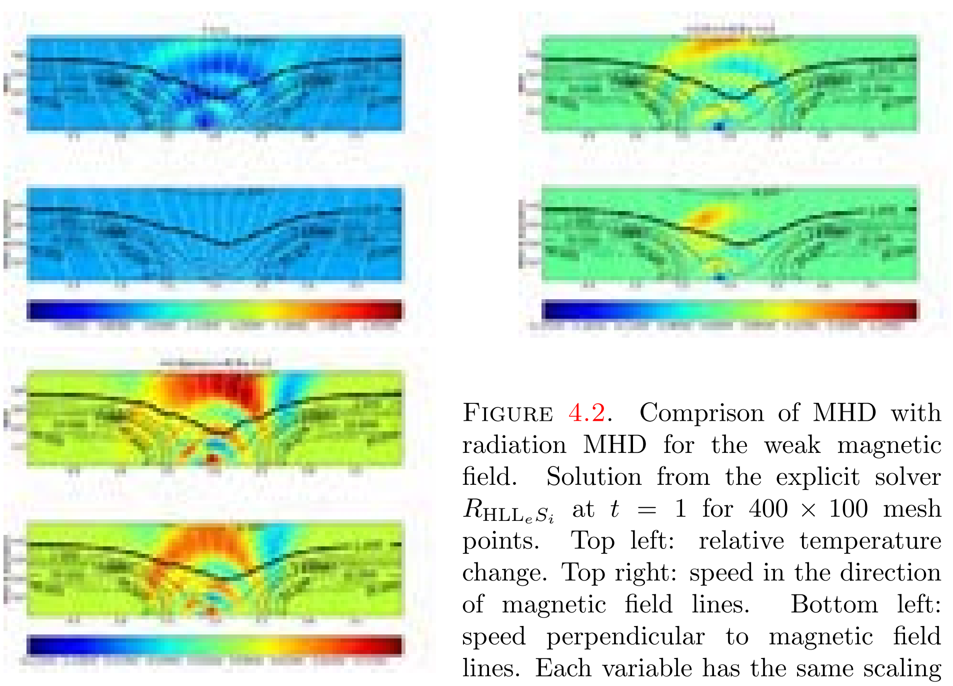


Figure 4.2. Comprison of MHD with radiation MHD for the weak magnetic field. Solution from the explicit solver $R_{\mathrm{HLL}_e S_i}$ at $t = 1$ for 400×100 mesh points. Top left: relative temperature change. Top right: speed in the direction of magnetic field lines. Bottom left: speed perpendicular to magnetic field lines. Each variable has the same scaling for MHD and RMHD.

In this example the magentic field is still of moderate strength. We therefor increase our the magnetic field (4.2) by a factor of 3. The results are presented in figure 4.2. In this case the waves are even more focused due to the action of the Lorentz force, and follow the magnetic field lines (white). As before, it becomes apparent from the temperature plot that the temperature is kept almost constant in the case of RMHD in contrast to the solution of the MHD equations. Looking at the velocity in the direction of the magnetic field, we can again conclude that the overall behaviour is the same, but the speed of the wave fronts as well as the

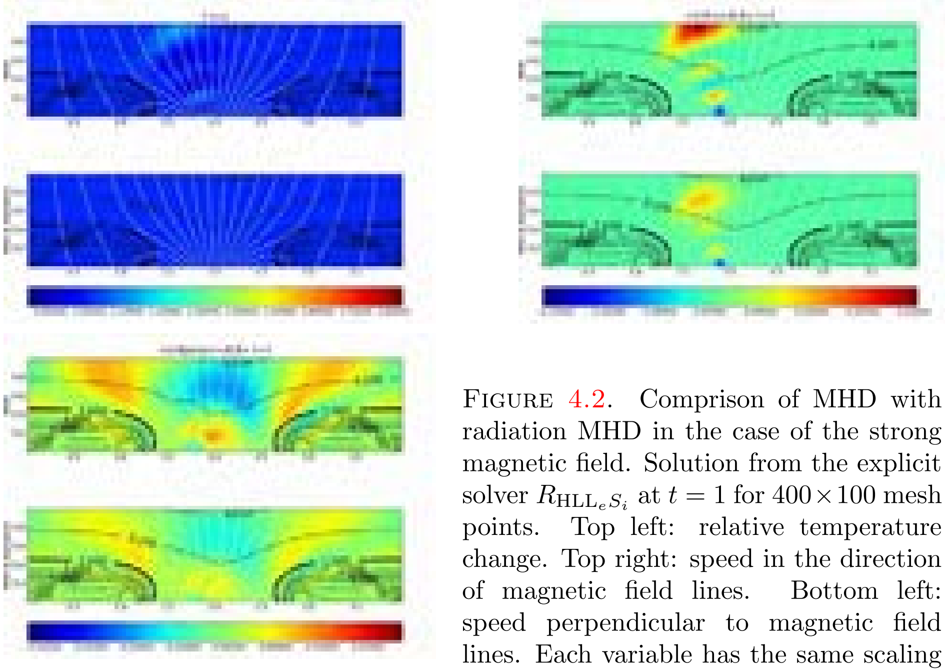


FIGURE 4.2. Comprison of MHD with radiation MHD in the case of the strong magnetic field. Solution from the explicit solver $R_{\mathrm{HLL}_e S_i}$ at $t = 1$ for 400×100 mesh points. Top left: relative temperature change. Top right: speed in the direction of magnetic field lines. Bottom left: speed perpendicular to magnetic field lines. Each variable has the same scaling for MHD and RMHD.

amplitude of them is reduced due to radiative transfer. This can also be seen in the plot for the velocity in the direction perpendicular to the magnetic field.

5. **Conclusion.** Wave propagation in the solar atmosphere is a very important mode of energy transport in the sun and plays an essential role in many interesting solar phenomena, particularly in chromospheric and coronal heating. Solar wave propagation can be modeled in terms of the equations of stratified radiative MHD. As the standard radiative transport equation is high-dimensional, reduced models are preferred. We focus on a particular reduced model, the so-called M1 model that accounts for radiation in terms of spectral radiative moments. The resulting M1-stratified MHD coupled system is then simulated using numerical schemes.

We consider two sets of numerical schemes in this paper. Both schemes are based on the coupling between the MHD and radiation parts in terms of source terms. Thus, standard HLLC and two-wave HLL solvers can be used to form the numerical fluxes in a finite volume framework. The simplest form of time stepping is the forward Euler time stepping. It has to augmented with implicit treatment of the stiff radiative source term. The resulting scheme works quite well. However, it is computationally very expensive as the relevant speed of the system, that is used in setting the time step through the CFL condition for the explicit scheme, is given by the speed of light. This is three to four orders of magnitude larger than the fastest magneto sonic waves of the system.

Consequently, we couple an explicit forward Euler discretization of the MHD flux and source terms with an implicit backward Euler discretization of the radiative flux and source terms. This coupled semi-implicit scheme allows us to determine the time step in terms of the magneto sonic waves and allows for time steps that are orders of magnitude larger than the fully explicit scheme. Numerical experiments show that the semi-implicit scheme is three to four orders of magnitude more efficient than the fully explicit scheme.

We conclude with a suit of numerical experiments for the propagation of waves in the sun. Both weak and strong magnetic fields are used and the numerical experiments indicate that the schemes work quite well with sharp resolution of the waves. Compared to the absence of radiation, adding radiation implies greater uniformity in temperature distributions as a consequence of radiative cooling. Furthermore, this cooling leads to the slowing down of propagating waves. Furthermore, the wave amplitude is also reduced as energy is taken out of the system on account of radiative cooling. This energy loss poses a considerable obstacle for wave propagation explaining coronal heating. More elaborate physical mechanisms are needed to explain this heating.

In terms of implementation, we use a Newton method to solve the resulting nonlinear algebraic system of equations at each time step. Currently, a direct method for inverting the linearized equations within each Newton step, is used. However, practical efficiency dictates the use of an iterative krylov type methods for solving the resulting linear equations. Such iterative methods suffer from ill-conditioning. The design of an efficient preconditioner is a prerequisite for the efficient solution of the nonlinear equations and will be the subject of a forthcoming paper.

REFERENCES

[1] T. J. Bogdan *et al.* Waves in the magnetized solar atmosphere II: Waves from localized sources in magnetic flux concentrations. *Astrophys. Jl,* 599, 2003, 626 - 660.

[2] F. Fuchs, A. D. McMurry, S. Mishra, N. H. Risebro and K. Waagan. Finite volume schemes for wave propagation in stratified magneto-atmospheres. *Comm. Comput. Phys.,* 7 (3), 2010, 473-509.

[3] F. Fuchs, A. D. McMurry, S. Mishra, N. H. Risebro and K. Waagan Well-balanced high-order finite volume methods for simulating wave propagation in stratified magneto-atmospheres. *Jl. Comput. Phys.,* 229 (11), 2010, 4033-4058.

[4] D. Mihalas and G. Mihalas. Foundation of Radiation Hydrodynamics. *Oxford university press,* Oxford, 1984.

[5] B. Dubroca and J. L. Feugeas. Entropic moment closure hierarchy for the radiative transfer equations. *C. R. Acad. Sci Paris*, Ser 1, 329, 915-920, 1999.

[6] C. Berthon, P. Charrier and B. Dubroca. An HLLC scheme to solve the M1 model of radiative transfer in two space dimensions. *Jl. Sci. Comp.,* 31 (3), 347-389, 2007.

IMPLICIT-EXPLICIT RUNGE-KUTTA SCHEMES FOR HYPERBOLIC SYSTEMS WITH STIFF RELAXATION AND APPLICATIONS

Sebastiano Boscarino and Giovanni Russo

Dipartimento di Matematica e Informatica - Viale
A. Doria n° 6 - 95125 Catania, Italy

Abstract. In this paper we give an overview of Implicit-Explicit Runge-Kutta schemes applied to hyperbolic systems with stiff relaxation. In particular, we focus on some recent results on the uniform accuracy for hyperbolic systems with stiff relaxation [6], and hyperbolic system with diffusive relaxation [7, 5, 4]. In the latter case, we present an original application to a model problem arising in Extended Thermodynamics.

1. **Introduction.** Many physical models are described by hyperbolic systems with relaxation of the form

$$\partial_t U + \partial_x F(U) = \frac{1}{\varepsilon} R(U), \quad x \in \mathbb{R}, \tag{1}$$

with $U = U(x,t) \in \mathbb{R}^N$, $F : \mathbb{R}^N \to \mathbb{R}^N$. Such systems are said hyperbolic if the Jacobian matrix $F'(U)$ has real eigenvalues and a basis of eigenvectors $\forall\, U \in \mathbb{R}^N$. Usually, the parameter ε is called the relaxation time, which is small in many physical situations. Here we use the term relaxation in the sense of Whitham [25] and Liu ([19]), which in practice means that if $\varepsilon \to 0$, the system formally relaxes to a quasilinear hyperbolic system with a smaller number of dimensions. Chen, Levermore, and Liu [10] provide the proper condition that ensures that the solution of the relaxation system actually converges to the solution of the *relaxed* system.

Typical examples of such systems are: gas dynamics with chemical reactions, shallow water with friction, discrete kinetic models, extended thermodynamics, hydrodynamical models for semiconductors, traffic flow models, granular gases (see [21] and references therein).

A simple prototype example of relaxation system is given by

$$\begin{aligned} \partial_t u + \partial_x v &= 0, \\ \partial_t v + \partial_x p(u) &= -\frac{1}{\varepsilon}(v - f(u)), \end{aligned}$$

which corresponds to $U = (u, v)$, $F(U) = (v, p(u))$, $R(U) = (0, f(u) - v)$. As $\varepsilon \to 0$ we get, formally, the local equilibrium $v = f(u)$ while u satisfies the conservation equation

$$\partial_t u + \partial_x f(u) = 0.$$

2000 *Mathematics Subject Classification.* Primary: 65C20, 65L06; Secondary: 65M06, 76D05, 82C40.

Key words and phrases. Runge-Kutta methods, stiff problems, hyperbolic systems with relaxation, diffusion equations.

In [10] the authors proved that the solution u actually converges to the solution of the relaxed equation if the characteristic speed of the relaxed equation is contained in the interval identified by the speed of the original system,i.e. $p'(u) \geq (f'(u))^2$, i.e.the *subcharacteristic condition.*

The most commonly used approach for the numerical solution of hyperbolic system with relaxation is based on the Method Of Line (MOL). First we discretize the system in space, leading to a large system of ODEs defined on a grid. The semi discrete scheme should be high resolution shock capturing, which provide correct shock location without numerical oscillations. Among space discretization techniques we mentioned several possibilities: Finite Volume (FV), Finite Difference (FD), Discontinuous Galerkin (DG). Method of lines based on conservative finite difference is the simplest choice for the construction of high order schemes in space and time [22, 21]. For example, in one space dimension, the scheme reads:

$$\frac{du_j}{dt} = -\frac{\hat{f}_{j+1/2} - \hat{f}_{j-1/2}}{dx} - g(u_j)$$

with

$$\hat{f}_{j+1/2} = \hat{f}^+_{j+1/2}(x^-_{j+1/2}) + \hat{f}^-_{j+1/2}(x^+_{j+1/2}).$$

The numerical flux $\{\hat{f}^{\pm}_{j+1/2}(x)\}$ being reconstructed from the fluxes $f^{\pm}(x_j)$, which in turn split the analytical flux: $f = f^+ + f^-$, $\lambda(\nabla f^+) \geq 0$, $\lambda(\nabla f^+) \leq 0$. High order reconstruction can be obtained, for example, by ENO or WENO reconstruction from cell averages to pointwise values,

$$\{f^{\pm}_j\} \xrightarrow{WENO} \hat{f}^{\pm}_j(x_{j\pm 1/2})$$

Since source term $g(u_j)$ is computed *pointwise* then the various cells are not coupled at the level of the source, and the implicit equations in each cell are independent from each other.

Applying MOL to hyperbolic system with relaxation, the PDEs become a system of ODEs of the form

$$u' = f(u) + \frac{1}{\varepsilon}g(u), \tag{2}$$

with initial vector $u_0 = (U(x_1,t_0), \cdots, U(x_N,t_0))^T$, where $\{x_i\}_{i=1}^N$ denote the spatial computational mesh. The solution at time t is $u(t) = (u(t_1), u_2(t), \cdots, u_N(t))^T$ where $u_i(t) \approx U(x_i,t)$. The term $f(u)$ represents the discretization of the convective derivative term, $-\partial_x F(U)$, while $g(u)$ represents the discrete approximation of the source term, $G(U)$, on the grid nodes (and possibly the boundary conditions). Then a suitable time integrator is used to solve ODEs.

In most cases $f(u)$ is non stiff and non linear while $\frac{1}{\varepsilon}g(u)$ contains the stiffness, so we look for numerical schemes which are explicit in f and implicit in g. In particular it is essential that the numerical scheme is accurate for $\varepsilon \to 0$ (possibly also for intermediate regimes of such parameters, i.e. when ε is not too small). Moreover some stability restrictions are required, i.e. for the convection term $\Delta t \leq \rho(\nabla_u F)\Delta x$ (CFL condition). The stiff term has to be treated implicitly to avoid restrictions $\Delta t \leq C\varepsilon$.

IMEX Runge-Kutta methods represents a very effective tool to guarantee the simplicity of the explicit treatment of the non-stiff term $f(u)$ and to avoid time restriction because of the stiffness in the source term $g(u)$.

An Implicit-Explicit (IMEX) Runge-Kutta scheme applied to system (2) takes the form

$$\begin{aligned} Y_i &= y_0 + h\sum_{j=1}^{i-1}\tilde{a}_{ij}f(t_0+\tilde{c}_jh, Y_j) + h\sum_{j=1}^{i}a_{ij}\frac{1}{\varepsilon}g(t_0+c_jh, Y_j), \\ y_1 &= y_0 + h\sum_{i=1}^{s}\tilde{b}_if(t_0+\tilde{c}_ih, Y_i) + h\sum_{i=1}^{s}b_i\frac{1}{\varepsilon}g(t_0+c_ih, Y_i). \end{aligned}$$

where $\tilde{A} = (\tilde{a}_{ij})$, $\tilde{a}_{ij} = 0$, $j \geq i$ and $A = (a_{ij})$ are $s \times s$ (lower triangular) matrices and $\tilde{c}, \tilde{b}, c, b \in \mathbb{R}^s$, coefficient vectors. A classical representation of a IMEX R-K method is given by

$$\text{Double Butcher } \textit{tableau}: \quad \begin{array}{c|c} \tilde{c} & \tilde{A} \\ \hline & \tilde{b}^T \end{array} \qquad \begin{array}{c|c} c & A \\ \hline & b^T \end{array}.$$

We restrict to consider IMEX schemes in which the implicit part is a diagonally implicit Runge-Kutta (DIRK). Besides it simplicity, this will ensure that f is always evaluated explicitly.

We can classify each IMEX Runge-Kutta scheme by considering the different structures of the matrix $A = (a_{ij})_{i,j=1}^s$, of the implicit scheme:

- (Methods of Type A) The matrix A is invertible.
- (Methods of Type CK)

$$A = \begin{pmatrix} 0 & 0 \\ a & \hat{A} \end{pmatrix}$$

  The submatrix $\hat{A}$ is invertible.

CK methods with $a = 0$ are called ARS methods [1]. Type A methods are somehow more difficult to construct, but easier to analyze than methods of type CK [9] or ARS.

The rest of the paper is organized as follows. Section 2 review some recent results on the development of high-order implicit-explicit (IMEX) Runge-Kutta (R-K) schemes suitable for time-dependent partial differential systems [6]. In section 3 we discuss hyperbolic systems with stiff diffusive relaxation. The last section is devoted to some applications to some models of diffusive relaxation, which confirm practice the advantageous effects of the approaches introduced the earlier sections. In particular, Sec. 4.2 is devoted to an original application of IMEX-I schemes without parabolic restriction to a one dimensional model problem arising in the context of Extended Thermodynamics.

2. On the uniform accuracy of IMEX Runge-Kutta schemes and applications to hyperbolic systems with relaxation.

Usually, under-resolved numerical schemes may yield spurious numerical solutions that are unphysical. Other times, in the case of hyperbolic systems with stiff terms, high order schemes may reduce to lower order when the time step fails to resolve the small relaxation time.

IMplicit-EXplicit (IMEX) Runge-Kutta (R-K) schemes have been widely used for the time evolution of hyperbolic partial differential equations but some of the schemes existing in literature do not exhibit uniform accuracy with respect to the relaxation time. Classical high-order IMEX R-K schemes fail to maintain the high-order accuracy in time in the whole range of the relaxation time and in particular in the asymptotic limit $\varepsilon \to 0$.

In [6] we developed new IMEX R-K schemes for hyperbolic systems with relaxation that present better uniform accuracy than the ones existing in the literature and in particular produce good behavior with high order accuracy in the asymptotic limit, i.e. when ε is very small. In particular, these schemes are able to handle the stiffness of the system (1), in a whole range of the relaxation time.

The schemes are obtained by imposing new additional conditions on their coefficients, in order to guarantee better accuracy over a wide range of the relaxation time. Following the same technique proposed in [11], the additional conditions are obtained by performing an asymptotic expansion of the exact and numerical solution in the small parameter ε (Hilbert expansion), and by matching the two solutions to various order in ε, [3].

The construction of a high-order accurate IMEX R-K scheme is obtained by imposing the extra order conditions, that ensure the agreement between exact and numerical solution up to a given order in ε. The scheme, called BHR(5,5,3), is presented in [3, 6]. Numerical tests on several ordinary differential systems and hyperbolic systems with relaxation term present better behavior for the new scheme BHR(5,5,3) over other IMEX R-K methods previously existing in literature [1, 9, 21]. For example, by imposing the additional order conditions to the zeroth-order in ε, the classical ARS(4,3,4) scheme can be modified (hereafter called Mod-ARS(3,4,3)), imposing its accuracy in the algebraic variable. Furthermore, by imposing conditions to terms up to fist order in ε and we constructed scheme RHR(5,5,3), a third order five stage scheme.

The construction of this type of IMEX R-K scheme is motivated by the order reduction of classical IMEX schemes observed when applying them to several stiff systems. An example of such behavior is illustrated in Figure 1, where the classical Van rer Pol equation is solved by ARS(3,4,3), Mod-ARS(3,4,3) and BHR(5,5,3) schemes derived in [1, 9, 21, 3, 6],

$$\begin{aligned} y' &= z, \\ \varepsilon z' &= (1-y^2)z - y, \end{aligned} \tag{3}$$

(for details of this problem and its initial conditions see, for example, [11]). The global error behaves like $C\Delta t^r$ with r the slope of the straight line and C is a constant. We observe that, while classical schemes, as ARS(3,4,3), are able to maintain the classical order of accuracy in the differential variable y, they lose accuracy in the algebraic variable z. BHR(5,5,3) method exhibits the better error estimate with respect to ARS(3,4,3) and Mod-ARS(3,4,3) schemes and no order reduction appears when ε is very small.

Concerning hyperbolic systems with stiff relaxation we report here a numerical test the Broadwell model equations

$$\begin{aligned} \rho_t + m_z &= 0, \\ m_t + z_x &= 0, \\ z_t + m_x &= \tfrac{1}{\varepsilon}(\rho^2 + m^2 - 2\rho z) \end{aligned} \tag{4}$$

(for details see [6]), which, in one space dimension, is a 3×3 semilinear hyperbolic system that, in in the relaxed limit, becomes a quasilinear hyperbolic system for the two two differential variables (ρ and m), while z becomes a function of the other two variables.

Figure 2 represents the convergence rate of some IMEX R-K scheme computed on a smooth test problem by grid refinement using three different grids. We have

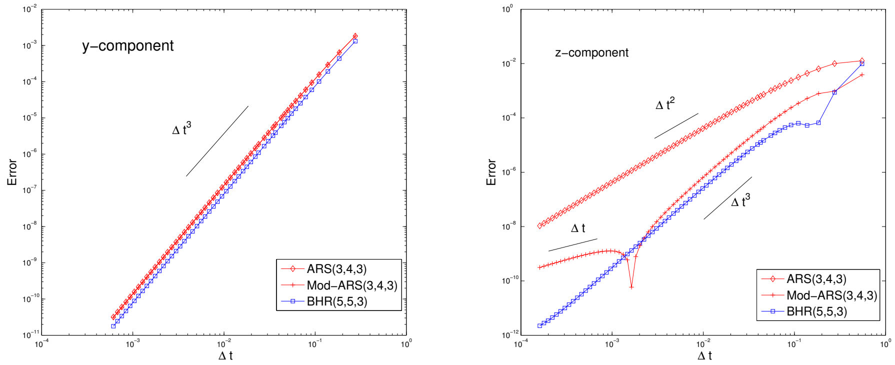


FIGURE 1. Global error versus the stepsize in the Van der Pol equation calculated with $\varepsilon = 10^{-6}$.

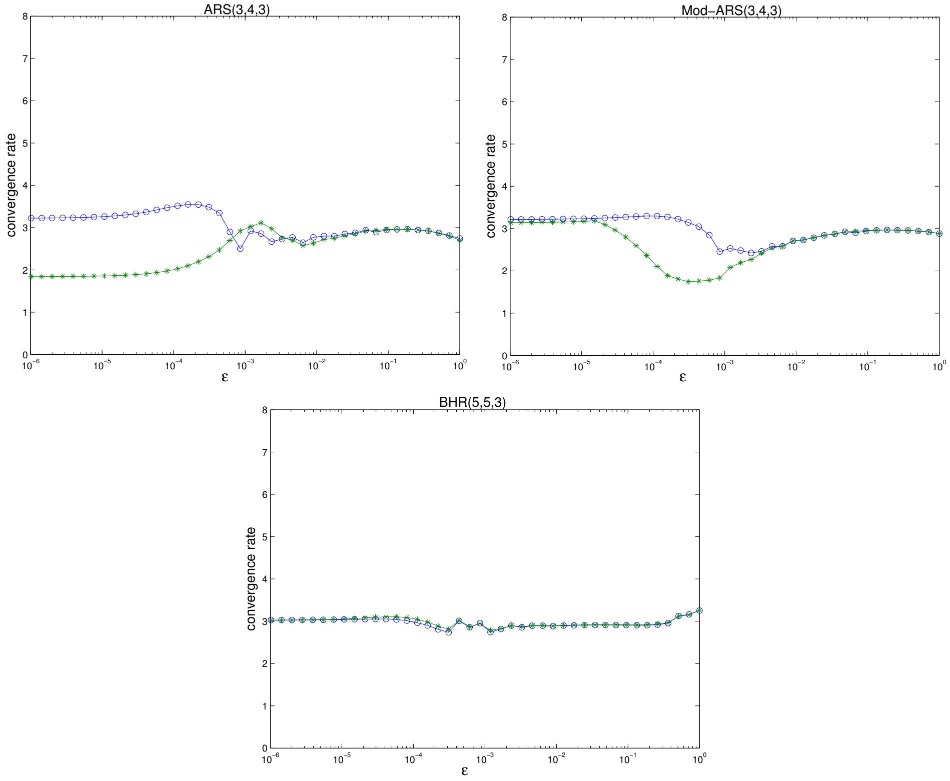


FIGURE 2. Convergence rate vs ε for the density ρ ($\circ$) (differential component) and the flux of the momentum z ($*$) (stiff component). Top: left panel ARS(3,4,3) scheme, right panel Mod-ARS(3,4,3) scheme. Botton: BHR(5,5,3) scheme.

obtained an improvement for the convergence of algebraic component for the Mod-ARS(4,3,4) scheme. In fact, on the left panel we have increased the convergence rate for sufficiently stiff parameters ($\varepsilon < 10^{-4}$). These results show a third-order accuracy for small and large values of ε and note that for intermediate values of the parameter ε ($10^{-4} < \varepsilon < 10^{-2}$) we have a slight deterioration of the accuracy. As it is evident in the right panel from the figure 2 BHR(5,5,3) shows an almost uniform third-order accuracy in the whole range of ε.

3. IMEX Runge-Kutta schemes for hyperbolic systems with diffusive relaxation.

The purpose of this section is to give a review on effective methods for the numerical solution of hyperbolic systems with diffusive relaxation.

As the relaxation parameter vanishes, the characteristic speeds of the system diverge, and the system reduces to a parabolic-type equation (typically a convection-diffusion equation).

A simple prototype of hyperbolic system with relaxation term is given by:

$$\partial_\tau u + \partial_\xi V = 0,$$
$$\partial_\tau V + \partial_\zeta p(u) = -\frac{1}{\varepsilon}(V - Q(u))$$

where $u = u(x,\tau), V = V(x,\tau) \in \mathbb{R}$, and $\varepsilon > 0$ is the relaxation time.

When looking for long time behavior of the solution of the previous system, it is more appropriate to rescale time and the variable V, according to the so called diffusive scaling:

$$\tau = t/\varepsilon, \quad V = \varepsilon v, \quad \xi = x, \quad q(u) = Q(u)/\varepsilon,$$

thus obtaining the general diffusive relaxation system given by:

$$\begin{array}{rcl} \partial_t u + \partial_x v &=& 0, \\ \partial_t v + \dfrac{1}{\varepsilon^2}\partial_x p(u) &=& -\dfrac{1}{\varepsilon^2}(v - q(u)) \end{array}, \tag{5}$$

where $p'(u) > 0$. This system is hyperbolic with two distinct real characteristics speed $\sqrt{p'(u)}/\varepsilon$.

In the small relaxation limit, $\varepsilon \to 0$ the system relaxes towards the system

$$\begin{array}{rcl} \partial_t u + \partial_x q(u) &=& \partial_{xx} p(u), \\ v &=& q(u) - \partial_x p(u). \end{array} \tag{6}$$

The sub characteristic conditions, [10], is automatically satisfied for small ε ($|q'(u)|^2 < p'(u)/\varepsilon^2$), i.e. the main stability condition for the diffusive relaxation system. The simplest form of (5) is to assume $p(u) = u$ and $q(u) = 0$, then from (6) we obtain the classical heat equation $u_t = u_{xx}$.

The attention is devoted to the construction of methods for the numerical solution of system (5) that are able to capture the asymptotic behavior as $\varepsilon \to 0$. Solving (5) numerically is challenging due to the stiffness of the problem both in the convection and in the relaxation terms.

In general, Implicit-Explicit (IMEX) Runge-Kutta schemes represent a powerful tool for the time discretization of stiff systems. Unfortunately, since the characteristic speed of the hyperbolic part is of order $1/\varepsilon$, standard IMEX Runge-Kutta schemes developed for hyperbolic systems with stiff relaxation [1, 9, 21, 6] fail in such parabolic scaling, because the CFL condition would require $\Delta t = \mathcal{O}(\varepsilon \Delta x)$. Of course, in the diffusive regime where $\varepsilon < \Delta x$, this is very restrictive since for an explicit method a parabolic condition $\Delta t = \mathcal{O}(\Delta x^2)$ should suffice.

Most previous work on asymptotic preserving schemes for hyperbolic systems and kinetic equations with diffusive relaxation focus on schemes which in the limit of in the finite stiffness become consistent explicit schemes for the diffusive limit equation [20, 14, 16, 18]. In those paper the authors separate the hyperbolic part into a non stiff and a stiff part and bring the stiff part to the r.h.s., treating it implicitly. As we shall see, this can be explicitly done in several diffusive relaxation models. In all above approaches the resulting schemes, the limit scheme as $\varepsilon \to 0$ are an explicit scheme for the diffusion-like equation, with the usual parabolic CFL restriction on the time step: $\Delta t \approx \Delta x^2$. Schemes that avoid such time step restriction and provide fully implicit solvers have been analyzed in [7, 5], where a new formulation of the problem (5) was introduced. In the next section we review two different approaches in order to treat problem (5) and some generalizations.

3.1. **Removing parabolic stiffness.** The schemes constructed with the approach outlined above converge to an explicit scheme for the limit diffusion equation, i.e. heat equation, and therefore they are subject to the classical parabolic CFL restriction $\Delta t \le C\Delta x^2$. In order to overcome such a restriction we adopt a penalization technique, based on adding two opposite terms to the first equation in (5), and treating one explicitly and one implicitly.

Let us consider the simplest example of hyperbolic system with parabolic relaxation, obtained by setting $q(u) = 0$ and $p(u) = u$ in Eqs.(5). By adding and subtracting the same term on the right hand side we obtain:

$$\begin{aligned} u_t &= -(v + \mu u_x)_x + \mu u_{xx}, \\ v_t &= -\tfrac{1}{\varepsilon^2}(u_x + v). \end{aligned} \tag{7}$$

In the first equation the term $-(v + \mu u_x)_x$ will be treated explicitly, while the second term is treated implicitly. IMEX schemes based on this approach will be called IMEX-I, to remind that the term containing u_x in the second equation is implicit, in the sense that it appears at the new time level.

Notice that the term $v + u_x$ appearing in the second equation is formally treated implicitly, but in practice u_x is computed at the new time from the first equation, so it can in fact explicitly computed.

The function $\mu : \mathbb{R}^+ \to [0, 1]$ must be such that $\mu(0) = 1$, so that in the limit $\varepsilon \to 0$ the quantity $(v + \mu u_x)_x$ vanishes. For $\varepsilon \ll 1$ such a quantity is very small, and so this term can be treated explicitly. As $\varepsilon \to 0$ the method becomes an *implicit scheme* for the limit equation, therefore the parabolic restriction on the time step is removed.

Linear stability analysis can be performed on this simple problem, for the first order IMEX scheme, i.e. a backward-forward Euler method, both in the space continuous case, and using classical central differencing to approximate the first derivatives. For small values of ε and for $\mu = 1$ one obtains the following stability conditions (in the continuous case in space)

$$\xi^2 \Delta t \le \frac{1 - 4\varepsilon^2\xi^2}{4\varepsilon^2\xi^2} \quad ,$$

the latter showing that there is almost no restriction for small values of ε, even if we use central differences coupled with forward Euler time discretization.

High order extensions of this approach are possible, by using high order IMEX (for details see [5]). However, if we want the scheme to be accurate also in the cases in which ε is not too small, then we need to add two main ingredients:

- no term should be added when not needed, i.e. for large enough values of ε, since in such cases the additional terms degrade the accuracy; this can be achieved by letting $\mu(\varepsilon)$ decrease as ε increases. A possible choice which is adopted in all numerical tests is, for example
$$\mu(\varepsilon) = \exp(-\varepsilon^2/\Delta x);$$
this choice guarantees $\mu(0) = 1$, $\mu(1) = \exp(-1/\Delta x) \approx 0$ for small Δx (we assume the equations are written in non dimensional form, so that Δx is a small pure number).
- when the stabilizing effect of the dissipation vanishes, i.e. as $\mu \to 0$, then central differencing is no longer suitable, and one should adopt some upwinding; this can be obtained for example by blending central and upwind differencing as
$$D_x = (1-\nu)D_x^{\text{upw}} + \nu D_x^{\text{cen}}.$$
with $\nu = \nu(\varepsilon)$ and $\nu(0) = 1$. A possible choice for $\nu(\varepsilon)$ is $\nu = \mu$, but other functions may be adopted.

The idea of blending between upwind and central difference has already been proposed in [13]. In such paper the authors use a blending function which in our notation becomes
$$\mu_{JL} = \frac{1}{1 + 2\varepsilon^2/\Delta x}$$
which is the [0/1]-Padé approximant of $\exp(-2\varepsilon^2/\Delta x)$.

In the IMEX-I approach, applying MOL, the diffusive system (7) can be written as a ODEs system of the form
$$\begin{aligned} u' &= f_1(u,v) + f_2(u), \\ \varepsilon^2 v' &= g(u,v). \end{aligned}$$
where $f_1(u,v)$ represents the discretization of the term $-\partial_x(v + \mu\partial_x p(u))$, $f_2(u)$ represents the discretization of $\mu\partial_{xx}p(u)$ and $g(u,v)$ the discretization of the term $(-\partial_x p(u) - v + q(u))$.

When $\varepsilon \to 0$ the solution is projected onto the manifold $\mathcal{M} = \{(u,v) \in \mathbb{R} | g(u,v) = 0\}$. If we assume that the equation $g(u,v) = 0$ can always be solved for v, and denote $v = G(u)$ the solution, then the differential variable u satisfies
$$u' = \hat{f}_1(u) + f_2(u),$$
with $\hat{f}_1(u) = f(u, G(u))$. The previous system is called the *reduced system.*

It would be desirable that the IMEX scheme projects the numerical solution onto the manifold $\mathcal{M}$ as $\varepsilon \to 0$. In paper [5] we proved that a sufficient condition for an IMEX scheme to project the solution onto the manifold $\mathcal{M}$ is that the scheme is *globally stiffly accurate.*

An implicit RK scheme is said *stiffly accurate* if the last row of the matrix A is equal to the weights b^T. This ensures that the last stage is equal to the numerical solution. This guarantees nice stability properties of the scheme for very stiff equations (for example it ensures that the absolute stability function vanishes at infinity).

In [5, 7] we extended the definition of stiff accuracy to IMEX schemes, and say that an IMEX scheme is *globally stiffly accurate* if the last row of both explicit and implicit RK schemes that define the IMEX are equal to the corresponding weights, i.e. $e_s^T A = b^T$, $e_s^T \tilde{A} = \tilde{b}^T$ with $e_s^T = (0, ..., 0, 1)$.

Usually the numerical solution (u_n, v_n) for all n when $\varepsilon \to 0$ will not lie on the manifold $g(u, v) = 0$ since the quantity $g(u_n, v_n)$ is not necessarily zero. The IMEX-I approach with a globally stuffy accurate scheme guarantees that in the limit $\varepsilon \to 0$ we obtain a globally stiffly accurate implicit scheme and therefore $g(u_n, v_n) = 0$ for all n.

Finally in [5] we derived additional order conditions, called algebraic conditions, that guarantee the correct behavior of the numerical solution in the limit ε maintaining the classical accuracy in time of the scheme. We obtained such algebraic order conditions using the classical technique by comparing the Taylor expansion in time of the numerical solution with the one of the exact solution. More details about this approach, as well as some rigorous analysis can be found in [5].

3.2. **Additive Approach.** In the previous approach there may be a subtle difficulty when it comes to applications, namely it is not clear how to identify the hyperbolic part of the system, i.e. what is the term that should be included in the numerical flux if I want to use my favorite shock capturing FV or FD scheme? We proposed in [7] an alternative approach, in which we treat the whole hyperbolic part explicitly. For practical applications, it would be very nice to treat the whole term containing the flux explicitly, while reserving the implicit treatment only to the source, according the scheme:

$$\begin{array}{rcl} \begin{array}{rcl} u_t & = \\ v_t & = \end{array} & \underset{\text{[Explicit]}}{\boxed{\begin{array}{c} -v_x \\ -u_x/\varepsilon^2 \end{array}}} & \underset{\text{[Implicit]}}{\boxed{\begin{array}{c} \\ - \quad v/\varepsilon^2 \end{array}}} \qquad \text{(Additive)} \end{array}$$

We call such an approach *additive* and the corresponding schemes are denoted IMEX-E, to emphasize that the hyperbolic part is treated explicitly.

Such schemes should be easier to apply, because the fluxes retain their original interpretation. However, the approach seems hopeless, because of the diverging speeds.

Similarly as for the IMEX-I approach, we proposed for this approach, in order to overcome the parabolic restriction $\Delta t \approx \Delta x^2$ the same penalization technique based on adding two opposite terms to the first equation, and treating one explicitly and one implicitly.

In this paper, the authors concentrated on developing IMEX R-K schemes of type A, since they are easier to analyze with respect to the other types. They started the analysis by introducing a property which is important in order to guarantee the asymptotic preserving property, i.e. the scheme possesses the correct zero-relaxation limit, in the sense that the numerical scheme applied to system (5) should be a consistent and stable scheme for the limit system (6) as the parameter ε approaches zero independently of the discretization parameters. IMEX R-K schemes that satisfy this property are globally stiffly accurate schemes. Several results and a rigorous analysis about that can be found in [7]. Most numerical tests are reported in [7] for IMEX-E approach and the results are compared with those obtained by other methods available in the literature.

4. **Applications.**

This section is devoted to the presentation of some applications of the previous two approaches for the treatment of hyperbolic systems with diffusive relaxation.

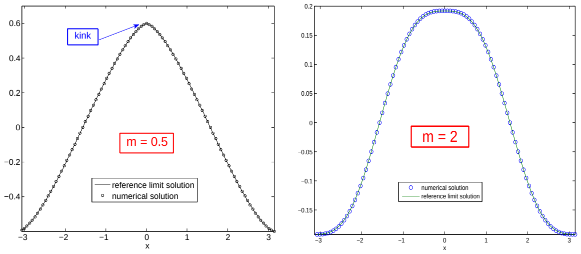


FIGURE 3. Numerical solution with $N = 96$ cells. Solid line: reference limit solution with $N = 384$ cells at time $T = 1$. IMEX-E approach, $\varepsilon = 10^{-4}$ and $\Delta t = C\Delta x^2$. $u(x, 0) = \cos(x)$, $v(x, 0) = \sin(x)$. Left $C = 1$. Right $C = 0.025$.

4.1. **Kawashima-LeFloch's nonlinear relaxation model.** Fully nonlinear relaxation terms arise, for instance in presence of nonlinear friction and, in this section we want numerically study the following non-linear relaxation model, first introduced by Kawashima and LeFloch [15], i.e.

$$\begin{aligned} u_t + v_x &= 0, \\ \varepsilon^2\, v_t + b(u)_x &= -|v|^{m-1}\, v + q(u). \end{aligned} \tag{8}$$

Provided $b'(u) > 0$, system (8) is strictly hyperbolic system of balance laws. In the stiff relaxation ε, ($\varepsilon \to 0$) we have

$$\begin{aligned} u_t &= (| - q(u) + b(u)_x|^{\alpha}(-q(u) + b(u)_x))_x, \\ |v|^{m-1}\, v &= q(u) - b(u)_x, \end{aligned}$$

which is a fully nonlinear parabolic equation in u with $\alpha = -1+1/m$. We distinguish between

$$\begin{aligned} \mathrm{sub-linear} &: 0 < m < 1, \\ \mathrm{linear} &: m = 1, \\ \mathrm{super-linear} &: m > 1. \end{aligned}$$

In its simplest form we assume $b(u) = u$, $q(u) = 0$ and we get:

$$\begin{aligned} u_t + v_x &= 0, \\ \varepsilon^2\, v_t + u_x &= -|v|^{m-1}\, v. \end{aligned}$$

As $\varepsilon \to 0$ this relaxes to

$$u_t = (|u_x|^{\alpha} u_x)_x\,, \qquad |v|^{m-1}\, v = -u_x. \tag{9}$$

Very interesting cases are both $m < 1\,(\alpha > 0)$ and $m \geq 1\,(\alpha \leq 0)$, The profile of the solution computed with $N = 96$ points is reported in Fig. 3. But by integrating for a longer time, the nonlinear parabolic equation (8) has regular solutions if $m > 1$, i.e. $\alpha \leq 0$, while it develops singularities in the derivatives if $0 < m < 1$, i.e. $\alpha > 0$. In fact, for $m = 2$ ($\alpha = -1/2$) integrating for a longer time $T = 1.77$, some instabilities appear (see Figure 4). The reason of such instabilities is that equation (9):

$$u_t = ((\alpha + 1)|u_x|^{\alpha})\, u_{xx},$$

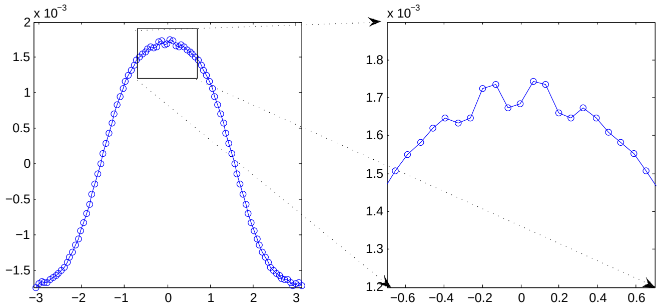


FIGURE 4. Instabilities for $m = 2$ ($\alpha = -1/2$), $\Delta t = C\Delta x^2$, $\varepsilon = 10^{-4}$ $N = 96$.

where the non-linear diffusion coefficient ν is

$$\nu = (1+\alpha)|u_x|^\alpha$$

which suggests the following condition in the nonlinear case

$$(1+\alpha)|u_x|^\alpha \frac{\Delta t}{\Delta x^2} \leq 1 \tag{10}$$

but the equation (9) diverges near local extrema when $\alpha < 0$ ($m > 1$). This condition (10) is used to determine the optimal time step for $m \leq 1$, no time step can guarantee stability near local extrema if $m > 1$. In [4], the same penalization technique proposed in [5, 7] in order to remove the parabolic stability restriction has been used. Then we write the system in the form

$$\begin{aligned} u_t &= -(v + \mu(\varepsilon)|u_x|^\alpha u_x)_x + \mu(\varepsilon)(|u_x|^\alpha u_x)_x \\ \varepsilon^2 v_t &= -u_x - |v|^{m-1} v. \end{aligned}$$

Now in order to treat this system by the IMEX-I or IMEX-E approach this requires that the term $(|u_x|^\alpha u_x)_x$ is treated implicitly. But some difficulty arises, in fact, when $\varepsilon \to 0$, the limit equation is non-linear parabolic and fully implicit would be very expensive.

In [4] a new approach has been used in order to solve the term $(|u_x|^\alpha u_x)_x$ where a very efficient method for the numerical solution of such an equation has been introduced. Indeed the idea is to write the equation as a system as

$$y' = F(y^*, y) \tag{11}$$

with F function non-stiff in the first variable and stiff in the second one. To be more specific, in our case $F(y^*, y)$ is given by $y = \left(\begin{array}{c} u \\ v \end{array} \right)$, $y^* = \left(\begin{array}{c} u^* \\ v^* \end{array} \right)$, and

$$F(y^*, y) = \left(\begin{array}{l} -(v_x^* + \mu(\varepsilon)(|u_x^*|^\alpha u_x^*)_x) + \mu(\varepsilon)(|u_x^*|^\alpha u_x)_x \\ -u_x + |v|^{m-1} v \end{array} \right).$$

Additive RK for this class of problems can be constructed, in particular we showed that in order to compute the numerical solution we need to require that $b_i = \tilde{b}_i$ for i (see [4]), then a good choice is to consider IMEX-I approach, whereas IMEX-E approach requires that the Runge-Kutta IMEX is globally stiffly accurate, i.e. $\tilde{b}_i \neq b_i$ for all i [7]. Using this new approach one can solve the relaxation system without parabolic CFL, i.e. $\Delta t = 0.25\Delta x$ and $T = 1.77$, Fig. 5.

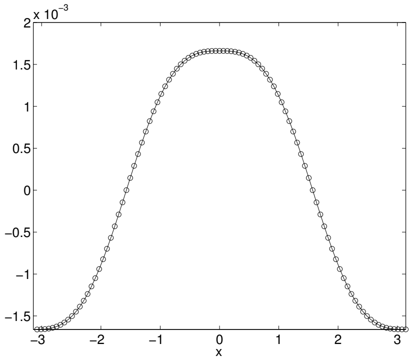


FIGURE 5. Numerical solution with $N = 96$ cells at time $T = 1.77$ for $m = 2$, $\varepsilon = 10^{-4}$ and $\Delta t = C\Delta x$, with $C = 0.25$.

- Time step is about 150 times larger than in the explicit method.
- The case $m = 2$, i.e. $\alpha = -1/2$, we set a TOL for computing $(|u|_x + TOL)^\alpha$, in order to avoid that the derivatives goes to infinity.

4.2. **R13: a regularized Grad's 13 moment method.** Grad's moment method is a technique used to close the infinite hierarchy of moments arising from the Boltzmann equation or rarefied gases. It is an example of hyperbolic relaxation: the Boltzmann equation relaxes to the hyperbolic system of Grad's equations. Sometimes parabolic systems provide more accurate physical description (e.g. Navier-Stokes equations are very successful in practice, although they are not hyperbolic). Some researchers, mainly Manuel Torrilhon and Henning Struchtrup [24] developed a parabolic extension of Grad's approach, called R13. When derived from the Boltzmann equation, this can be viewed as a parabolic relaxation.

In this section we present some results for the asymptotic accuracy for boundary value problems, which emerges from a 1D simplification of the R13 system that describes a Poiseuille-flow [17]. The system takes the form

$$U_\tau + F(U)_\xi = -\frac{1}{\varepsilon}P(U) + G \tag{12}$$

Here, the variables are $U = (u, v, w)^T$ with velocity u, shear stress v and parallel heat w. Furthermore, we have

$$F(U) = AU, \quad A = \begin{pmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix}, \quad P(U) = \begin{pmatrix} 0 \\ v \\ w \end{pmatrix}, \quad G = \begin{pmatrix} g \\ 0 \\ 0 \end{pmatrix} \tag{13}$$

where here the parameters g and ε are the external force and the relaxation time. Explicitly, we write system (12) as

$$\begin{aligned} u_\tau + v_\xi &= g, \\ v_\tau + \frac{1}{2}(u+w)_\xi &= -\frac{v}{\varepsilon}, \\ w_\tau + v_\xi &= -\frac{w}{\varepsilon}. \end{aligned} \tag{14}$$

We consider a bounded domain $\xi \in [-1, 1]$ where we have to prescribe boundary conditions. In [17], the authors used the following boundary conditions for v:

$$v|_{\xi=\pm 1} = \pm(\alpha u + \beta w)_{\xi=\pm 1}, \tag{15}$$

with $\alpha > \beta > 0$ some parameters. In the numerical experiments we chose the following values, $g = 1$, $\alpha = 0.7$, $\beta = 0.3$. A steady state solution for system (12) is given by

$$u_s(\xi) = g\left(\frac{1+\varepsilon\beta}{\alpha} + \frac{1}{\varepsilon}(1-\xi^2)\right), \quad v_s(\xi) = g\xi, \quad w_s(\xi) = -\varepsilon g. \tag{16}$$

We consider numerical tests whose solution converges to such steady state. We note that as we use high order reconstruction for the fuxes, then we need two layers of ghost cells that can be obtained using the boundary values. This part of the discretization is most important, because the efficiency of the whole method heavily depends on the choice of the correct boundary values and extrapolation methods.

Now we will focus our attention to the following system

$$\begin{aligned} \tilde{u}_t + v_x &= g, \\ v_t + \frac{1}{2}(\frac{\tilde{u}}{\varepsilon^2} + \tilde{w})_x &= -\frac{v}{\varepsilon^2}, \\ w_t + \frac{v_x}{\varepsilon^2} &= -\frac{\tilde{w}}{\varepsilon^2}. \end{aligned} \tag{17}$$

obtained by (14) under the diffusive scaling $t = \varepsilon\tau$, $x = \xi$, $\tilde{u} = \varepsilon u$ and $\tilde{w} = w/\varepsilon$.

Concerning the space discretization we consider a finite volume discretization as done in [17]. In our diffusive approach the matrix A in (13) has the following expression

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1/2\varepsilon^2 & 0 & 1/2 \\ 0 & 1/\varepsilon^2 & 0 \end{pmatrix} \tag{18}$$

In the small relaxation (or diffusion) limit, i.e. when $\varepsilon \to 0$, the behaviour of the solution to (17) is governed by

$$\tilde{w} = -v_x, \qquad v = \frac{-\tilde{u}_x}{2}, \tag{19}$$

and

$$\tilde{u}_t = \frac{\tilde{u}_{xx}}{2} + g. \tag{20}$$

Now consider boundary conditions for (17) which are consistent to the limit system (19,20).

4.2.1. *Boundary Treatment.* In this section we derive boundary conditions which are in agreement with the stationary solution.

From the steady state condition of Eq. (17) we get

$$\begin{aligned} v_x &= g, \\ \frac{\tilde{u}_x/\varepsilon^2 + \tilde{w}_x}{2} &= -\frac{v}{\varepsilon^2}. \\ v_x &= -\tilde{w} \end{aligned}$$

We observe that compatibility with stationary solutions implies:

$$\begin{aligned} \tilde{w}|_{\pm 1} &= -g, && (21)\\ v_x|_{\pm 1} &= g, && (22)\\ \left(\tilde{u}_x + \varepsilon^2 \tilde{w}_x\right)|_{\pm 1} &= -2v|_{\pm 1}. && (23) \end{aligned}$$

Such conditions are compatible with condition

$$\tilde{u}|_{\pm 1} = \pm \frac{\epsilon v \mp \beta \tilde{w} \epsilon^2}{\alpha}. \tag{24}$$

for the stationary solution (16). Therefore one can solve the system with boundary conditions (21), (22), (23) or (21), (22) and (24). In both cases one obtains convergence to the stationary solution.

System (12) is discretized by second order finite volume for the internal points. Ghost points are used out of the boundary to impose boundary conditions. Such ghost points are computed by extrapolation. For instance for the calculation of the boundary values considering (21), (22) and (24), we can write

$$\begin{aligned} \tilde{w}^W_{1/2} &= -g, && (25)\\ \tilde{w}_0 &= (8w^W_{1/2} - 6\tilde{w}_1 + \tilde{w}_2)/3, && (26)\\ v_0 &= v_1 - g\Delta x, && (27)\\ v^W_{1/2} &= \frac{3}{8}v_0 + \frac{3}{4}v_1 - \frac{1}{8}v_2, && (28)\\ \tilde{u}^W_{1/2} &= -\varepsilon(v^W_{1/2} + \varepsilon\beta\tilde{w}^W_{1/2})/\alpha, && (29)\\ u_0 &= (8\tilde{u}^W_{1/2} - 6\tilde{u}_1 + \tilde{u}_2)/3, && (30)\\ U_{-1} &= 3U_0 - 3U_1 + U_2, && (31) \end{aligned}$$

where $U = (\tilde{u}, v, \tilde{w})$. We do the same for the other part of the wall $x_{N+1/2}$.

We remark that we can improve the order of the extrapolation to the ghost cells by the following considerations. We consider the Lagrange polynomial

$$L_n(x; U) = \sum_{i=0}^{n} U_i \ell_i(x)$$

where $U = (\tilde{u}, v, \tilde{w})$ and

1. $\tilde{w}_0 \ell_0(x_{1/2}) = -g - \sum_{i=1}^{n} \tilde{w}_i \ell_i(x_{1/2})$,

2. $v_0 \ell_0^{'}(x_{1/2}) = g - \sum_{i=1}^{n} v_i \ell_i^{'}(x_{1/2})$,

3. $\tilde{u}_0 \ell_0(x_{1/2}) = -\frac{\varepsilon}{\alpha}(v(x_{1/2}) + \beta\tilde{w}(x_{1/2})\varepsilon) - \sum_{i=1}^{n} \tilde{u}_i \ell_i(x_{1/2})$.

and we can compute

$$U_k = \sum_{i=0}^{n} U_i \ell(x_k), \quad k = -1, -2, ...$$

Similarly for the other side of the wall.

We remark that we tested this approach performing also a numerical simulation setting different initial conditions, i.e. introducing a little perturbations to the initial data, and we observed that after a short time the numerical solution converge to the stationary solution.

4.2.2. *Removing parabolic stiffness.* We rewrite system (17) in the following form

$$
\begin{aligned}
\tilde{u}_t &= -v_x - \underbrace{\mu(\varepsilon)\frac{\tilde{u}_{xx}}{2} + \mu(\varepsilon)\frac{\tilde{u}_{xx}}{2}} + g,\\
v_t &= -\frac{\tilde{w}_x}{2} - \frac{1}{2}\frac{\tilde{u}_x}{\varepsilon^2} - \frac{v}{\varepsilon^2},\\
\tilde{w}_t &= -\frac{v_x}{\varepsilon^2} - \frac{\tilde{w}}{\varepsilon^2}.
\end{aligned}
\tag{32}
$$

where we added and subtracted the term $\mu(\varepsilon)\tilde{u}_{xx}/2$ in order to overcome the stability restriction that usually we have for hyperbolic system with diffusive relaxation. Here $\mu(\varepsilon)$ is such that $\mu : \mathbb{R}^+ \to [0,\ 1]$ and $\mu(0) = 1$. When ε is not small there is no reason to add and subtract the term $\mu(\varepsilon)u_{xx}$, therefore $\mu(\varepsilon)$ will be small in such a regime, i.e. $\mu(\varepsilon) \approx 0$. For a detailed analysis on this topic we report to [5]. Furthermore this reformulation allows us to design a class of IMEX Runge-Kutta schemes that work with high order accuracy in time in the zero-diffusion limit, i.e. when ε is very small, and in a wide range of the parameter ε such that the scheme maintains the accuracy uniformly for each value of ε. Now we want to apply an IMEX Runge-Kutta scheme with these features to this system considering IMEX-I approach, [5]. In our numerical test we consider the stiffly accurate IMEX-SSP2(3,3,2) which satisfies all the conditions described above.

Then, by (32), we treat the quantities

$$
(-v_x - \mu(\varepsilon)\frac{u_{xx}}{2}, -\frac{\tilde{w}_x}{2}, 0)^T \tag{33}
$$

explicitly and

$$
(\mu(\varepsilon)\frac{u_{xx}}{2} + g, -\frac{1}{2}\frac{\tilde{u}}{\varepsilon^2} - \frac{v}{\varepsilon^2}, -\frac{\tilde{v}_x}{\varepsilon^2} - \frac{\tilde{w}}{\varepsilon^2})^T \tag{34}
$$

implicitly, respectively.

4.2.3. *Convergence Results.* In order to ensure the second order convergence for the IMEX-SSP(3,3,2) scheme with the previous boundary conditions proposed, we simulate the same periodic test case proposed in [17]. We chose $g = 0$ and the initial conditions are $u = \sin(\pi x) + 0.5\sin(5\pi x)$, $v = 0$, $w = 0$. We simulate until $t_{end} = \varepsilon\tau_{end}$ with $\tau_{end} = 4$, $\varepsilon = 0.01$.

Numerical convergence rate is calculated by the formula

$$
p = \log_3(E_{\Delta t_1}/E_{\Delta t_2}), \tag{35}
$$

where $E_{\Delta t_1}$ and $E_{\Delta t_2}$ are the global errors associated to time steps Δt_1 and Δt_2, respectively. $E_{\Delta t_1}$ is obtained by comparing a solution with $N = 50$ with a solution obtained using $N = 150$ points, while for $E_{\Delta t_2}$ we use two solutions obtained, respectively, with $N = 150$ and $N = 450$ points. The number of points is tripled each time, because in this way it is easier to compare solutions in the same location using finite volume discretization. In Table 1 we show that a second order is reached for IMEX-SSP(3,3,2) scheme for all three components.

We note that we have obtained these convergence results considering the system (17) without adding and subtracting any term. It is clear from the previous considerations that a time step $\Delta t = \mathcal{O}(\Delta x^2)$ must be chosen. It is possible to obtain similar results considering the reformulated system (32) and choosing a time step $\Delta t = \mathcal{O}(\Delta x)$, although a special care has to be taken when imposing boundary conditions in the implicit step.

N	$Error_u$	$Error_v$	$Error_w$
50	$-$	$-$	$-$
150	$8.062e-04$	$2.530e-03$	$1.089e-02$
450	$7.838e-05$	$2.879e-04$	$1.162e-03$
Order	2.121	1.978	2.036

TABLE 1.

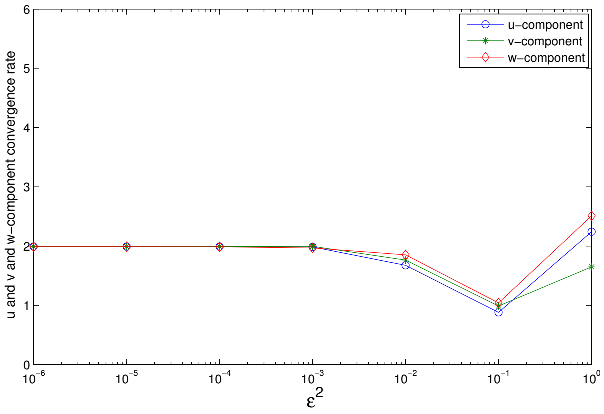


FIGURE 6. Convergence rate for the u, v, and w component versus ϵ^2

We now investigate numerically the convergence rate for a wide range of ε considering system (32) and choosing a time step $\Delta t = \mathcal{O}(\Delta x)$. To this aim we consider the previous test problem with the second order IMEX-SSP(3,3,2) scheme introduced before. Numerical convergence rate is calculated by (35) and time step $\Delta t = 0.3\Delta x$. We simulate until $t_{end} = 1$.

Figure 6 shows the convergence rates as a function of ε^2 using different values of ε ranging from 10^{-6} to 1. The second order scheme tested has the prescribed order of accuracy uniformly in ε^2 until ε is small. Instead, for values of ε large, say 10^{-1}, a degradation of accuracy is observed. This phenomenon requires further investigation as mentioned in [5].

4.2.4. *Convergence to the steady state solutions.* In this numerical test we show how starting form arbitrary initial conditions and considering the stiffly accurate IMEX-SSP2(3,3,2), the IMEX-I approach proposed in section (4.2.2), (see for details [5]), provides a numerical solution that converges to the steady state solution (16) in a number of time steps much smaller than the one needed by classical IMEX methods.

We consider $g = 1$, $\alpha = 0.7$, $\beta = 0.3$ and we choose $\varepsilon = 10^{-4}$ (*diffusive regime*). The final time is $\tau = 10$, the domain is $I = \{x : x \in [-1, 1]\}$ and $\Delta t_H = 2.5\Delta x$ with $N = 50$ grid points. This CFL number has been empirically adjusted to approximate the largest one that maintains stability. As initial data we consider

$$u_0 = \frac{\varepsilon}{\alpha}\left((C + \beta\varepsilon)x - g\right) \quad v_0 = gx + C \quad w_0 = -x^2. \tag{36}$$

This initial conditions are compatible with the boundary conditions (21), (22), (24). We plot the numerical solution at different final times 0.5, 1, 1.5, 3 and 10. At the final time the numerical solution is in perfect agreement with the steady state solution (16) after 100 time steps. We remark that the steady state solution is in practice reached a smaller time, say $t = 5$. We chose a long time in order to show that the numerical solution reaches the steady state with no oscillations. IMEX-I approach with the penalization technique described in Sec. 4.2.2 allows a time step Δt with a hyperbolic stability restriction rather than the parabolic one typical of explicit schemes for diffusion problems. Indeed, if we compute the numerical solutions u, v and w of system (17) without adopting the penalization technique, when ε is very small a stability parabolic restriction like $\Delta t_P = CFL\Delta x^2$ is required because the IMEX R-K method becomes an explicit one in the limit case $\varepsilon \to 0$. In this case we consider $CFL = 2.5$ and we note that thanks to the better stability properties of the new approach, the time step Δt_H is about 25 times bigger then Δt_P.

5. **Conclusions.**

We gave a brief review of modern IMEX Runge-Kutta schemes for hyperbolic systems in presence of stiff relaxation. Both hyperbolic and parabolic relaxations are considered, in the framework of conservative finite difference space discretization, which is the simplest approach to construct high order shock capturing schemes for such problems.

In the hyperbolic relaxation case, most IMEX schemes in the literature are able to capture the correct relaxed limit, converging to explicit schemes for the relaxed system. If high accuracy is required for a wide range of values of the relaxation parameter, then suitable conditions have to be imposed on the coefficients of the scheme in order to guarantee uniform accuracy, based on the analysis developed in [2].

The parabolic case is more subtle, since the characteristic speeds of the hyperbolic part diverge as the stiffness parameter vanishes. Numerical schemes commonly found in the literature for this family of problems converge to an explicit scheme for the limit parabolic equation, thus requiring a parabolic type CFL restriction on the time step. Recently developed schemes overcome such problem, using a penalization technique, and providing IMEX schemes that relax to an implicit scheme for the limit diffusion equation, of to an IMEX scheme for the limit convection-diffusion equation (according to the form of the relaxation term) [5, 7]. Suitable modification of such schemes can be adapted to problems that relax to genuinely non-linear diffusion equations [4]. IMEX schemes with the penalization techniques are applied here to a model problem coming from Extended Thermodynamics, providing a much more efficient tool to solve the problem with a number of time steps considerably smaller than the one required by other schemes present in the literature.

Several open problems remain. In particular we mention two problems that may attract the attention of researchers in this area. The first one is the extension of the uniform accuracy analysis performed in the case of hyperbolic relaxation to the more difficult problem of the parabolic relaxation. The second problem consists in exploiting the stabilization effect of the penalization technique adopted to improve the stability properties of the IMEX schemes for the parabolic relaxation to more a more general framework, extending the work already performed in [23] and [12] in specific cases.

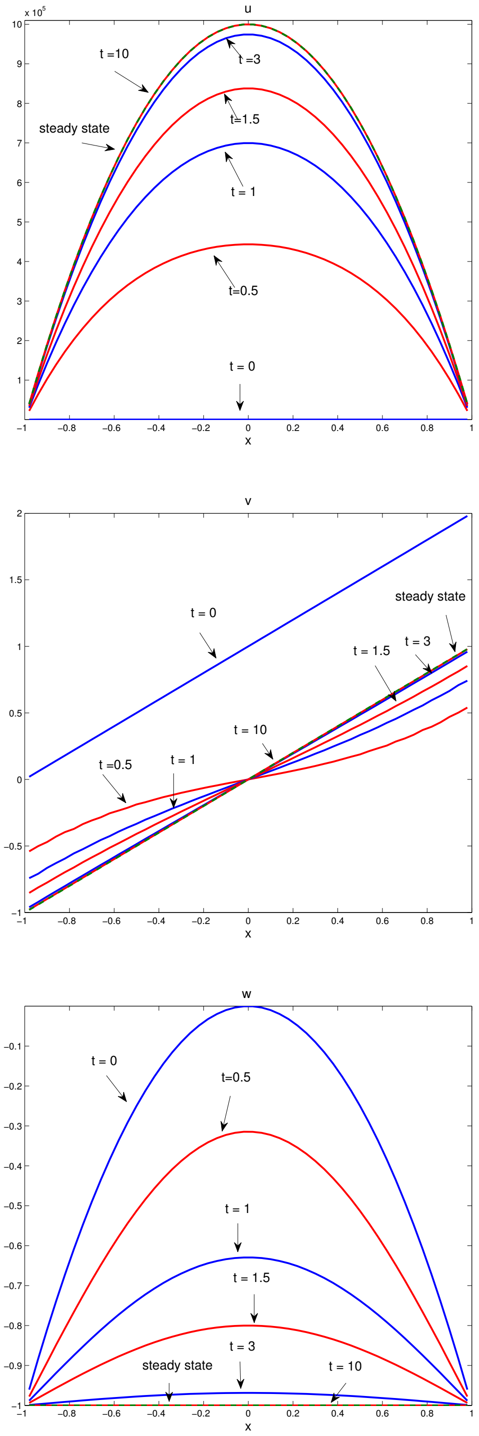


Figure 7. Convergence to the steady state for the R13 model problem. From top to bottom: u, v, and w profiles at different times. Number of grid points $N = 50$. Time step $\Delta t_H = 2.5\Delta x$.

Acknowledgments. The research was partially supported by Italian Ministry for University and Research *PRIN 2009: Innovative numerical methods for hyperbolic systems, with application to fluid dynamics, kinetic theory, and computational biology.*

REFERENCES

[1] U. Ascher, S. Ruuth, and R.J. Spiteri, *Implicit-explicit Runge-Kutta methods for timedependent partial differential equations*, Appl. Numer. Math., 25 (1997), pp. 151?67.

[2] S. Boscarino *Error analysis of IMEX Runge-Kutta methods derived from differential-algebraic systems*, SIAM J. Numer. Anal. Vol. 45, No. 4, pp. 1600-1621

[3] S. Boscarino *On an accurate third order implicit-explicit Rungeutta method for stiff problems*, Applied Numerical Mathematics 59 (2009) 1515?528.

[4] S. Boscarino, P. G. LeFloch and G. Russo, *High-order asymptotic-preserving methods for fully nonlinear relaxation problems*, submitted to SIAM J. on Sci. Comput. Preprint: arxiv.org/pdf/1210.4761.

[5] S. Boscarino, L. Pareschi and G. Russo, *Implicit-Explicit Runge-Kutta schemes for hyperbolic systems and kinetic equations in the diffusion limit*, forthcoming publication on SIAM J. Sci. Comput., preprint, http://arxiv.org/abs/1110.4375v2.

[6] S. Boscarino, G. Russo *On a class of uniformly accurate IMEX Runge-Kutta schemes and applications to hyperbolic systems with relaxation*, SIAM J. Sci. Comput., Vol. 31. No **3**, (2009), 1926–1945.

[7] S. Boscarino, G. Russo *Flux-Explicit IMEX Runge-Kutta schemes for hyperbolic to parabolic relaxation problems*, forthcoming publication on SIAM J. Numer. Anal.

[8] R. E. Caflisch, S. Jin, and G. Russo, *Uniformly accurate schemes for hyperbolic systems with relaxation*, SIAM J. Numer. Anal., 34 (1997), pp. 246?81.

[9] M. H. Carpenter and C. A. Kennedy, Additive Runge-Kutta schemes for convectiondiffusion-reaction equations, Appl. Numer. Math., 44 (2003), pp. 139?81.

[10] C. Q. Chen, C. D. Levermore, and T. P. Liu, *Hyperbolic conservation laws with relaxation terms and entropy*, Comm. Pure Appl. Math., 47 (1994), pp. 787?30.

[11] E. Hairer and G. Wanner, Solving Ordinary Differential Equation II: Stiff and Differential Algebraic Problems, 2nd ed., Springer Ser. Comput. Math. 14, Springer-Verlag, New York, 1991, 1996.

[12] F. Filbet and S. Jin, *A class of asymptotic-preserving schemes for kinetic equations and related problems with stiff sources* Journal of Computational Physics Volume: 229 Issue: 20, pp. 7625-7648

[13] S. Jin and D. Levermore, *Numerical schemes for hyperbolic conservation laws with source stiff Relaxation Terms*, J. Comp. Phys., 1996.

[14] S. Jin, L. Pareschi and G. Toscani *Diffusive relaxation for multiscale Discrete-Velocity Kinetic Equations.* SIAM. J. Num. Anal. Vol. 35, No. 6 (1998), pp. 2405-2439.

[15] S. Kawashima and P.G. LeFloch, in preparation.

[16] A. Klar *An asymptotic-induced scheme for nonstationary transport equations in the diffusive limit*, SIAM J Numer. Anal. Vol. 35, No. 3, pp. 1073-1094 (1998).

[17] J. Kollermeier, M. Torrilhon *Asymptotic Accuracy for Boundary Value Problems. Hyperbolic Gas Poiseuille Flow Model.* private communication.

[18] M. Lemou, L. Mieussens, *A new asymptotic preserving scheme based on micro-macro dormulation for linear kinetic equations in the diffusion limit*, SIAM Journal on Scientific Computing archive Volume 31 Issue 1, October 2008 Pages 334-368 .

[19] T. P. Liu, *Hyperbolic conservation laws with relaxation*, Comm. Math. Phys., 108 (1987) pp. 153?75.

[20] G. Naldi, L. Pareschi *Numerical Schemes for hyperbolic systems of conservation laws with stiff Diffusive relaxation.* SIAM. J. Num. Anal. Vol. 37, No. 4 (2000), pp. 1246-1270.

[21] L. Pareschi and G. Russo, *Implicit-explicit Runge-Kutta schemes and applications to hyperbolic systems with relaxations*, J. Sci. Comput., 25 (2005), pp. 129?55.

[22] C. W. Shu, *Essentially Non Oscillatory and Weighted Essentially Non Oscillatory Schemes for Hyperbolic Conservation Laws, in Advance Numerical Approximation of Nonlinear Hyperbolic Equations*, Lecture Notes in Math. 1697, (2000).

[23] P. Smereka, *Semi-implicit level set methods for curvature and surface diffusion motion* JOURNAL OF SCIENTIFIC COMPUTING Volume: 19 Issue: 1-3, pp. 439-456

[24] M. Torrilhon and H. Struchtrup *Regularized 13-moment equations: shock structure calculations and comparison to Burnett model*, J. Fluid. Mech. 2004, Vol. 513, pp. 171–198.

[25] G. B. Whitham, *Linear and non-linear waves*, Wiley, New York, 1974.

E-mail address: `boscarino@dmi.unict.it`
E-mail address: `russo@dmi.unict.it`

NEW INTERFACE METHODS FOR TRACKING MULTIPHASE MULTIPHYSICS

Robert Saye and James Sethian

Department of Mathematics, University of California
Berkeley and Lawrence Berkeley National Laboratory
Berkeley, California 94720, USA

Abstract. We discuss the theory and application of new methods to track moving interfaces in multiphase settings. These problems are characterized by multiple regions connected together, and moving under complex physics. We review work on a new mathematical and algorithmic methodology, known as the "Voronoi Implicit Interface Method", to track such problems.

1. **Introduction.** Propagating interfaces appear in wide variety of contexts, including fluid mixing, combustion dynamics, image segmentation, robotic navigation and path planning, and medical image analysis. One class of techniques to model and approximate such problems come from initial value partial differential equations, and results from embedding the evolving interface as the zero level set of a higher-dimensional function defined in an Eulerian setting, and whose resulting equations of motion are approximated using techniques borrowed from hyperbolic conservation laws. These techniques include level set methods [5], and their efficient adaptive implementation known as Narrow Band Level Set Methods [1]. They depend in part on the theory of curve and surface evolution, see [8, 9, 10].

These techniques are built for propagating interface problems in which there are two phases, which precludes the existence of complex structures such as triple points where three regions meet. These problems occur in a host of more complex interface problems, including grain metal boundaries, biological cell evolution and industrial foams. Considerable work has been aimed at moving these techniques to these more complex settings. These approaches typically involve multiple level set functions, essentially scaling the computational work with the number of phases, and relying on reconstruction techniques to ameliorate difficulties inherent in treating each region as less than fully coupled to its neighbors.

Recently, a new mathematical and computational methodology has been introduced, known as the "Voronoi Implicit Interface Method" [6, 7]. This method has a variety of features, including:

- *Accuracy, consistency, efficiency*: The method works in any number of dimensions, using a fixed Eulerian mesh, and a single function plus an indicator function to track the entire multiphase system. Geometric quantities and constraints are accurately computed, and phases are coupled together in a consistent fashion, with no gaps, overlaps, or ambiguities.

2000 *Mathematics Subject Classification.* Primary: 65M06.

Key words and phrases. Interface Methods, Voronoi Implicit Interface Methods.

- *Multiple junctions and topological change*: Multiple junctions, such as triple points, are all handled naturally and automatically, as well as breakage, merger, creation, and disappearance of phases. No special attention is paid to discontinuous topological change.
- *Coupling with time-dependent physics*: The method uses a physical time step, which then allows coupling complex physics into the interface evolution. Feedback from the physics affects the interface, and changes to the interface affects the physics.

In this review, we briefly discuss the development and application of these methods.

2. **Level Set Methods for Two-Phase Physics.** Level Set Methods, introduced by Osher and Sethian ("Fronts Propagating with Curvature-Dependent Speeds" [5]), were devised to accurately track interfaces evolving under a variety of complex speed laws in two and three dimensions. They rely in part on the theory of curve and surface evolution given in [9] and on the link between front propagation and hyperbolic conservation laws discussed in [10]. They recast interface motion as a time-dependent Eulerian initial value partial differential equation, and rely on viscosity solutions to the appropriate differential equations to update the position of the front, using an interface velocity that is derived from the relevant physics both on and off the interface. These viscosity solutions are obtained by exploiting schemes from the numerical solution of hyperbolic conservation laws. Level set methods are specifically designed for problems involving topological change, dependence on curvature, formation of singularities, and host of other issues that often appear in interface propagation techniques.

Briefly, let $\Gamma(t)$ be a moving interface, and let F be the speed normal to the interface, derived from solving the equations describing the appropriate physics. Typically, this involves a PDE on either side of the interface, augmented by jump conditions provided by the interface location and geometry, as well as solution values on the interface itself. One first constructs the initial value for the level set function $\phi(x, t = 0)$, obtained by evaluating the signed distance function at each point in the computational domain to the initial front position given by $\Gamma(t = 0)$. This extends the initial front Γ and constructs a function $\phi(x, t = 0)$ defined everywhere, such that the zero level set of this function corresponds to the initial position of the front.

By a suitable construction (see [2]), one may also extend the speed function F to the entire computational domain to obtain the extension velocity F_{ext} with the property that F_{ext} is constant along lines orthogonal to the level sets of ϕ, that is,

$$\nabla\phi \cdot \nabla F_{ext} = 0.$$

All that remains is to derive an evolution equation for the level set function ϕ under the extension velocity F_{ext} such that the evolving zero level set always matches the evolving interface $\Gamma(t)$. An application of the chain rule results in the initial value PDE

$$\phi_t + F_{ext}|\nabla\phi| = 0.$$

The entire level set interface evolution approach stems from approximating the solution of the above two PDEs.

2.1. **Sample Application of Two-Phase Physics.** As an application, we show the results from a numerical simulation of the ejection process associated with ink jet printing and two-phase microjetting in manufacturing and industrial devices. On a large scale, we built a fully three-dimensional computational model simulating both Newtonian and Oldroyd-B viscoelastic fluids in two-phase immiscible incompressible flows with surface tension, together with viscosity and density jumps across interfaces separating viscoelastic fluid from air[12, 13, 14]. This model includes a contact model for air/wall/fluid interactions, and incorporates complex geometries.

The coupled algorithm uses a projection method to enforce fluid incompressibility, a level set method to implicitly capture the moving interface, high order Godunov schemes for convection terms in the momentum and level set equations, a first-order upwind algorithm for convective viscoelastic stress and higher order central schemes for viscosity, surface tension, and upper-convected derivatives, and a slipping line contact model at air/wall/fluid triple points. The algorithm and software works on an arbitrary logically rectangular 3D mesh, and adaptive mesh refinement is employed to resolve fine scale features.

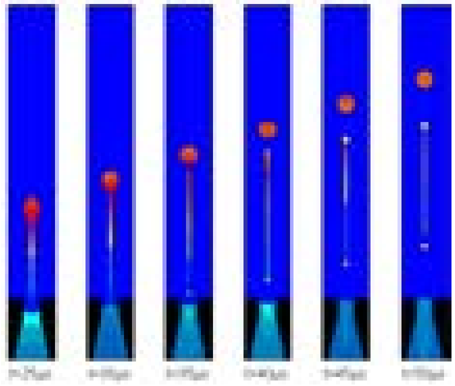


Figure 1. Axisymmetric inkjet ejection (see [12, 13, 14])

For smaller scales, we built a detailed combination level set/boundary element method to study micro-breakup in fluid drops[4]. To go past droplet breakup required a new mathematical model; briefly, one starts with a potential formulation on the evolving interface, and embed this potential function as an evolving PDE on a background fixed Eulerian mesh. Together with the embedded interface and extension velocity, this allows calculation of fluid breakup beyond singularity formation. Results show a remarkably close comparison with experiments, as well as quantitative match of scaling exponents for velocity and pinch-off times.

2.2. **Multiple Phases.** The above level set methodology works well for two phases. However, the situation becomes considerably more challenging when considering three phases or more. As illustration, imagine an interface separating three phases, as in the figure below.

Suppose we consider the seemingly straightforward problem of flow under curvature. It seems clear, at least from a mathematical point of view, that the phase with label "C" should pull more, since the angle formed by region C at the triple point is much smaller, and hence the curvature, as seen from C, is higher. Thus, as C retracts, its sides should open up, until an equilibrium is reached between all three phases, resulting in a stationary configuration with 120 degrees for each angle. However, this is a difficult flow to analyze mathematically, since the curvature is zero everywhere except at the triple junction itself, where it is not defined. While

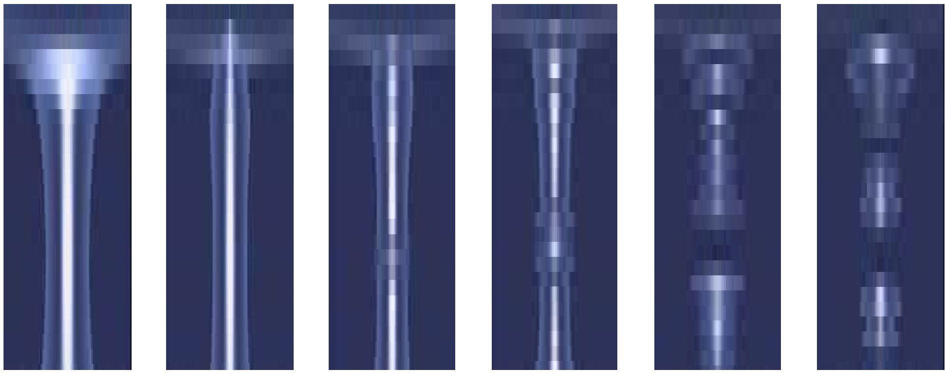


FIGURE 2. Microscale droplet breakup (see [4])

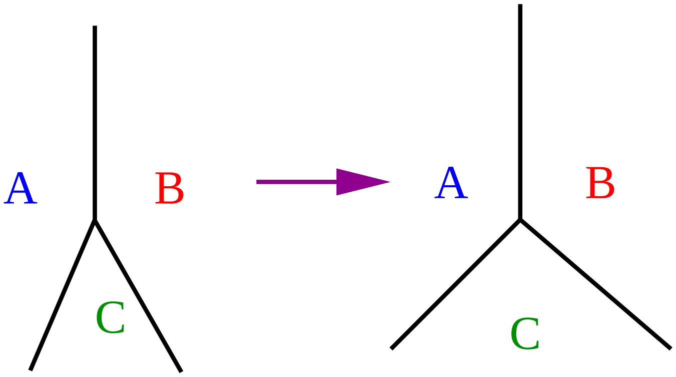


FIGURE 3. Evolution of a triple junction

it is possible in this case to pose a reasonable mathematical theory by, for example, choosing a viable mathematical solution based on minimizing area in the special case of this curvature flow, such analysis cannot be easily extended to the time-dependent complex physics where local geometric properties interact with physics away from the interface.

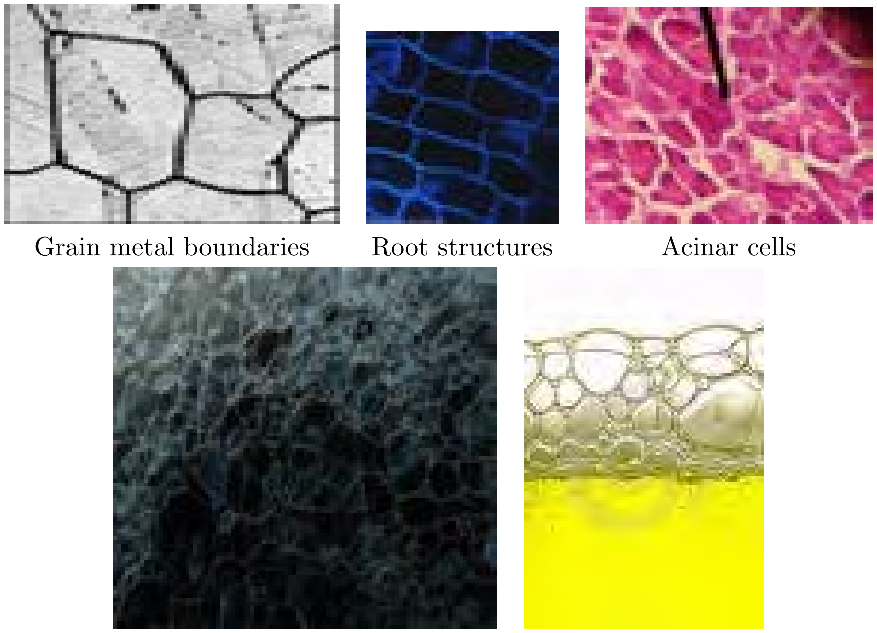
From a computational point of view, the problem is equally challenging. The lack of differentiability at the interface junction, the large range of possibilities for multiphase contact points, and the complications that come from solving physics in all of the domain, while needing accurate information on the interface itself, all pose special challenges. These issues become even more challenging in three dimensions.

3. Beyond Two Phases: The Voronoi Implicit Interface Method.

Imagine a collection of phases which share common boundaries: examples are shown in the figure below.

The basic idea of the Voronoi Implicit Interface Method, also known as the VIIM, is to combine three ideas.

- We consider the original interface Γ, which represents any boundary between two or more neighboring phases. Using this boundary, we then build the



Grain metal boundaries

Root structures

Acinar cells

Soap film

Foamy fluids

FIGURE 4. Examples of multiphase problems

unsigned distance function to each point in the domain. We additionally tag each point with an integer indicator flag corresponding to the phase.

- We now note that the ϵ level sets, $\epsilon > 0$, exist in one and only one phase.
- We also note that the *Voronoi Interface*, defined as the set of all points equidistant from at least two ϵ-level sets, returns, to a close approximation, the original interface Γ. As ϵ goes to zero, the limit of these Voronoi Interface is the original interface Γ.

These three ideas allow us to build a mathematical framework and computational methodology for following multiphase interfaces using a single evolving function on an Eulerian mesh. With time step Δt, mesh size h, and choice of $\epsilon > 0$, the steps are as follows:

1. Build the unsigned distance ϕ from the interface at each mesh point given on a Cartesian mesh with mesh size h.
2. Solve the appropriate physics to obtain a speed function, and then extend the solution to obtain the speed function F_{ext} at each mesh point.
3. Compute one step of level set evolution, namely by solving for one time step Δt the initial value PDE
$$\phi_t + F_{ext}|\nabla\phi| = 0.$$
4. Reconstruct the interface by finding the Voronoi Interface from the ϵ level set using the above solution.
5. Loop to the top.

This is the most straightforward implementation of the method. More efficient and sophisticated techniques include the use of narrow banding [2] to limit computational labor to a small region near the interface, a fast Eikonal solver [11, 3] to

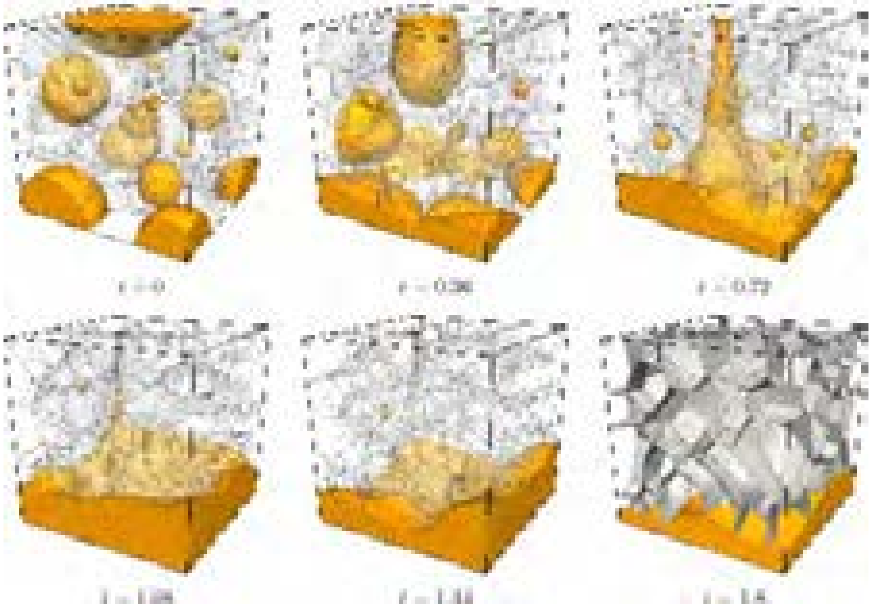


FIGURE 5. Results of a fluid flow simulation in three dimensions with gravity, in which the orange colored phase is more viscous and more dense than the other phases. The bulk foam is rendered mostly transparent except for the last frame, where it is rendered opaque to make the structure more prominent.

find the new unsigned distance from the ϵ level sets without explicitly constructing the front, and careful data structures which allow any non-negative value for ϵ, including $\epsilon = 0^+$. For details, see [6, 7].

4. **An Example.** As application, Figure 5 illustrates the results for a three-dimensional simulation of a variable density fluid flow, computed on a 128^3 grid with slip boundary conditions, using $\epsilon = 0^+$. The simulation starts with 15 heavy phases and approximately 100 less dense phases. The incompressible Navier-Stokes equations are solved, using a second order projection method, and coupled to the Voronoi Implicit Interface Method. For all but the last snapshot in Figure 5, the heavier phase is dark, while the other phases have been rendered mostly transparent, together with the triple line junctions as a network of curves. In the last snapshot, at time $t = 1.8$, we have rendered the bulk foam opaque, to make the structure of the foam more obvious.

Acknowledgments. This research was supported in part by the Applied Mathematical Sciences subprogram of the Office of Energy Research, US Department of Energy, under contract number DE-AC02-05CH11231, by the Division of Mathematical Sciences of the National Science Foundation, and by National Cancer Institute U54CA143833. Some computations used the resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. J.A.S. was also supported by the Miller Foundation at University of California, Berkeley, and as an Einstein Visiting Fellow of the Einstein Foundation, Berlin. R.I.S. was also supported by an American Australian Association Sir Keith Murdoch Fellowship.

REFERENCES

[1] D. Adalsteinsson, and J.A. Sethian, *A Fast Level Set Method for Propagating Interfaces*, J. Comp. Phys., **118**, 1995, pp. 269–277.

[2] D. Adalsteinsson, and J.A. Sethian, *The Fast Construction of Extension Velocities in Level Set Methods*, J. Comp. Phys., **148**, 1999, pp. 2–22.

[3] D.L. Chopp, *Some Improvements of the Fast Marching Method*, SIAM Journal Scientific Computing, **23**, 2001.

[4] Garzon, M., L.J., Gray, and J.A. Sethian, *Numerical simulation of non-viscous liquid pinch off using a coupled level set boundary integral method*, Journal Computational Physics, **228**, 2009, pp. 6079-6106.

[5] S. Osher, and J.A. Sethian, *Fronts Propagating with Curvature-Dependent Speed: Algorithms based on Hamilton-Jacobi Formulations*, J. Comp. Phys., **79**, 1988, pp.12-49.

[6] R. Saye, and J.A. Sethian, *The Voronoi Implicit Interface Method for Computing Multiphase Physics*, Proceedings of the National Academy of Sciences, **108**, 2011, pp. 19498-19503.

[7] R. Saye, and J.A. Sethian, *Analysis and Applications of the Voronoi Implicit Interface Method*, Journal Computational Physics, **231**, 2012, pp. 6051-6085.

[8] Sethian, J.A., *An Analysis of Flame Propagation*, Ph.D. Dissertation, Dept. of Mathematics, University of California, Berkeley, CA, 1982.

[9] J.A. Sethian, *Curvature and the Evolution of Fronts*, Comm. in Math. Phys., **101**, 1985, pp. 487–499.

[10] J.A. Sethian, *Numerical Methods for Propagating Fronts*, in Variational Methods for Free Surface Interfaces, Eds. P. Concus and R. Finn, Springer-Verlag, NY, 1987.

[11] J.A. Sethian, "Level Set Methods and Fast Marching Methods", Cambridge Univ. Press, 1999.

[12] J.D Yu, S. Sakai, and Sethian, J.A., *A Coupled Level Set Projection Method Applied to Ink Jet Simulation*, Interfaces and Free Boundaries, **193**, 2003, pp. 275-305.

[13] J.D Yu, S. Sakai, and Sethian, J.A., Yu, J., Sakai, S., and Sethian, J.A., *A Coupled Quadrilateral Grid Level Set Projection Method Applied to Ink Jet Simulation*, J. Computational Physics, **206**, 2005, pp. 227-251.

[14] J.D Yu, S. Sakai, and Sethian, J.A., *Two-Phase Viscoelastic Jetting*, Journal of Computational Physics, **220**, 2007, pp. 568-585.

E-mail address: saye@math.berkeley.edu
E-mail address: sethian@math.berkeley.edu

OPTIMAL SENSOR LOCATION FOR WAVE AND SCHRÖDINGER EQUATIONS

Yannick Privat
IRMAR, ENS Cachan Bretagne, Univ. Rennes 1, CNRS, UEB
av. Robert Schuman
35170 Bruz, France

Emmanuel Trélat
Université Pierre et Marie Curie (Univ. Paris 6)
and Institut Universitaire de France
CNRS UMR 7598, Laboratoire Jacques-Louis Lions
F-75005, Paris, France

Enrique Zuazua
BCAM - Basque Center for Applied Mathematics
Mazarredo, 14 E-48009 Bilbao-Basque Country, Spain
and
Ikerbasque, Basque Foundation for Science
Alameda Urquijo 36-5, Plaza Bizkaia
48011, Bilbao-Basque Country, Spain

Abstract. This paper summarizes the research we have carried out recently on the problem of the optimal location of sensors and actuators for wave equations, which has been the object of the talk of the third author at the *Hyp2012 Conference* held in Padova (Italy). We also address the same issues for the Schrödinger equations and present some possible perspectives of future research.

We consider the multi-dimensional wave or Schrödinger equations in a bounded domain Ω, with usual boundary conditions (Dirichlet, Neumann or Robin). We investigate the problem of optimal sensor location, in other words, the problem of designing what is the best possible subdomain of a prescribed measure on which one can observe the solutions. We present two mathematical problems modeling this question. The first one, in which the initial data under consideration are fixed, leads to optimal sets whose complexity depends on the regularity of the initial data. In the second one, the optimal set is searched so as to be uniform with respect to all initial data, and leads to a criterium of spectral nature, the answer being intimately related to the concentration properties of the eigenfunctions of the Laplacian. Under quantum ergodicity assumptions on the domain Ω we compute the optimal value of this problem, and show that this optimal value can be interpreted as the best possible observability

2000 *Mathematics Subject Classification.* Primary: 35P20, 93B07; Secondary: 58J51, 49K20.

Key words and phrases. Wave equation, Schrödinger equation, observability inequality, optimal design, spectral decomposition, ergodic properties, quantum ergodicity.

The third author acknowledges the hospitality of support of the organizers of the *Hyp2012 Conference* held in Padova (Italy) in which the content of the present work was presented in a plenary talk. The first author was partially supported by the ANR project OPTIFORM. The third author was partially supported by Grant MTM2011-29306-C02-00, MICINN, Spain, ERC Advanced Grant FP7-246775 NUMERIWAVES, ESF Research Networking Programme OPTPDE and Grant PI2010-04 of the Basque Government.

constant of a corresponding time-asymptotic or randomized observability inequality. Although optimal sets do exist in some specific situations, we show that the existence of an optimal set cannot be expected in general. Finally, we study a spectral approximation of that problem and construct a maximizing sequence of sets.

1. **Introduction.**

1.1. **Preliminaries: the problem of optimal observation.**

Let $T>0$, $n\in\mathbb{N}^*$, and $\Omega\subset\mathbb{R}^n$ be a bounded open connected subset. In this article we consider both the homogeneous wave equation

$$\partial_{tt}y=\triangle y, \tag{1}$$

and the Schrödinger equation

$$i\partial_t y=\triangle y, \tag{2}$$

for almost all $(t,x)\in(0,T)\times\Omega$, with Dirichlet boundary conditions for the sake of simplicity (other conditions are considered at the end of the article).

For any measurable subset ω of Ω of positive Lebesgue measure, we consider in both cases the observable variable

$$z(t,x)=\chi_\omega(x)y(t,x), \tag{3}$$

where χ_ω denotes the characteristic function of ω.

In this article we investigate the question of knowing whether there exists a best possible subset ω in order to observe the equation (1) or (2). To make the problem more precise, throughout the article we fix a real number $L\in(0,1)$, and from now on we restrict our search to all measurable subsets $\omega\subset\Omega$ which are of Lebesgue measure $|\omega|=L|\Omega|$. This determines the volume fraction of sensors that one would like to place in the domain Ω, in the best possible way.

Let us next model and define what the wording "best possible way" can mean.

1.2. **Mathematical modeling of two optimal design problems.**

In this context there are several possible ways of defining a concept of domain optimization. Certainly, the first problem that can be raised is the following.

First problem: best observation domain for fixed initial data.

- **Wave equation (1):** *given fixed initial data* $(y^0,y^1)\in L^2(\Omega,\mathbb{C})\times H^{-1}(\Omega,\mathbb{C})$*, we investigate the problem of maximizing the functional*

$$G_T(\chi_\omega)=\int_0^T\int_\omega|y(t,x)|^2\,dx\,dt, \tag{4}$$

  over all possible measurable subsets ω *of* Ω *of Lebesgue measure* $|\omega|=L|\Omega|$*, where* $y\in C^0(0,T;L^2(\Omega,\mathbb{C}))\cap C^1(0,T;H^{-1}(\Omega,\mathbb{C}))$ *is the unique solution of (1) such that* $y(0,\cdot)=y^0(\cdot)$ *and* $\partial_t y(0,\cdot)=y^1(\cdot)$*.*
- **Schrödinger equation (2):** *given* $y^0\in L^2(\Omega,\mathbb{C})$*, we investigate the problem of maximizing the functional* G_T *defined by (4) over all possible measurable subsets* ω *of* Ω *of Lebesgue measure* $|\omega|=L|\Omega|$*, where* $y\in C^0(0,T;L^2(\Omega,\mathbb{C}))$ *is the unique solution of (2) such that* $y(0,\cdot)=y^0(\cdot)$*.*

This problem appears as a mathematical benchmark, and is the first problem that one can raise in order to give a sense to the notion of best observation. However, this problem is not well suited in view of practical applications since it depends on the initial conditions. In applications, obviously, the location of sensors should be independent on the initial data. This problem is however interesting from an analytical point of view. As we will see, solving this problem is easy and optimal sets are level sets of a given function, that depends on the solution under consideration in a very sensitive way.

Let us now come to the definition of a uniform optimal design problem, independent on the initial data. In view of defining such a problem, relevant for practical issues, let us first recall the notion of observability inequality.

The system (1)-(3) is said to be observable on ω in time T if and only if there exists $C_T^{(W)}(\chi_\omega) > 0$ such that

$$C_T^{(W)}(\chi_\omega)\|(y^0,y^1)\|^2_{L^2(\Omega,\mathbb{C})\times H^{-1}(\Omega,\mathbb{C})} \leq \int_0^T \int_\omega |y(t,x)|^2\, dxdt, \tag{5}$$

for all $(y^0,y^1) \in L^2(\Omega,\mathbb{C})\times H^{-1}(\Omega,\mathbb{C})$. This is the so-called *observability inequality.* It is well known that within the class of $\mathcal{C}^\infty$ domains Ω, this observability property holds if the pair (ω,T) satisfies the *Geometric Control Condition* in Ω (see [3]), according to which every ray of Geometric Optics that propagates in Ω and is reflected on its boundary $\partial\Omega$ intersects ω within time T.

Similarly, system (2)-(3) is said to be observable on ω in time T if and only if there exists $C_T^{(S)}(\chi_\omega) > 0$ such that

$$C_T^{(S)}(\chi_\omega)\|y^0\|^2_{L^2(\Omega,\mathbb{C})} \leq \int_0^T \int_\omega |y(t,x)|^2\, dxdt, \tag{6}$$

for every $y^0 \in L^2(\Omega,\mathbb{C})$. If there exists T^* such that the pair (ω,T^*) satisfies the *Geometric Control Condition* then the observability inequality (35) holds for every $T>0$ (see [18]). In some sense the Schrödinger equation can be viewed as a wave equation with an infinite speed of propagation.

In the sequel, the quantities $C_T^{(W)}(\chi_\omega)$ and $C_T^{(S)}(\chi_\omega)$ denote the largest possible nonnegative constants for which the inequalities (34) and (35) hold, that is,

$$C_T^{(W)}(\chi_\omega) = \inf_{\|(y^0,y^1)\|_{L^2\times H^{-1}}=1} \int_0^T \int_\omega |y(t,x)|^2\, dx\, dt, \tag{7}$$

and

$$C_T^{(S)}(\chi_\omega) = \inf_{\|y^0\|_{L^2}=1} \int_0^T \int_\omega |y(t,x)|^2\, dx\, dt. \tag{8}$$

They are called the *observability constants.*

These remarks being done, in view of defining a uniform optimal design problem for the observability of wave or Schrödinger equations, it is natural to raise the problem of maximizing the above observability constants over all possible subsets ω of Ω of Lebesgue measure $|\omega| = L|\Omega|$. However, this problem appears to be:

1. Very difficult to handle: indeed when considering spectral expansions of the solutions, difficulties arise due to crossed terms, as in the interesting open problem of determining the best constants in Ingham's inequalities (see [13, 14], see also [23] for such considerations in the one-dimensional case);
2. Finally, not so relevant. Indeed the above inequalities are *deterministic*, and hence, in some sense, the observability constants are pessimistic, since they give an account for the worst possible observability scenario. In practice one is led to handle a large number of solutions but not all of them, and the deterministic observability constant will rarely be reached. We are then going

to define a randomized version of the observability constant, which appears to be more relevant.

These two points appeal further comments.

Let us first present the Fourier expansion of solutions of the spectral basis of the Laplacian. Let $(\phi_j)_{j\in\mathbb{N}^*}$ be a Hilbertian basis of $L^2(\Omega)$ consisting of eigenfunctions[1] of the Dirichlet Laplacian operator on Ω, associated with the negative eigenvalues $(-\lambda_j^2)_{j\in\mathbb{N}^*}$. Then, for given initial data $(y^0, y^1) \in L^2(\Omega,\mathbb{C}) \times H^{-1}(\Omega,\mathbb{C})$, the corresponding solution of (1) is

$$y(t,x) = \sum_{j=1}^{+\infty} \left(a_j e^{i\lambda_j t} + b_j e^{-i\lambda_j t}\right) \phi_j(x), \tag{9}$$

where the sequences $(a_j)_{j\in\mathbb{N}^*}$ and $(b_j)_{j\in\mathbb{N}^*}$ belong to $\ell^2(\mathbb{C})$ and are defined by

$$\begin{aligned} a_j &= \frac{1}{2}\left(\int_\Omega y^0(x)\phi_j(x)\,dx - \frac{i}{\lambda_j}\int_\Omega y^1(x)\phi_j(x)\,dx\right), \\ b_j &= \frac{1}{2}\left(\int_\Omega y^0(x)\phi_j(x)\,dx + \frac{i}{\lambda_j}\int_\Omega y^1(x)\phi_j(x)\,dx\right). \end{aligned} \tag{10}$$

for every $j \in \mathbb{N}^*$. Moreover,

$$\|(y^0,y^1)\|^2_{L^2\times H^{-1}} = 2\sum_{j=1}^{+\infty}(|a_j|^2+|b_j|^2). \tag{11}$$

With such a spectral expansion, note that

$$G_T(\chi_\omega) = \sum_{j,k=1}^{+\infty} \alpha_{jk}\int_\omega \phi_i(x)\phi_j(x)\,dx, \tag{12}$$

where the coefficients α_{jk}, $(j,k) \in (\mathbb{N}^*)^2$ (which can be easily computed) depend only on the initial data (y^0,y^1) and the observation time T. It can be noted that, since ω is a proper subset of Ω, there holds in general $\int_\omega \phi_i(x)\phi_j(x)\,dx \neq 0$. Because of these crossed terms, the observability constant $C_T^{(W)}(\chi_\omega)$ defined by (7) can be interpreted as the infimum of eigenvalues of an infinite dimensional nonnegative symmetric matrix (called *Gramian*), which is far from diagonal due to nonzero nondiagonal terms.

The observability constant $C_T^{(W)}(\chi_\omega)$ could be easily expressed if the Gramian were to be a diagonal matrix. This is actually one of the nice consequences of the randomization procedure mentioned in the second point. Let us explain briefly this procedure (full details are provided in [26]). Following [6], we randomize some initial data determined by their Fourier coefficients (10), by defining $a_j^\nu = \beta_{1,j}^\nu a_j$ and $b_j^\nu = \beta_{2,j}^\nu b_j$, where $(\beta_{1,j}^\nu)_{j\in\mathbb{N}^*}$ and $(\beta_{2,j}^\nu)_{j\in\mathbb{N}^*}$ are two sequences of independent Bernoulli random variables on a probability space $(\mathcal{X},\mathcal{A},\mathbb{P})$, satisfying

$$\mathbb{P}(\beta_{1,j}^\nu = \pm 1) = \mathbb{P}(\beta_{2,j}^\nu = \pm 1) = \frac{1}{2} \qquad \text{and} \qquad \mathbb{E}(\beta_{1,j}^\nu \beta_{2,k}^\nu) = 0$$

for all j and k in $\mathbb{N}^*$ and every $\nu \in X$. Here, the notation $\mathbb{E}$ stands for the expectation over the space $\mathcal{X}$ with respect to the probability measure $\mathbb{P}$. Let y_ν

[1] Note that this Hilbertian basis is not necessarily unique in case of multiple eigenvalues. What follows depends a priori on the specific choice of the basis of eigenfunctions which is done at this step of our analysis.

denote the corresponding solution,

$$y_\nu(t,x) = \sum_{j=1}^{+\infty} \left(\beta_{1,j}^\nu a_j e^{i\lambda_j t} + \beta_{2,j}^\nu b_j e^{-i\lambda_j t} \right) \phi_j(x).$$

Then, instead of considering the deterministic observability inequality (34), we consider the randomized one

$$C_{T,\text{rand}}^{(W)}(\chi_\omega) \|(y^0, y^1)\|_{L^2\times H^{-1}}^2 \leq \mathbb{E}\left(\int_0^T \int_\omega |y_\nu(t,x)^2| \, dx \, dt \right), \tag{13}$$

for all $y^0(\cdot) \in L^2(\Omega, \mathbb{C})$ and $y^1(\cdot) \in H^{-1}(\Omega, \mathbb{C})$. Here, the constant $C_{T,\text{rand}}^{(W)}(\chi_\omega)$, called the *randomized observability constant* for the wave equation, is a new constant which is a priori different from its deterministic counterpart $C_T^{(W)}(\chi_\omega)$. A similar consideration is done for the Schrödinger equation, with a randomized observability constant $C_{T,\text{rand}}^{(S)}(\chi_\omega)$.

It is proved in [26] that, for every measurable subset ω of Ω, there holds

$$2\, C_{T,\text{rand}}^{(W)}(\chi_\omega) = C_{T,\text{rand}}^{(S)}(\chi_\omega) = T \inf_{j\in\mathbb{N}^*} \int_\omega \phi_j(x)^2 \, dx = T J(\chi_\omega). \tag{14}$$

In other words, the randomization procedure sketched permits to kill the crossed terms and hence, up to considering random initial data and an averaged version of the observability inequality, to provide a concept of randomized Gramian, which is a diagonal infinite dimensional matrix.

As mentioned above, the randomized observability inequality (13) appears to be more relevant than its classical deterministic version (34) in view of applications. The first problem in which the initial data are given and fixed is not very relevant. But in practice one does not need to consider all possible solutions either. The above randomization procedure, provides a reasonable mathematical modeling of this practical optimal design problem.

It follows from all above considerations that a way to define a relevant uniform optimal design problem is the following.

> **Second problem: uniform optimal design problem.** *We investigate the problem of maximizing the functional*
>
> $$J(\chi_\omega) = \inf_{j\in\mathbb{N}^*} \int_\omega \phi_j(x)^2 \, dx, \tag{15}$$
>
> *over all possible subsets ω of Ω of Lebesgue measure $|\omega| = L|\Omega|$.*

This problem consists of maximizing an eigenfunction energy concentration criterion. As we will see, solving this problem leads to highly interesting mathematical considerations related to quantum ergodicity properties of the domain Ω.

Remark 1. It is proved in [26] that, if the domain Ω is such that every eigenvalue of A is simple, then, similarly to (14), there holds

$$2\, C_\infty^{(W)}(\chi_\omega) = C_\infty^{(S)}(\chi_\omega) = \inf_{j\in\mathbb{N}^*} \int_\omega \phi_j(x)^2 \, dx = J(\chi_\omega), \tag{16}$$

for every measurable subset ω of Ω, where $C_\infty^{(W)}(\chi_\omega)$ and $C_\infty^{(S)}(\chi_\omega)$ are *time asymptotic observability constants*, defined respectively as the largest possible nonnegative constant for which the time asymptotic observability inequality

$$C_\infty^{(W)}(\chi_\omega)\|(y^0,y^1)\|_{L^2\times H^{-1}}^2 \leq \lim_{T\to+\infty}\frac{1}{T}\int_0^T\int_\omega |y(t,x)^2|\,dx\,dt, \tag{17}$$

holds for all $(y^0,y^1)\in L^2(\Omega,\mathbb{C})\times H^{-1}(\Omega,\mathbb{C})$, for the wave equation, and

$$C_\infty^{(S)}(\chi_\omega)\|y^0\|_{L^2}^2 \leq \lim_{T\to+\infty}\frac{1}{T}\int_0^T\int_\omega |y(t,x)^2|\,dx\,dt, \tag{18}$$

holds for every $y^0(\cdot)\in L^2(\Omega,\mathbb{C})$, for the Schrödinger equation.

1.3. **Some bibliographical comments.** The problem of optimal measurement locations for state estimation in linear partial differential equations has been widely considered in engineering problems (see e.g. [9, 15, 16, 21, 28, 29] and the many references therein), the aim being to optimize the number, the place and the type of sensors or actuators in order to improve the estimation or more generally some performance index. Fields of applications are very numerous and concern for example active structural acoustics, piezzoelectric issues, vibration control in mechanical structures, damage detection processes, chemical reactions, just to name a few of them. A usual approach popular in the engineering community consists of recasting the optimal sensor location problem for distributed systems as an optimal control problem with an infinite dimensional Riccati equation, having a statistical model interpretation, and then of computing approximations with optimization techniques. However, on the one part, their techniques rely on an exhaustive search over a predefined set of possible candidates and are faced with combinatorial difficulties due to the selection problem and thus with the usual flaws of combinatorial optimization methods. On the other part, in all these references approximations are used to determine the optimal sensor or actuator location. The optimal performance and the corresponding sensor or actuator location of the approximating sequence are then expected to converge to the exact optimal performance and location. Among the possible approximation processes, the closest one to our present study consists of considering Fourier expansion representations and using modal approximation schemes.

However, in these references there is no systematic mathematical study of the optimal design problem. The search of optimal domains relies on finite-dimensional approximations and no convergence analysis is led. However, in the present article we show that modal approximation procedures may fail and Γ-convergence properties may not hold when passing to the limit from a finite number of eigenfunction components to all of them.

Although the optimal design problems under consideration in this article have been widely studied in the engineering community, in particular because of their great importance in practical problems, there exist only few mathematical results. An important difficulty arising when focusing on an optimal shape problem is the generic non-existence of classical solutions, as explained and surveyed in [2], thus leading to consider relaxation procedures. In [4] the authors investigate the problem modeled in [27] of finding the best possible distributions of two materials (with different elastic Young modulus and different density) in a rod in order to minimize the vibration energy in the structure. For this optimal design problem in wave propagation, the authors of [4] prove existence results and provide relaxation and optimality conditions. The authors of [1] also propose a relaxation formulation of eigenfrequency optimization problems applied to optimal design. In [7] the authors discuss several possible criteria for optimizing the damping of abstract wave equations in Hilbert spaces, and derive optimality conditions for a certain criterion related to a Lyapunov equation. In [11, 12], the authors consider the problem of determining the best possible shape and position of the damping subdomain of given measure for a 1D wave equation. In [20, 22] the authors investigate numerically the optimal location of the support of the control for the 1-D wave equation. Their numerical methods are then mostly based on gradient techniques or level set methods combined with shape and topological derivatives (we refer the reader e.g. to [5] for a survey on variational methods in shape optimization problems). In [23] we investigated the second problem presented previously in the one-dimensional case, and in [24] we studied the related dual problem of finding the optimal location of the support of the control for the one-dimensional wave equation. In [25] we solved in a complete way the first problem (optimal observation domain for the problem with fixed initial data), and in [26] we solved the second problem (uniform with respect to initial data), emphasizing close connections with the quantum chaos theory, as explained further.

2. Statement of the main results.

2.1. First problem: best observation domain for fixed initial data.

Consider fixed initial data $(y^0, y^1) \in L^2(\Omega, \mathbb{C}) \times H^{-1}(\Omega, \mathbb{C})$ (resp., $y^0 \in L^2(\Omega, \mathbb{C})$) for the wave equation (1) (resp., for the Schrödinger equation (2)), and let y be their corresponding solution. We define the integrable function

$$\varphi(x) = \int_0^T |y(t,x)|^2 dt, \tag{19}$$

for every $x \in \Omega$. Note that $G_T(\chi_\omega) = \int_\omega \varphi(x)\, dx$ for every measurable subset $\omega \subset \Omega$.

Theorem 2.1. [25] *There exists at least one measurable subset ω of Ω, solution of the first problem, characterized as follows. There exists a real number λ such that every optimal set ω is contained in the level set $\{\varphi \geq \lambda\}$, where the function φ defined by (19) is integrable on Ω.*

Moreover, if there exists $R > 0$ such that

$$\sum_{j=0}^{+\infty} \frac{R^j}{j!} \left(\|A^j y^0\|_{L^2}^2 + \|A^{j-1} y^1\|_{L^2}^2 \right)^{1/2} < +\infty, \tag{20}$$

in the case of the wave equation, and

$$\sum_{j=0}^{+\infty} \frac{R^j}{j!} \|A^j y^0\|_{L^2} < +\infty, \tag{21}$$

in the case of the Schrödinger equation, where $A = \sqrt{-\triangle}$ (square root of the Dirichlet-Laplacian), then the first problem has a unique[2] solution χ_ω, where ω is a measurable subset of Ω of measure $L|\Omega|$, satisfying moreover the following properties:

- *there exists $\eta > 0$ such that $d(\omega, \partial\Omega) > \eta$, where d denotes the Euclidean distance on $\mathbb{R}^n$;*

[2] Similarly to the definition of elements of L^p-spaces, the subset ω is unique within the class of all measurable subsets of Ω quotiented by the set of all measurable subsets of Ω of zero measure.

- *ω is semi-analytic*[3]*, and has a finite number of connected components;*
- *if Ω is symmetric with respect to an hyperplane and $y^0 \circ \sigma = y^0$ and $y^1 \circ \sigma = y^1$, where σ denotes the symmetry operator with respect to this hyperplane, then ω enjoys the same symmetry property.*

Remark 2. The optimal set is not necessarily unique, whenever the function φ is constant on some subset of Ω of positive measure. We refer to [23, 25] for explicit examples.

Theorem 2.1 states that, if the initial data belong to some analyticity spaces, then the (unique) optimal set ω is the union of a finite number of connected components. Using a careful harmonic analysis construction, it is proved in [25] that there exist C^∞ initial data for which the optimal set ω may have a fractal structure and, more precisely, may be of Cantor type. More precisely, one has the following result.

Theorem 2.2. [25] *Let $\Omega = (0, 2\pi)$ and let $T > 0$ be an integer multiple of 4π. There exist C^∞ initial data (y^0, y^1) defined on Ω for which the first problem has a unique solution ω; moreover ω has a fractal structure and in particular it has an infinite number of connected components.*

2.2. **Uniform optimal design.** In this section, we focus on the second problem, defined as

$$\sup_{\chi_\omega \in \mathcal{U}_L} J(\chi_\omega), \tag{22}$$

with

$$J(\chi_\omega) = \inf_{j \in \mathbb{N}^*} \int_\omega \phi_j(x)^2 \, dx,$$

and

$$\mathcal{U}_L = \{\chi_\omega \mid \omega \text{ is a measurable subset of } \Omega \text{ of measure } |\omega| = L|\Omega|\}. \tag{23}$$

2.2.1. *Convexification.* To ensure compactness properties, we consider the convex closure of $\mathcal{U}_L$ for the weak star topology of L^∞,

$$\overline{\mathcal{U}}_L = \left\{ a \in L^\infty(\Omega, [0,1]) \mid \int_\Omega a(x)\, dx = L|\Omega| \right\}. \tag{24}$$

The convexified version of the second problem (22) is

$$\sup_{a \in \overline{\mathcal{U}}_L} J(a), \tag{25}$$

where

$$J(a) = \inf_{j \in \mathbb{N}^*} \int_\Omega a(x) \phi_j(x)^2 \, dx. \tag{26}$$

By upper semi-continuity of J for the weak star topology of L^∞, it is clear that the problem (25) has at least one solution. For instance in dimension one there is an infinite number of solutions, characterized through their Fourier coefficients (see

[3] A subset ω of a real analytic finite dimensional manifold M is said to be semi-analytic if it can be written in terms of equalities and inequalities of analytic functions, that is, for every $x \in \omega$, there exists a neighborhood U of x in M and $2pq$ analytic functions g_{ij}, h_{ij} (with $1 \leq i \leq p$ and $1 \leq j \leq q$) such that

$$\omega \cap U = \bigcup_{i=1}^{p} \{y \in U \mid g_{ij}(y) = 0 \text{ and } h_{ij}(y) > 0,\ j = 1, \ldots, q\}.$$

We recall that such semi-analytic (and more generally, subanalytic) subsets enjoy nice properties, for instance they are stratifiable in the sense of Whitney.

[23]). Note that taking $a(\cdot) = L$ yields $\sup_{a\in\overline{\mathcal{U}}_L} J(a) \geq L$, and note that a priori, $\sup_{\chi_\omega\in\mathcal{U}_L} J(\chi_\omega) \leq \sup_{a\in\overline{\mathcal{U}}_L} J(a)$. The question of knowing if this inequality is an equality or not (gap or no-gap) is not obvious, and cannot be treated using standard Γ-convergence arguments due to the lack of lower semi-continuity of J.

2.2.2. *Main results.* We make the following assumptions on the Hilbertian basis $(\phi_j^2)_{j\in\mathbb{N}^*}$ of eigenfunctions under consideration.

> ***Weak Quantum Ergodicity on the base (WQE) property.*** *There exists a subsequence of the sequence of probability measures $\mu_j = \phi_j^2\, dx$ converging vaguely to the uniform measure $\frac{1}{|\Omega|}\, dx$.*
>
> ***Uniform L^∞-boundedness property.*** *There exists $A > 0$ such that*
> $$\|\phi_j\|_{L^\infty(\Omega)} \leq A, \tag{27}$$
> *for every $j \in I\!N^*$.*

These assumptions above imply what we call the L^∞*-Weak Quantum Ergodicity on the base (L^∞-WQE) property*, that is, there exists a subsequence of $(\phi_j^2)_{j\in\mathbb{N}^*}$ converging to $\frac{1}{|\Omega|}$ for the weak star topology of $L^\infty(\Omega)$. This property obviously implies that

$$\sup_{a\in\overline{\mathcal{U}}_L} J(a) = \sup_{a\in\overline{\mathcal{U}}_L} \inf_{j\in\mathbb{N}^*} \int_\Omega a(x)\phi_j(x)^2\, dx = L, \tag{28}$$

and moreover the supremum is reached with the constant function $a = L$ on Ω.

Theorem 2.3. [26] *If the WQE and uniform L^∞-boundedness properties hold, then*

$$\sup_{\chi_\omega\in\mathcal{U}_L} \inf_{j\in I\!N^*} \int_\omega \phi_j(x)^2\, dx = L, \tag{29}$$

for every $L \in (0,1)$. In other words, under these assumptions there is no gap between the original problem (22) and the convexified one.

As a consequence, the maximal value of the randomized observability constants $2\, C_{T,\text{rand}}^{(W)}(\chi_\omega) = C_{T,\text{rand}}^{(S)}(\chi_\omega)$ over the set $\mathcal{U}_L$ is equal to TL. Moreover if the spectrum of A is simple then the maximal value of the time asymptotic observability constants $2C_\infty^{(W)}(\chi_\omega) = C_\infty^{(S)}(\chi_\omega)$ over the set $\mathcal{U}_L$ is equal to L.

We now define the set $\mathcal{U}_L^b = \{\chi_\omega \in \mathcal{U}_L \mid |\partial\omega| = 0\}$, and we make the following assumptions.

> ***Quantum Unique Ergodicity on the base (QUE) property.*** *The whole sequence of probability measures $\mu_j = \phi_j^2\, dx$ converges vaguely to the uniform measure $\frac{1}{|\Omega|}\, dx$.*
>
> ***Uniform L^p-boundedness property.*** *There exist $p \in (1, +\infty]$ and $A > 0$ such that*
> $$\|\phi_j\|_{L^{2p}(\Omega)} \leq A, \tag{30}$$
> *for every $j \in I\!N^*$.*

Theorem 2.4. [26] *If $\partial\Omega$ is Lipschitz and if the QUE and uniform L^p-boundedness properties hold, then*

$$\sup_{\chi_\omega\in\mathcal{U}_L^b} \inf_{j\in I\!N^*} \int_\omega \phi_j(x)^2\, dx = L, \tag{31}$$

for every $L \in (0,1)$.

Actually the statement of Theorem 2.4 holds true as well whenever the set $\mathcal{U}_L^b$ is replaced by the set of all measurable subsets ω of Ω, of measure $|\omega| = L|\Omega|$, that are moreover open either with Lipschitz boundary or bounded perimeter.

The ergodicity assumptions made above are sufficient but are not sharp. For instance it is proved in [26] that, if Ω is the unit disk of the Euclidean two-dimensional space, then, for every $p \in (1, +\infty]$ and for any basis of eigenfunctions, the uniform L^p-boundedness property is not satisfied, and QUE does not hold as well; however (29) and (31) hold true. And this, in spite of the phenomenon of whispering galleries, which gives an account for the existence of certain semi-classical measures (weak limits of the probability measures $\phi_j^2\,dx$) such as the Dirac measure along the boundary.

Remark 3. The assumptions made in the above theorems obviously hold in dimension one (Dirichlet-Laplacian on a bounded interval). In higher dimensions they are related to deep questions arising in mathematical physics (indeed, in quantum mechanics $\mu_j = \phi_j^2\,dx$ is the probability of being in the state ϕ_j), related to Shnirelman's Theorem. This celebrated result asserts that, if the domain Ω is a convex ergodic billiard with piecewise smooth boundary, then there exists a subsequence of the sequence of probability measures $\mu_j = \phi_j^2\,dx$ of density one converging vaguely to the uniform measure $\frac{1}{|\Omega|}dx$ (see [10, 31]). This property is referred to as *Quantum Ergodicity on the base* (in short, QE on the base). Actually the result is stronger and holds in the full phase space, for pseudo-differential operators (see [30] for a recent survey). Of course, QUE implies QE which in turn implies WQE.

Note that Shnirelman Theorem lets open the possibility of having an exceptional subsequence of μ_j converging vaguely to some measure different from the uniform one, for instance, to a measure carried by closed geodesics (concentration phenomenon known as *scar*, see e.g. [8]). The QUE assumption made above postulates that this scarring phenomenon does not occur. Up to now there is no example of a domain in dimension more than one in which QUE has been proved to hold, and this is a deep open question in this thematics. We refer the reader to [26] for a more detailed discussion on such quantum ergodicity issues in relation with shape optimization problems.

Remark 4. In general we do not expect the supremum in (29) or (31) to be reached. This is an open question. But it is reached in several very particular situations. This is the case for instance in dimension one for a very specific value of L: when $\Omega = [0, \pi]$, then the supremum of J over $\mathcal{U}_L$ (which is equal to L) is reached if and only if $L = 1/2$; in that case, it is reached for all measurable subsets $\omega \subset [0, \pi]$ of measure $\pi/2$ such that ω and its symmetric image $\omega' = \pi - \omega$ are disjoint and complementary in $[0, \pi]$ (see [26]).

3. Spectral approximation of the uniform optimal design problem.

Given the functional J defined by (15), in view of designing a spectral approximation it is natural to consider the truncated functional defined by

$$J_N(\chi_\omega) = \min_{1\leq j\leq N} \int_\omega \phi_j(x)^2\,dx, \tag{32}$$

for every $N \in \mathbb{N}^*$ and every measurable subset ω of Ω. The spectral approximation of the second problem (uniform optimal design problem) is then

$$\sup_{\chi_\omega\in\mathcal{U}_L} J_N(\chi_\omega). \tag{33}$$

Accordingly, J_N is extended to $\overline{\mathcal{U}}_L$ by $J_N(a) = \min_{1\leq j\leq N} \int_\Omega a(x)\phi_j(x)^2\,dx$ for every $a \in \overline{\mathcal{U}}_L$.

Theorem 3.1. [26]

1. *For every measurable subset ω of Ω, the sequence $(J_N(\chi_\omega))_{N\in\mathbb{N}^*}$ is non-increasing and converges to $J(\chi_\omega)$.*
2. *There holds*
$$\lim_{N\to+\infty} \max_{a\in\overline{\mathcal{U}}_L} J_N(a) = \max_{a\in\overline{\mathcal{U}}_L} J(a).$$
Moreover, if $(a^N)_{n\in\mathbb{N}^*}$ is a sequence of maximizers of J_N in $\overline{\mathcal{U}}_L$, then up to a subsequence, it converges to a maximizer of J in $\overline{\mathcal{U}}_L$ for the weak star topology of L^∞.
3. *For every $N \in \mathbb{N}^*$, the problem (33) has a unique solution χ_{ω^N}, where $\omega^N \in \mathcal{U}_L$. Moreover, ω^N is semi-analytic (see Footnote 3) and thus has a finite number of connected components.*

Remark 5. It is proved in [12, 23] that, in the one-dimensional case, the optimal set ω_N maximizing J_N is the union of N intervals concentrating around equidistant points and that ω_N is actually the worst possible subset for the problem of maximizing J_{N+1}. This is the so-called *spillover phenomenon* which is a serious drawback from the practical point of view since it makes it impossible the implementation of a spectral approximation procedure.

The next numerical simulations, based on the above spectral approximation, confirm this pathological behavior. Consider $\Omega = [0,\pi]^2$. The normalized eigenfunctions of the Dirichlet-Laplacian are $\phi_{j,k}(x_1,x_2) = \frac{2}{\pi}\sin(jx_1)\sin(kx_2)$, for every $(x_1,x_2) \in [0,\pi]^2$. Let $N \in \mathbb{N}^*$. We use an interior point line search filter method to solve the optimization problem $\sup_{\chi_\omega\in\mathcal{U}_L} J_N(\chi_\omega)$, with
$$J_N(\chi_\omega) = \min_{1\leq j,k\leq N} \int_0^\pi\int_0^\pi \chi_\omega(x_1,x_2)\phi_{j,k}(x_1,x_2)^2\,dx_1\,dx_2.$$
Some results are provided on Figure 1 in the Dirichlet case. They show very clearly that the number of connected components of the optimal set increases as N grows. We have thus constructed a maximizing sequence of sets for the second problem (uniform optimal design problem) which is evidently far from converging in any reasonable sense.

4. Further comments and perspectives.

4.1. Generalization to other boundary conditions.

Up to now we have restricted ourselves to Dirichlet boundary conditions. Actually, as shown in [26], our analysis can be developed in the more general framework where Ω is an open bounded connected subset of M, and (M,g) is a smooth n-dimensional Riemannian manifold, with $n \geq 1$. In that case, the Dirichlet-Laplacian is replaced with the Laplace-Beltrami operator $\triangle_g$ on M for the metric g. The boundary of Ω can be empty: in this case, Ω is a compact connected n-dimensional Riemannian manifold. If $\partial\Omega \neq \emptyset$ then we consider boundary conditions $By = 0$ on $(0,T)\times\partial\Omega$, where B can be either:

- the usual Dirichlet trace operator, $By = y_{|\partial\Omega}$,
- or Neumann, $By = \frac{\partial y}{\partial n}_{|\partial\Omega}$, where $\frac{\partial}{\partial n}$ is the outward normal derivative on $\partial\Omega$,

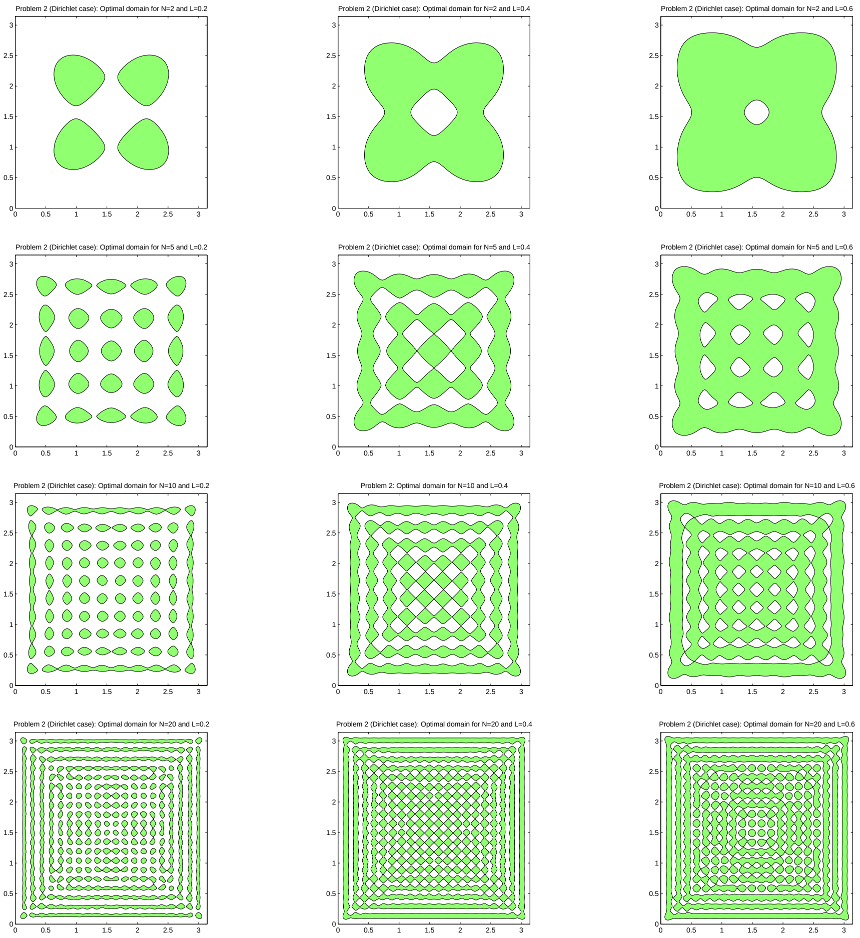


FIGURE 1. On this figure, $\Omega = [0,\pi]^2$. Line 1, from left to right: optimal domain (in green) in the Dirichlet case for $N = 2$ (4 eigenmodes) and $L \in \{0.2, 0.4, 0.6\}$. Line 2, from left to right: optimal domain (in green) for $N = 5$ (25 eigenmodes) and $L \in \{0.2, 0.4, 0.6\}$. Line 3, from left to right: optimal domain (in green) for $N = 10$ (100 eigenmodes) and $L \in \{0.2, 0.4, 0.6\}$. Line 4, from left to right: optimal domain (in green) for $N = 20$ (400 eigenmodes) and $L \in \{0.2, 0.4, 0.6\}$.

- or mixed Dirichlet-Neumann, $By = \chi_{\Gamma_0} y_{|\partial\Omega} + \chi_{\Gamma_1} \frac{\partial y}{\partial n}_{|\partial\Omega}$, where $\partial\Omega = \Gamma_0 \cup \Gamma_1$ with $\Gamma_0 \cap \Gamma_1 = \emptyset$, and χ_{Γ_i} is the characteristic function of Γ_i, $i = 0, 1$,
- or Robin, $By = \frac{\partial y}{\partial n}_{|\partial\Omega} + \beta y_{|\partial\Omega}$, where β is a nonnegative bounded measurable function defined on $\partial\Omega$, such that $\int_{\partial\Omega} \beta > 0$.

The Lebesgue measure dx must be replaced with the canonical measure dV_g induced by the canonical Riemannian volume V_g on M.

Also, to encompass all possible boundary conditions settled above, we replace the observability inequalities (34) and (35) with

$$C_T^{(W)}(\chi_\omega)\|(y^0,y^1)\|^2_{D(A^{1/2})\times X} \leq \int_0^T \int_\omega |\partial_t y(t,x)|^2 \, dV_g \, dt, \tag{34}$$

for all $(y^0,y^1) \in D(A^{1/2}) \times X$, and

$$C_T^{(S)}(\chi_\omega)\|y^0\|^2_{D(A)} \leq \int_0^T \int_\omega |\partial_t y(t,x)|^2 \, dV_g \, dt, \tag{35}$$

for every $y^0 \in D(A)$. Here, the following notations are used: $A = -\triangle_g$ is the Laplace operator defined on $D(A) = \{y \in X \mid Ay \in X \text{ and } By = 0\}$ with one of the above boundary conditions whenever $\partial\Omega \neq \emptyset$, and X is the space $L^2(\Omega,\mathbb{C})$ in the case of Dirichlet, mixed or Robin boundary conditions, and otherwise

$$X = L^2_0(\Omega,\mathbb{C}) = \{y \in L^2(\Omega,\mathbb{C}) \mid \int_\Omega y(x)\, dV_g = 0\}.$$

Defined in this space, the operator A is then selfadjoint and positive definite. In the case of Dirichlet boundary conditions, one has $D(A) = H^2(\Omega,\mathbb{C}) \cap H^1_0(\Omega,\mathbb{C})$ and $D(A^{1/2}) = H^1_0(\Omega,\mathbb{C})$. For Neumann boundary conditions, one has

$$D(A) = \{y \in H^2(\Omega,\mathbb{C}) \mid \frac{\partial y}{\partial n}_{|\partial\Omega} = 0 \text{ and } \int_\Omega y(x)\, dV_g = 0\}$$

and

$$D(A^{1/2}) = \{y \in H^1(\Omega,\mathbb{C}) \mid \int_\Omega y(x)\, dV_g = 0\}.$$

In the mixed Dirichlet-Neumann case (with $\Gamma_0 \neq \emptyset$), one has

$$D(A) = \{y \in H^2(\Omega,\mathbb{C}) \mid y_{|\Gamma_0} = \frac{\partial y}{\partial n}_{|\Gamma_1} = 0\},$$

and

$$D(A^{1/2}) = H^1_{\Gamma_0}(\Omega,\mathbb{C}) = \{y \in H^1(\Omega,\mathbb{C}) \mid y_{|\Gamma_0} = 0\}$$

(see e.g. [17]).

4.2. **An intrinsic variant of the uniform optimal design problem.**

As said before, the second problem (15) depends a priori on the orthonormal Hilbertian basis $(\phi_j)_{j\in\mathbb{N}^*}$ of $L^2(\Omega)$ which has been fixed at the beginning of the analysis, at least whenever the spectrum of A is not simple. If the eigenvalues $(\lambda_j^2)_{j\in\mathbb{N}^*}$ of A are multiple, then the choice of the basis $(\phi_j)_{j\in\mathbb{N}^*}$ is an issue. One possible way to get rid of this dependence is to consider the infimum of the criteria J defined by (15) over all possible choices of orthonormal bases of eigenfunctions. This leads to the following intrinsic variant of the second problem. We adopt the framework and the notations of the previous section.

Intrinsic uniform optimal design problem. *We investigate the problem of maximizing the functional*

$$J_{int}(\chi_\omega) = \inf_{\phi\in\mathcal{E}} \int_\omega \phi(x)^2 \, dV_g, \tag{36}$$

over all possible subsets ω of Ω of measure $V_g(\omega) = LV_g(\Omega)$, where $\mathcal{E}$ denotes the set of all normalized eigenfunctions of A.

Note that $C_T^{(W)}(\chi_\omega) \leq \frac{T}{2} J_{\text{int}}(\chi_\omega) \leq C_{T,\text{rand}}^{(W)}(\chi_\omega)$ and $C_T^{(S)}(\chi_\omega) \leq T J_{\text{int}}(\chi_\omega) \leq C_{T,\text{rand}}^{(S)}(\chi_\omega)$. As before, the functional J_{int} is extended to $\overline{\mathcal{U}}_L$ by setting $J_{\text{int}}(a) = \inf_{\phi\in\mathcal{E}} \int_\Omega a(x)\phi(x)^2\, dV_g$ for every $a \in \overline{\mathcal{U}}_L$. The following results are the intrinsic counterpart of Theorems 2.3 and 2.4.

Theorem 4.1. [26] *Assume that the uniform measure $\frac{1}{V_g(\Omega)}\, dV_g$ is a closure point of the family of probability measures $\mu_\phi = \phi^2\, dV_g$, $\phi \in \mathcal{E}$, for the vague topology, and that the whole family of eigenfunctions in $\mathcal{E}$ is uniformly bounded in $L^\infty(\Omega)$. Then*

$$\sup_{\chi_\omega\in\mathcal{U}_L} \inf_{\phi\in\mathcal{E}} \int_\omega \phi(x)^2\, dV_g = \sup_{a\in\overline{\mathcal{U}}_L} \inf_{\phi\in\mathcal{E}} \int_\Omega a(x)\phi(x)^2\, dV_g = L, \tag{37}$$

for every $L \in (0,1)$. In other words, there is no gap between the intrinsic uniform optimal design problem and its convexified version.

Theorem 4.2. [26] *Assume that the uniform measure $\frac{1}{V_g(\Omega)}\, dV_g$ is the unique closure point of the family of probability measures $\mu_\phi = \phi^2\, dV_g$, $\phi \in \mathcal{E}$, for the vague topology, and that the whole family of eigenfunctions in $\mathcal{E}$ is uniformly bounded in $L^{2p}(\Omega)$, for some $p \in (1,+\infty]$. Then*

$$\sup_{\chi_\omega\in\mathcal{U}_L^b} \inf_{\phi\in\mathcal{E}} \int_\omega \phi(x)^2\, dV_g = L, \tag{38}$$

for every $L \in (0,1)$.

Remark 6. We are able to provide examples where there is a gap between the intrinsic second problem (36) and its convexified version. This occurs in any of the two following examples (see [26]):

- $\Omega = S^2$, the unit sphere in $\mathbb{R}^3$, endowed with the usual flat metric;
- Ω is the unit half-sphere in $\mathbb{R}^3$, endowed with the usual flat metric, and Dirichlet conditions are imposed on the great circle (boundary of Ω).

In both cases, if L is close enough to 1 then $\sup_{\chi_\omega\in\mathcal{U}_L} J(\chi_\omega) < L$, and hence there is a gap between the problem (36) and its convexified version.

4.3. **Optimal location of internal controllers.** By duality, our previous results provide an answer to the question of determining the shape and location of the control domain for wave or Schrödinger equations that minimizes the L^2 norm of the controllers realizing null controllability. For simplicity we restrict ourselves to the internally controlled wave equation on Ω with Dirichlet boundary conditions,

$$\begin{array}{ll} \partial_{tt}y(t,x) - \triangle_g y(t,x) = h_\omega(t,x), & (t,x) \in (0,T)\times\Omega, \\ y(t,x) = 0, & (t,x) \in [0,T]\times\partial\Omega, \\ y(0,x) = y^0(x),\ \partial_t y(0,x) = y^1(x), & x \in \Omega, \end{array} \tag{39}$$

where h_ω is a control supported in $[0,T]\times\omega$ and ω is a measurable subset of Ω. Note that the Cauchy problem (39) is well posed for all initial data $(y^0,y^1) \in H_0^1(\Omega,\mathbb{C})\times L^2(\Omega,\mathbb{C})$ and every $h_\omega \in L^2((0,T)\times\Omega,\mathbb{C})$, and its solution y belongs to $C^0(0,T;H_0^1(\Omega,\mathbb{C}))\cap C^1(0,T;L^2(\Omega,\mathbb{C}))\cap C^2(0,T;H^{-1}(\Omega,\mathbb{C}))$. The exact null controllability problem settled in these spaces consists of finding a control h_ω steering the control system (39) to

$$y(T,\cdot) = \partial_t y(T,\cdot) = 0. \tag{40}$$

It is well known that, for every subset ω of Ω of positive measure, the exact null controllability problem is by duality equivalent to the fact that the observability inequality

$$C\|(\phi^0,\phi^1)\|^2_{L^2(\Omega,\mathbb{C})\times H^{-1}(\Omega,\mathbb{C})} \leq \int_0^T\int_\omega |\phi(t,x)|^2\,dV_g\,dt, \tag{41}$$

holds, for all $(\phi^0,\phi^1)\in L^2(\Omega,\mathbb{C})\times H^{-1}(\Omega,\mathbb{C})$, for a positive constant C (only depending on T and ω), where ϕ is the (unique) solution of the adjoint system

$$\begin{array}{ll} \partial_{tt}\phi(t,x)-\triangle_g\phi(t,x)=0, & (t,x)\in(0,T)\times\Omega,\\ \phi(t,x)=0, & (t,x)\in[0,T]\times\partial\Omega,\\ \phi(0,x)=\phi^0(x),\ \partial_t\phi(0,x)=\phi^1(x), & x\in\Omega. \end{array} \tag{42}$$

The Hilbert Uniqueness Method (HUM, see [19]) provides a way to design the unique control solving the control problem (39)-(40) and having moreover a minimal $L^2((0,T)\times\Omega,\mathbb{C})$ norm. This control is referred to as the HUM control and is characterized as follows. Define the HUM functional J_ω by

$$J_\omega(\phi^0,\phi^1)=\frac{1}{2}\int_0^T\int_\omega\phi(t,x)^2\,dV_g\,dt-\langle\phi^1,y^0\rangle_{H^{-1},H_0^1}+\langle\phi^0,y^1\rangle_{L^2}. \tag{43}$$

The notation $\langle\cdot,\cdot\rangle_{H^{-1},H_0^1}$ stands for the duality bracket between $H^{-1}(\Omega,\mathbb{C})$ and $H_0^1(\Omega,\mathbb{C})$, and the notation $\langle\cdot,\cdot\rangle_{L^2}$ stands for the usual scalar product of $L^2(\Omega,\mathbb{C})$. If (41) holds then the functional J_ω has a unique minimizer (still denoted (ϕ^0,ϕ^1)) in the space $L^2(\Omega,\mathbb{C})\times H^{-1}(\Omega,\mathbb{C})$, for all $(y^0,y^1)\in H_0^1(\Omega,\mathbb{C})\times L^2(\Omega,\mathbb{C})$. The HUM control h_ω steering (y^0,y^1) to $(0,0)$ in time T is then given by

$$h_\omega(t,x)=\chi_\omega(x)\phi(t,x), \tag{44}$$

for almost all $(t,x)\in(0,T)\times\Omega$, where ϕ is the solution of (42) with initial data (ϕ^0,ϕ^1) minimizing J_ω.

The HUM operator Γ_ω is defined by

$$\begin{array}{rrcl} \Gamma_\omega: & H_0^1(\Omega,\mathbb{C})\times L^2(\Omega,\mathbb{C}) & \longrightarrow & L^2((0,T)\times\Omega,\mathbb{C})\\ & (y^0,y^1) & \longmapsto & h_\omega \end{array}$$

Optimal design control problem. We investigate the problem of minimizing the norm of the operator Γ_ω,

$$\|\Gamma_\omega\|=\sup_{\|(y^0,y^1)\|_{H_0^1)\times L^2}=1}\|h_\omega\|_{L^2((0,T)\times\Omega,\mathbb{C})} \tag{45}$$

over the set $\mathcal{U}_L$.

Here, we formulate the optimal design control problem in terms of minimizing the operator norm of Γ_ω in order to discard the dependence with respect to the initial data (y^0,y^1) and improve the robustness of the cost function.

By a duality argument, it is proved in [26] that, for every measurable subset ω of Ω, if $C_T^{(W)}(\chi_\omega)>0$ then

$$\|\Gamma_\omega\|=\frac{1}{C_T^{(W)}(\chi_\omega)},$$

and if $C_T^{(W)}(\chi_\omega)=0$, then $\|\Gamma_\omega\|=+\infty$. It follows that, for the optimal design control problem,

$$\inf_{\chi_\omega\in\mathcal{U}_L}\|\Gamma_\omega\|=\left(\sup_{\chi_\omega\in\mathcal{U}_L}C_T^{(W)}(\chi_\omega)\right)^{-1},$$

and therefore the problem is equivalent to the problem of maximizing the observability constant. Then, all considerations before can be applied as well to the optimal design control problem.

4.4. **Conclusions and perspectives.** We have provided a mathematical rigorous modeling of the problem of optimizing the shape and placement of sensors over a domain in which one considers the wave or the Schrödinger equation, with Dirichlet, Neumann, mixed or Robin boundary conditions whenever the boundary is nonempty.

First, when a specific choice of the initial data is given and therefore we deal with a particular solution, we have shown that the problem always admits at least one solution that can be regular or of fractal type depending on the regularity of the initial data.

In view of practical applications, we have defined a uniform optimal design problem, which does not depend on the initial data. Through spectral decompositions, we have motivated a second problem which consists of maximizing a spectral functional that can be viewed as a measure of eigenfunction concentration. Roughly speaking, the subset ω has to be chosen so to maximize the minimal trace of the squares of all eigenfunctions. This spectral criterion can be obtained and interpreted in two ways: on the one hand, it corresponds to a time asymptotic observability constant as the observation time interval tends to infinity, and on the other hand, to a randomized version of the deterministic observability inequality. We have also considered the convexified formulation of the problem. Under appropriate quantum ergodicity assumptions on Ω, we have a no-gap result between the initial problem and its convexified version, and we have computed the optimal value.

We have then provided spectral approximations, permitting to construct a maximizing sequence, and presented some numerical simulations that show the increasing complexity of the optimal sets.

Overall, our results highlight precise connections between optimal observability issues and quantum ergodic properties of the domain under consideration.

Our results open new directions for future research. We mention hereafter some of them.

1. As mentioned before, we expect that the second problem (uniform optimal design problem) not to have any optimal solution in general, except in very particular (degenerate) situations. In other words, in general, an optimal set probably does not exist. Besides, when implementing spectral approximations of the second problem the spillover phenomenon has been underlined and the increasing complexity has been put in evidence on numerical simulations. This indicates the lack of suitability of this spectral approximation procedure of common use in engineering applications. Further investigation is needed to formulate variants of these problems not presenting these instabilities. We mention here two possibilities:
   (a) In [26] we propose a slight modification of the observability inequality under consideration, which consists, e.g. in the Dirichlet case, of replacing the H_0^1 norm by the full H^1 one. Surprisingly enough, we show that the situation is then very different and that, if L is not too small then under QUE type assumptions there exists an optimal set. Consequently, the reinforcement of the observed norm by a compact term contributes to the existence of optimal sets. This can be even achieved by every value of

the volume fraction L by means of a suitable modification of the observed norm (essentially by adding to the H_0^1-norm the L^2 one multiplied by a sufficiently large positive constant). Note however that, when reinforcing the observed norm, the corresponding observability constant decreases. It would then be natural to look for a compromise between ensuring the existence of optimal sets but at the price of deteriorating the observability constant.

(b) A second idea is to define a variant of the criterion (15), by using Cesaro means. This idea is close to the filtering procedures used in [32] in the context of the numerical approximation of controls. The use of Cesaro means should also permit to weaken ergodicity assumptions (see [10]).

In any case, an interesting direction for research is to model and define other kinds of spectral criteria permitting to avoid the spillover phenomenon to recover the existence of an optimal set.

2. In this work we considered wave and Schrödinger equations. In an ongoing work, we are studying the case of the heat equation. As it could be expected, the conclusion is then very different since optimal sets then exist much more easily, due to the intrinsic strong damping of the heat equation.
3. A crucial from the point of view of applications but fully open question is that of the numerical approximation of the optimal sets or densities. Two approaches are then to be considered, the continuous and the discrete one. In this setting a natural question is as follows: do the numerical optimal designs corresponding to discrete dynamics obtained by numerical approximation of the wave equation converge to the continuous optimal design as the mesh size tends to 0? According to the results of [32], one can expect the answer to be negative because of the effect of high-frequency spurious numerical solutions.

   If this were the case the numerical optimal design problem should be reformulated by means of suitable high-frequency filtering techniques.
4. Similar issues can be formulated in the context of homogenization. For instance, we could consider the optimal design problem above on a perforated domain Ω_ε, a rapidly oscillating manifold M_ϵ or for elliptic operators with rapidly oscillating coefficients. The question would then be to know whether, as ϵ tends to zero, the optimal designs do converge in some suitable sense to the optimal design of the limit homogenization problem. Once again one expects the result not to be true in general, due to the distortion that the high-frequency solutions may introduce in the highly heterogeneous medium, with respect to the limit homogeneous one. These issues have been the object of intensive research in the context of controllability problems (see [33]), but, as far as we know, have not been treated so far in the frame of the optimal design problems discussed in this paper.

Acknowledgments. The third author acknowledges the hospitality of support of the organizers of the *Hyp2012 Conference* held in Padova (Italy) in which the contents of the present work were presented in a plenary talk.

REFERENCES

[1] G. Allaire, S. Aubry, F. Jouve, *Eigenfrequency optimization in optimal design*, Comput. Methods Appl. Mech. Engrg. **190** (2001), 3565–3579.

[2] G. Allaire, A. Henrot, *On some recent advances in shape optimization*, C. R. Acad. Sci. Paris, t. 329, Série II b (2001), 383–396.

[3] C. Bardos, G. Lebeau, J. Rauch, *Sharp sufficient conditions for the observation, control, and stabilization of waves from the boundary*, SIAM J. Control Optim., **30** (1992), no. 5, 1024–1065.

[4] J.C. Bellido, A. Donoso, *An optimal design problem in wave propagation*, J. Optim. Theory Appl. **134** (2007), 339–352.

[5] D. Bucur, G. Buttazzo, *Variational methods in shape optimization problems*, Progress in Nonlinear Differential Equations **65**, Birkhäuser Verlag, Basel (2005).

[6] N. Burq, N. Tzvetkov, *Random data Cauchy theory for supercritical wave equations. I. Local theory*, Invent. Math. **173** (2008), no. 3, 449–475.

[7] F. Fahroo, K. Ito, *Optimum damping design for an abstract wave equation*, New directions in control and automation, I (Limassol, 1995), Kybernetika (Prague) 32 (1996), no. 6, 557–574.

[8] F. Faure, S. Nonnenmacher, S. De Bièvre, *Scarred eigenstates for quantum cat maps of minimal periods*, Comm. Math. Phys. **239** (2003), no. 3, 449–492.

[9] M.I. Frecker, *Recent advances in optimization of smart structures and actuators*, Journal of Intelligent Material Systems and Structures **14** (2003), 207–216.

[10] P. Gérard, E. Leichtnam, *Ergodic properties of eigenfunctions for the Dirichlet problem*, Duke Math. J. **71** (1993), 559–607.

[11] P. Hébrard, A. Henrot, *Optimal shape and position of the actuators for the stabilization of a string*, Syst. Cont. Letters **48** (2003), 199–209.

[12] P. Hébrard, A. Henrot, *A spillover phenomenon in the optimal location of actuators*, SIAM J. Control Optim. **44** 2005, 349–366.

[13] A.E. Ingham, *Some trigonometrical inequalities with applications to the theory of series*, Math. Zeitschrift **41** (1936), 367–379.

[14] S. Jaffard, M. Tucsnak, E. Zuazua, *On a theorem of Ingham*, J. Fourier Anal. Appl. **3** (1997), 577–582.

[15] C.S. Kubrusly, H. Malebranche, *Sensors and controllers location in distributed systems - a survey*, Automatica **21** (1985), 117–128.

[16] S. Kumar, J.H. Seinfeld, *Optimal location of measurements for distributed parameter estimation*, IEEE Trans. Autom. Contr. **23**?1978), 690–698.

[17] I. Lasiecka, R. Triggiani, *Exact controllability of the wave equation with Neumann boundary control*, Appl. Math. Optim. **19** (1989), 243–290.

[18] G. Lebeau, *Contrôle de l'equation de Schrödinger*, J. Math. Pures Appl., **71** (1992), 267–291.

[19] J.-L. Lions, *Exact controllability, stabilizability and perturbations for distributed systems*, SIAM Rev. **30** (1988), 1–68.

[20] A. Münch, *Optimal location of the support of the control for the 1-D wave equation: numerical investigations*, Comput. Optim. Appl. **42** (2009), 443–470.

[21] S.L. Padula, R.K. Kincaid, *Optimization strategies for sensor and actuator placement*, Report NASA TM-1999-209126, 1999.

[22] F. Periago, *Optimal shape and position of the support for the internal exact control of a string*, Syst. Cont. Letters **58**?(2009), no. 2, 136–140.

[23] Y. Privat, E. Trélat, E. Zuazua, *Optimal observation of the one-dimensional wave equation*, J. Fourier Anal. Appl. **19** (2013), no. 3, 514–544.

[24] Y. Privat, E. Trélat, E. Zuazua, *Optimal location of controllers for the one-dimensional wave equation*, to appear in Ann. Inst. H. Poincaré Anal. Non Linéaire (2013).

[25] Y. Privat, E. Trélat, E. Zuazua, *Complexity and regularity of maximal energy domains for the wave equation with fixed initial data*, Preprint Hal (2013), 17 pages. Submitted.

[26] Y. Privat, E. Trélat and E. Zuazua, *Optimal observability of the multi-dimensional wave and Schrödinger equations in quantum ergodic domains*, Preprint Hal (2013), 63 pages. Submitted.

[27] O. Sigmund, J.S. Jensen, *Systematic design of phononic band-gap materials and structures by topology optimization*, R. Soc. Lond. Philos. Trans. Ser. A Math. Phys. Eng. Sci. **361** (2003), no. 1806, 1001–1019.

[28] D. Ucinski, M. Patan, *Sensor network design fo the estimation of spatially distributed processes*, Int. J. Appl. Math. Comput. Sci. **20** (2010), no. 3, 459–481.

[29] M. van de Wal, B. Jager, *A review of methods for input/output selection*, Automatica **37** (2001), no. 4, 487–510.

[30] S. Zelditch, *Local and global analysis of eigenfunctions on Riemannian manifolds*, Preprint (2012).

[31] S. Zelditch, M. Zworski, *Ergodicity of eigenfunctions for ergodic billiards*, Comm. Math. Phys. **175** (1996), no. 3, 673–682.

[32] E. Zuazua, Propagation, observation, control and numerical approximation of waves approximated by finite difference method, SIAM Review, **47** (2005), no. 2, 197–243.

[33] E. Zuazua Controllability and Observability of Partial Differential Equations: Some results and open problems in *Handbook of Differential Equations: Evolutionary Equations, vol. 3*, C. M. Dafermos and E. Feireisl eds., Elsevier Science, 2006, 527–621.

E-mail address: yannick.privat@math.cnrs.fr
E-mail address: emmanuel.trelat@upmc.fr
E-mail address: zuazua@bcamath.org

Part 2

Invited Lectures

ORDINARY DIFFERENTIAL EQUATIONS AND SINGULAR INTEGRALS

Gianluca Crippa

Departement Mathematik und Informatik, Universität Basel
Mathematisches Institut, Rheinsprung 21, CH-4051 Basel, Switzerland

Abstract. We present an informal review of some concepts and results from the theory of ordinary differential equations in the non-smooth context, following the approach based on quantitative a priori estimates introduced in [9] and [7].

1. **Introduction.** In this note we give an informal overview on some results from [9] (collaboration with Camillo De Lellis) and [7] (collaboration with François Bouchut) regarding an approach to non-smooth ordinary differential equations based on quantitative a priori estimates.

Given the *velocity field*

$$b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d \tag{1}$$

we consider the *ordinary differential equation*

$$\begin{cases} \dot{X}(t,x) = b(t, X(t,x)) \\ X(0,x) = x\,, \end{cases} \tag{2}$$

where we denote with the "dot" the differentiation with respect to the time variable t. The solution $X : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ is called the *flow* of the velocity field b. We are thus looking for characteristic (or integral) curves of the given velocity field b, i.e., curves with the property that at each point the tangent vector coincides with the value of the given vector field at such point.

The classical Cauchy-Lipschitz theory deals with the case in which the velocity field b is regular enough (Lipschitz with respect to the space variable uniformly with respect to time, see (3)). After a brief review of this smooth theory, in this note we motivate the extension to non-smooth contexts, and we consider first of all the case of $W^{1,p}$ (with $p > 1$) velocity fields, then the case of $W^{1,1}$ velocity fields, and and finally the case of velocity fields whose derivative can be represented as a singular integral operator of an L^1 function. This stratified presentation has the advantage to present the main conceptual and technical differences between these different cases.

The presentation will be *very* informal and only the key points of the proofs will be indicated, with the aim to catch the interest of the reader for the general context and to motivate him or her to further readings on this topic. Emphasis will be put on the ideas, rather than on the details.

2000 *Mathematics Subject Classification.* Primary: 34A12; Secondary: 42B37.

Key words and phrases. Ordinary Differential Equations, Sobolev and Bounded Variation Functions, Singular Integrals.

For this reason, the only references given will be those strictly related to our line of presentation. For a wider presentation of the subject and a detailed bibliography the reader is referred for instance to [5] or [2]. Moreover, the "partial differential equations side" of this problem (well posedness of the transport and the continuity equations in the non-smooth context) will not be addressed here. The interested reader is referred to the two most important papers in this area, namely [10] for the Sobolev case and [1] for the bounded variation case, and again to the bibliographical references in [5] or [2].

2. **The Lipschitz case.** As a warm up, let us start by considering the case of a vector field which is Lipschitz with respect to the space variable uniformly with respect to the time. This means that we assume the existence of a constant L such that

$$|b(t,x)-b(t,y)| \leq L|x-y| \tag{3}$$

for every x, $y \in \mathbb{R}^d$ and every $t \in [0,T]$. Under this assumption, it is known from the classical Cauchy-Lipschitz theorem that a unique solution to (2) exists for every initial point $x \in \mathbb{R}^d$, and moreover the flow $X(t,x)$ inherits the Lipschitz regularity with respect to x.

Uniqueness can be easily proven with the following argument. Consider two (possibly distinct) flows X_1 and X_2. Then for every given $x \in \mathbb{R}^d$ one may compute

$$\begin{aligned}\frac{d}{dt}|X_1(t,x)-X_2(t,x)| \leq & |b(t,X_1(t,x))-b(t,X_2(t,x))| \\ \leq & L|X_1(t,x)-X_2(t,x)|\,,\end{aligned}$$

where in the last inequality we have used (3). Using Gronwall Lemma (and recalling that $X_1(0,x)=X_2(0,x)$) we deduce immediately that $X_1(t,x)=X_2(t,x)$ for every $t \in [0,T]$, i.e., the desired uniqueness.

The proof of the Lipschitz regularity of the flow $X(t,x)$ with respect to x goes along the same line. Fix two points x, $y \in \mathbb{R}^d$ and compute

$$\begin{aligned}\frac{d}{dt}|X(t,x)-X(t,y)| \leq & |b(t,X(t,x))-b(t,X(t,y))| \\ \leq & L|X(t,x)-X(t,y)|\,.\end{aligned}$$

Applying again Gronwall Lemma and observing that $|X(0,x)-X(0,y)|=|x-y|$ we obtain

$$|X(t,x)-X(t,y)| \leq e^{Lt}|x-y|\,, \tag{4}$$

i.e., $X(t,x)$ is Lipschitz with respect to x, and the Lipschitz constant depends exponentially on the Lipschitz constant of the given velocity field b.

3. **Towards non-Lipschitz velocity fields: The regular Lagrangian flow.** After some reflections on the very simple theory presented in the previous section, a natural question arises: how much of such a theory survives when the velocity field b is less regular than Lipschitz?

We immediately realize that, if we stick to "classical" statements (for instance, if we look for uniqueness of the flow for *every* initial point), then the answer is negative. A possible example is very well known: consider in $\mathbb{R}$ the (Hölder but not Lipschitz) vector field $b(x)=\sqrt{|x|}$. Then it is readily checked that $X_1(t,0)\equiv 0$ and $X_2(t,0)=\frac{1}{4}t^2$ are two *distinct* solutions of (2), with the same value ($x=0$) at the initial time. Indeed, it is easy to construct an infinite family of distinct solutions.

One may be discouraged by having a "counterexample" in a still fairly simple situation (an Hölder time-independent vector field in one dimension!). However, non-regular transport phenomena do appear in an ubiquitous fashion in physical models: fluid dynamics, conservation laws, kinetic equations... The reader is referred again to [5] and to [2] for a list of references.

The hope is to find some "milder" issues (some "weakened" version of the pointwise uniqueness for (2), i.e., uniqueness of the flow for *every* point $x \in \mathbb{R}^d$, or of the regularity of the flow with respect to the initial position), together with some "reasonable" context in which such new question may allow a positive answer. The two elements of the new theory will be the following:

(1) The velocity field b may be non-Lipschitz, but it must have "a first-order derivative" in some suitable weak sense. The "bad" velocity field $b(x) = \sqrt{|x|}$ is merely $1/2$-Hölder, hence it possesses only "half a derivative" at the origin.

(2) We content ourselves with showing uniqueness of (almost) measure preserving flow solutions of (2). That is, we drop the pointwise framework, and we just consider as "admissible" solutions to (2) those flows $X(t,x)$ for which, at every time $t \in [0,T]$, the map $X(t,\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ does not squeeze or expand sets in a crazy fashion. The non-unique trajectories produced by $b(x) = \sqrt{|x|}$ do indeed "compress" long segments into one point, the origin of $\mathbb{R}$. (The non-uniqueness is dynamically due to the stopping of the trajectories at the origin). The reader will notice that the origin is precisely the point at which the regularity of b is degenerating.

We now specify what we mean with "measure preserving flow solution":

Definition 3.1 (Regular Lagrangian flow)**.** We say that a map $X : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ is a *regular Lagrangian flow* associated to the vector field b if

(i) For $\mathcal{L}^d$-a.e. $x \in \mathbb{R}^d$ the map $t \mapsto X(t,x)$ is a distributional solution to the ordinary differential equation $\dot\gamma(t) = b(t,\gamma(t))$, with $\gamma(0) = x$;

(ii) There exists some constant $M > 0$ such that the *compressibility condition*

$$X(t,\cdot)_{\#}\mathcal{L}^d \leq M\mathcal{L}^d \qquad \text{for every } t \in [0,T] \tag{5}$$

holds.

The condition in (5) involves the push-forward of the d-dimensional Lebesgue measure $\mathcal{L}^d$ and can be equivalently reformulated as follows: there exists some constant $M > 0$ such that for every $t \in [0,T]$ and every $\varphi \in C_c(\mathbb{R}^d)$ with $\varphi \geq 0$ there holds

$$\int_{\mathbb{R}^d} \varphi(X(t,x))\,dx \leq M \int_{\mathbb{R}^d} \varphi(x)\,dx\,.$$

This means that we require *a priori*, i.e., as a sort of "selection condition" for our notion of solution, a quantitative control on how much the flow compresses d-dimensional sets. For simplicity, in the following presentation, we shall restrict our attention to those regular Lagrangian flow which *exactly* preserve the Lebesgue measure (in the smooth context, this corresponds to the condition of b having zero divergence, thanks to Liouville's Theorem). We can formulate it by saying that "changes of variable along the flow are performed for free", that is, for every $\varphi \in C_c(\mathbb{R}^d)$ we can compute

$$\int_{\mathbb{R}^d} \varphi(X(t,x))\,dx = \int_{\mathbb{R}^d} \varphi(x)\,dx\,. \tag{6}$$

4. **A stable formal estimate, and a new integral quantity.** In order to make our computations typographically more clear, in the rest of this note we shall only consider time-independent vector fields. The passage to the time-dependent case does not give rise to any complication in the argument.

The natural attempt is to rephrase the strategy of §2 in a way that will be robust when lowering the regularity of the velocity field from Lipschitz to "weakly differentiable". Denoting by ∇ the gradient with respect to the x variable, we can *formally* compute as follows:

$$\begin{aligned} \frac{d}{dt}\log|\nabla X| \leq & \frac{1}{|\nabla X|}\left|\frac{d}{dt}\nabla X\right| \\ = & \frac{1}{|\nabla X|}\left|\nabla\big(b(X)\big)\right| = |\nabla b|(X)\,. \end{aligned} \tag{7}$$

Notice that this computation is effective in the case of a smooth velocity field b possessing a smooth flow X. Anyhow, in the Lipschitz context, it allows us to recover the estimate for the regularity of the flow with respect to the initial position already established in (4). Indeed, if b satisfies (3), then $|\nabla b| \leq L$, and so by integrating (7) we deduce

$$\log|\nabla X| \leq Lt + \log|\nabla \mathrm{Id}| = Lt\,,$$

from which (4).

We apply a similar strategy in order to show uniqueness. For this we fix a small parameter $\delta > 0$. If X_1 and X_2 are flows of b, then we compute

$$\frac{d}{dt}\log\left(1+\frac{|X_1-X_2|}{\delta}\right) \leq \frac{\delta}{\delta+|X_1-X_2|}\,\frac{|b(X_1)-b(X_2)|}{\delta} \leq L\,,$$

where L is the Lipschitz constant of b. Hence

$$\log\left(1+\frac{|X_1-X_2|}{\delta}\right) \leq Lt + \log\left(1+\frac{|X_1(0,\cdot)-X_2(0,\cdot)|}{\delta}\right) = Lt\,,$$

and finally

$$\frac{|X_1-X_2|}{\delta} \leq e^{Lt}\,.$$

Since $\delta > 0$ can be chosen arbitrarily small, we deduce that $X_1 = X_2$.

The remarkable advantage of this argument is that it allows an integral version, which can be used for non-Lipschitz vector fields. In the rest of this note, we focus on the uniqueness issue for the regular Lagrangian flow associated to a given velocity field, as defined in Definition 3.1. Given a velocity field b, two (possibly distinct) associated regular Lagrangian flows X_1 and X_2, and a small parameter $\delta > 0$ we consider

$$\Phi_\delta(t) = \int \log\left(1+\frac{|X_1(t,x)-X_2(t,x)|}{\delta}\right)dx\,. \tag{8}$$

Notice that suitable truncations are necessary in order to make this integral convergent, but for the sake of clarity in this exposition we will ignore this technical issue.

This integral functional has been first considered in a joint paper with De Lellis [9], where we were inspired by some similar computations due to Ambrosio, Lecumberry and Maniglia [3]. Although we are now focussing our presentation on the uniqueness issue, we remark that similar integral quantities are useful to prove regularity, compactness and quantitative stability rates for regular Lagrangian flows.

5. **A condition for uniqueness.** In the unlucky situation of non-uniqueness of the regular Lagrangian flow, that is, when there are two distinct regular Lagrangian flows X_1 and X_2, we easily discover that there is a set $A \subset \mathbb{R}^d$ of measure at least $\alpha > 0$ such that $|X_1(t,x) - X_2(t,x)| \geq \gamma > 0$ for some $t \in [0,T]$ and for all $x \in A$. Hence we can estimate the integral functional $\Phi_\delta(t)$ from below as follows:

$$\Phi_\delta(t) \geq \int_A \log\left(1 + \frac{\gamma}{\delta}\right) dx \geq \alpha \log\left(1 + \frac{\gamma}{\delta}\right) .$$

We then discover that a condition guaranteeing uniqueness is:

$$\frac{\Phi_\delta}{\log\left(\frac{1}{\delta}\right)} \to 0 \qquad\qquad \text{as } \delta \downarrow 0 . \tag{9}$$

This means that a good strategy to prove uniqueness is to derive upper bounds for the integral functional $\Phi_\delta(t)$. The natural computation starts with a time differentiation, aimed at making the difference quotients of the velocity field b appear. We calculate

$$\begin{aligned} \Phi'_\delta(t) &\leq \int \frac{\partial_t |X_1 - X_2|}{\delta + |X_1 - X_2|}\, dx \leq \int \frac{|b(X_1) - b(X_2)|}{\delta + |X_1 - X_2|}\, dx \\ &\leq \int \min\left\{ \frac{2\|b\|_{L^\infty}}{\delta} \ ; \ \frac{|b(X_1) - b(X_2)|}{|X_1 - X_2|} \right\} dx . \end{aligned} \tag{10}$$

For a Lipschitz velocity field, it is sufficient to estimate

$$\frac{|b(X_1) - b(X_2)|}{|X_1 - X_2|} \leq L$$

in (10) to obtain that $\Phi'_\delta(t)$ (and thus $\Phi_\delta(t)$) is bounded by a constant. We recover again uniqueness in the Lipschitz case.

But a milder condition to obtain boundedness of $\Phi_\delta(t)$ would be the difference-quotients estimate

$$\frac{|b(x) - b(y)|}{|x - y|} \leq \psi(x) + \psi(y) \tag{11}$$

for some function $\psi \in L^1_{\mathrm{loc}}$. Indeed, getting back to (10), we estimate

$$\Phi'_\delta(t) \leq \int (\psi(X_1) + \psi(X_2))\, dx = 2 \int \psi(x)\, dx ,$$

where in the last equality we change variables as in (6), and we conclude again that $\Phi_\delta(t)$ is bounded by a constant. Notice that the first term in the minimum in (10) has been simply neglected. A smarter computation allowing for its use will be explained in §7.

6. **Maximal functions, strong and weak estimates, and uniqueness for $W^{1,p}$ velocity fields with $p > 1$.** In the paper [9] with De Lellis we realized that condition (11) is satisfied (and so uniqueness holds) in the case of velocity fields with Sobolev $W^{1,p}$ regularity, for any $p > 1$.

Indeed, in such case, the estimate for the difference quotients

$$\frac{|b(x) - b(y)|}{|x - y|} \leq C_{p,d}\Big(MDb(x) + MDb(y)\Big) \tag{12}$$

holds, where the *maximal function* of a locally summable function f is defined by

$$Mf(x) = \sup_{r>0} \frac{1}{\mathcal{L}^d(B(x,r))} \int_{B(x,r)} |f(y)|\, dy . \tag{13}$$

It is classical (see for instance [11]) that the maximal function enjoys the *strong estimate*

$$\|Mf\|_{L^p} \leq C_{d,p} \|f\|_{L^p} \tag{14}$$

for any $1 < p \leq \infty$, but unfortunately this fails for $p = 1$. In that case, only the *weak estimate*

$$\mathcal{L}^d\Big(\Big\{x \ : \ |Mf(x)| > \lambda\Big\}\Big) \leq C_{d,1} \frac{\|f\|_{L^1}}{\lambda} \qquad\qquad \text{for any } \lambda > 0 \tag{15}$$

is available.

We see from (12) that (11) holds if we take

$$\psi = MDb\,, \tag{16}$$

and using the strong estimate (14) we deduce from the assumption $Db \in L^p$ that $\psi \in L^p$, and uniqueness follows. The failure of the strong estimate (14) for $p = 1$ is precisely the reason why the uniqueness theorem in [9] was limited to the case $p > 1$. The cases of $W^{1,1}$ or even of BV velocity fields were missing.

7. Uniqueness for $W^{1,1}$ velocity fields.

Together with Bouchut, we discovered in [7] how to extend this argument to the case of $W^{1,1}$ velocity fields. The proof uses some more elaborate tools from harmonic analysis (as a general reference the interested reader can consult [11]).

Introducing the quantity

$$|||f|||_{M^1} = \sup \Big\{ \lambda\, \mathcal{L}^d\big(\{|f| > \lambda\}\big) \ : \quad \lambda > 0 \Big\}\,, \tag{17}$$

we see that (15) can be rewritten as

$$|||Mf|||_{M^1} \leq C_{d,1} \|f\|_{L^1}\,. \tag{18}$$

The space M^1 consisting of all functions for which the quantity in (17) is finite is called weak Lebesgue space (or alternatively Lorentz space or Marcinkiewicz space). It is endowed with the natural pseudo-norm $|||f|||_{M^1}$, which is however not a norm, lacking the subadditivity property. Notice that M^1 is strictly bigger than L^1.

Going back to (16), and observing that we are now concerned with the case when $Db \in L^1$, we discover that condition (11) now is satisfied for some $\psi \in M^1$. In general ψ does not belong to $L^1_{\rm loc}$: we need some additional considerations in order to conclude uniqueness.

Let us go back to (10). Using (11) and changing variable using (6) we obtain

$$\Phi'_\delta(t) \leq \int_{\mathbb{R}^d} \min \left\{ \frac{2\|b\|_{L^\infty}}{\delta} \ \ ; \ \ 2\psi \right\} dx\,. \tag{19}$$

None of the two terms inside the minimum suffices by itself to deduce (9). The first term is L^∞, but with a norm which blows up as $\delta \downarrow 0$, while the second term is merely M^1. However, an interpolation inequality between M^1 and L^∞ is at our disposal (see [7] for a proof):

$$\|f\|_{L^1} \leq |||f|||_{M^1} \left[1 + \log\left(C \frac{\|f\|_{L^\infty}}{|||f|||_{M^1}} \right) \right]\,.$$

We apply this interpolation inequality to

$$f = \min \left\{ \frac{2\|b\|_{L^\infty}}{\delta} \ \ ; \ \ 2\psi \right\}\,,$$

and we observe that

$$\|f\|_{L^\infty} = \frac{2\|b\|_{L^\infty}}{\delta} \leq \frac{C}{\delta}$$

$$\text{and} \qquad |||f|||_{M^1} = 2|||\psi|||_{M^1} = 2|||MDb|||_{M^1} \leq C\|Db\|_{L^1} \,,$$

by (18). We go back to (19) and employing these estimates we deduce

$$\Phi'_\delta(t) \leq C\|Db\|_{L^1}\left[1 + \log\left(\frac{C}{\delta\|Db\|_{L^1}}\right)\right] . \tag{20}$$

Remember our criterion for uniqueness (9): the bound (20) is exactly the critical growth of the functional $\Phi_\delta(t)$ which is relevant for the uniqueness! In fact, the ratio that criterion (9) requires to be infinitesimal for $\delta \downarrow 0$, is now merely bounded. Still, we cannot conclude uniqueness with this information only.

It is at this point that we exploit the information that Db is an L^1 function, and not just a Radon measure. (Notice that all arguments carried out until now would work verbatim if we substitute $\|\cdot\|_{L^1}$ with the total variation norm $\|\cdot\|_{\mathcal{M}}$, i.e., for b being a BV velocity field). Up to a remainder in L^2, we can assume that Db not only *belongs to* L^1, but also that it is *small in* L^1. (The existence of such a decomposition is due to the equi-integrability of L^1 functions). This smallness allows to fullfill the criterion (9), while the residual part of the functional originated by the L^2 remainder can be treated with the arguments of §6. This allows to conclude uniqueness for $W^{1,1}$ velocity fields, but it is still far from giving any result for BV velocity fields: a measure does *not* allow a decomposition in a small L^1 part plus an L^2 remainder!

8. **Vector fields whose derivative is a singular integral of an L^1 function.** The strategy described in the previous section extends in a (technical but) natural way to the case in which the derivatives of the velocity field b can be expressed as

$$\partial_j b^i = \sum_k S_{ijk} g_{ijk} \,,$$

where $g_{ijk} \in L^1(\mathbb{R}^d)$ and every S_{ijk} is a *singular integral operator.* In more details, we assume that any of these operators can be expressed as a convolution

$$S_{ijk} g_{ijk} = K_{ijk} * g_{ijk} \,,$$

where the singular kernel K_{ijk} is smooth away of the origin of $\mathbb{R}^d$, is homogeneous of degree $-d$ and satisfies the usual cancellation property.

Observe that this class of vector fields includes $W^{1,1}$. However, it does neither include BV, nor it is included in BV. The relevance of this class of vector fields is due to their appearance in some physical problems: for instance, in two dimensional incompressible fluid dynamics, this is the regularity enjoyed by fluid velocities with L^1 vorticity.

It is well known (see again [11]) that singular integrals enjoy the same estimates as maximal functions: namely, strong estimates for $1 < p < \infty$

$$\|Sf\|_{L^p} \leq C_{d,p}\|f\|_{L^p}$$

(the case $p = \infty$ has now to be excluded), and the weak estimates for the case $p = 1$

$$|||Sf|||_{M^1} \leq C_{d,1}\|f\|_{L^1} \,.$$

Also in this case, no strong estimate for $p = 1$ is available.

One basic consequence of the cancellation property assumed for the singular kernels under consideration is the weak estimate for the composition of two singular integral operators. Namely, if we consider a composition $S = S_2 \circ S_1$, the associated singular kernel is given by the convolution $K = K_2 * K_1$, and it is again a singular kernel. Thus we still have

$$|||Sf|||_{M^1} = |||S_2 \circ S_1 f|||_{M^1} \leq C\|f\|_{L^1} \,. \tag{21}$$

Note carefully that estimate (21) *cannot* be obtained by composing the two analogue estimates (from L^1 to M^1) which hold for the two singular integral operators S_1 and S_2 separately. At a formal level, (21) requires *cancellations* in the convolutions.

9. **Back to the proof of the uniqueness.** We describe now how to modify the strategy described in §7 in order to prove uniqueness of the regular Lagrangian flow associated to vector fields with the regularity described in §8. This result is contained in [7].

Going back to (16), we realise that in the present context we have

$$\psi = MSg \,,$$

for $g \in L^1$. We thus need a bound of the type

$$|||\psi|||_{M^1} \leq C\|g\|_{L^1} \,, \tag{22}$$

in order to conclude the proof along the lines of §7. In general, however, estimate (22) does not hold if the classical maximal function (13) is considered. Inspired by the cancellation phenomenon which allows (21), we can prove that (22) holds if we consider instead a smooth version of the maximal function, defined as

$$M_\rho f(x) = \sup_{r>0} \left| \int_{\mathbb{R}^d} \rho_r(x-y) f(y)\, dy \right| \,,$$

where ρ is a given smooth convolution kernel. This smooth version of the maximal function is well known in the context of Hardy spaces, under the name of grand maximal function. It is possible to prove that

$$|||\psi|||_{M^1} = |||M_\rho Sg|||_{M^1} \leq C\|g\|_{L^1} \,,$$

and this estimate is sufficient to conclude using the strategy in §7, yielding uniqueness of the regular Lagrangian flow for the class of vector fields considered in §8.

Acknowledgments. I thank the scientific committee of the conference HYP2012 for the invitation to give the talk that gave rise to this note. The topics presented here are part of very enjoyable collaborations with Camillo De Lellis and François Bouchut. I thank Anna Bohun and Laura Keller for some comments on a preliminary version of this note. I also acknowledge support by the Swiss National Science Foundation through the grant 140232

REFERENCES

[1] L. Ambrosio, *Transport equation and Cauchy problem for BV vector fields*, Invent. Math., **158** (2004), 227–260.

[2] L. Ambrosio and G. Crippa, *Existence, uniqueness, stability and differentiability properties of the flow associated to weakly differentiable vector fields*, in "Transport equations and multi-D hyperbolic conservation laws", Lect. Notes Unione Mat. Ital., **5** (2008), 1–41.

[3] L. Ambrosio, M. Lecumberry and S. Maniglia, *Lipschitz regularity and approximate differentiability of the DiPerna–Lions flow*, Rend. Sem. Mat. Univ. Padova, **114** (2005), 29–50.

[4] G. Crippa, *The ordinary differential equation with non-Lipschitz vector fields*, Boll. Unione Mat. Ital., **1** (2008), 333–348.

[5] G. Crippa, "The flow associated to weakly differentiable vector fields", Theses of Scuola Normale Superiore di Pisa (New Series), **12**, Edizioni della Normale, Pisa, 2009.

[6] F. Bouchut and G. Crippa, *Équations de transport à coefficient dont le gradient est donné par une intégrale singulière*, Sémin. Équ. Dériv. Partielles, Exp. No. I, École Polytech., Palaiseau, 2009.

[7] F. Bouchut and G. Crippa, *Lagrangian flows for vector fields with gradient given by a singular integral*, J. Hyperbolic Differ. Equ., **10** (2013), 235–282.

[8] G. Crippa and C. De Lellis, *Regularity and compactness for the DiPerna-Lions flow*, Hyperbolic problems: theory, numerics, applications, Springer, Berlin (2008), 423–430.

[9] G. Crippa and C. De Lellis, *Estimates and regularity results for the DiPerna-Lions flow*, J. Reine Angew. Math., **616** (2008), 15–46.

[10] R. J. DiPerna and P.-L. Lions, *Ordinary differential equations, transport theory and Sobolev spaces*, Invent. Math., **98** (1989), 511–547.

[11] E. Stein, "Singular integrals and differentiability properties of functions", Princeton University Press, 1970.

E-mail address: gianluca.crippa@unibas.ch

ENTROPY VISCOSITY FOR THE EULER EQUATIONS AND QUESTIONS REGARDING PARABOLIC REGULARIZATION

Jean-Luc Guermond and Bojan Popov

Department of Mathematics
Texas A&M University
College Station, TX 77843-3368, USA

Abstract. This note describes a general class of regularizations for the compressible Euler equations. A unique regularization is identified that is compatible with all the generalized entropies à la [8] and satisfies the minimum entropy principle. All the results announced herein will be reported in detail in [5].

1. **Introduction.** A new numerical method for approximating nonlinear conservation laws using an artificial viscosity based on entropy production has been described in [4, 6, 16]. This so-called entropy viscosity method uses finite elements, either continuous or discontinuous, and consists of augmenting the numerical discretization at hand with a parabolic regularization where the nonlinear viscosity is based on the local size of a discrete entropy production. The idea of using the entropy to design numerical methods for nonlinear conservation equations is not new. For instance it is shown in [11] that the entropy production can be used as an a posteriori error indicator and therefore is useful for adaptive strategies. The main originality of the entropy viscosity method is that one directly uses the entropy production to construct an artificial viscosity. This strategy makes an automatic distinction between shocks and contact discontinuities. This method is simple to program and does not use any flux or slope limiters. The method can be reasonably justified for scalar conservation equations. For instance it is now well established that the solution of the parabolic regularization of a scalar conservation equation converges to the entropy solution as the regularization parameter goes to zero. This fundamental fact is the key justification for constructing approximation techniques based on artificial viscosity. Stability results have been established in [10] and [1] for fully discrete versions of the entropy viscosity method for scalar conservation equations using simple entropies. The extension of this strategy to hyperbolic systems is not so clear, since the question of how parabolic regularizations should be constructed for hyperbolic systems is still an open problem. In particular, our experience is that the Navier-Stokes system is not a robust regularization of the Euler system,

2000 *Mathematics Subject Classification.* Primary: 35L65, 65M60; Secondary: 76N15.

Key words and phrases. Conservation equations, hyperbolic systems, parabolic regularization, entropy, viscosity solutions.

This material is based upon work supported in part by the National Science Foundation grants DMS-1015984 and DMS-1217262, by the Air Force Office of Scientific Research, USAF, under grant/contract number FA9550-09-1-0424, FA99550-12-0358, and by Award No. KUS-C1-016-04, made by King Abdullah University of Science and Technology (KAUST).

one key reason being that there is no mechanism therein to help the density to stay positive, another one being that the Navier-Stokes regularization is known to violate the minimum entropy principle if the thermal diffusivity is zero.

The objective of this note is to investigate a nonstandard family of regularization of the Euler system that can serve as a reasonable starting point for an entropy viscosity technique. We identify a single family that preserves positivity of the density, satisfies a minimum entropy principle (see [14]), and is compatible with the largest class of generalized entropies inequalities of [8].

2. **Statement of the problem.** Consider the compressible Euler equations in conservative form in $\mathbb{R}^d$,

$$\partial_t \boldsymbol{U} + \nabla{\cdot}\boldsymbol{F}(\boldsymbol{U}) = 0, \qquad \boldsymbol{U}(\boldsymbol{x},0) = (\rho_0(x), \boldsymbol{m}_0(\boldsymbol{x}), E(\boldsymbol{x}))^T, \tag{1}$$

where $\boldsymbol{U} = (\rho, \boldsymbol{m}, E)^T$, $\boldsymbol{F}(\boldsymbol{U}) = (\boldsymbol{m}, \boldsymbol{u}\otimes\boldsymbol{m} + p\mathbb{I}, \boldsymbol{u}(E+p))^T$. The dependent variables are the density, ρ, the momentum, $\boldsymbol{m}$ and the total energy, E. We adopt the usual convention that for any vectors $\boldsymbol{a}$, $\boldsymbol{b}$, with entries $\{a_i\}_{i=1,\dots,d}$, $\{b_i\}_{i=1,\dots,d}$, the following holds: $(\boldsymbol{a}\otimes\boldsymbol{b})_{ij} = a_i b_j$ and $\nabla{\cdot}\boldsymbol{a} = \partial_{x_j} a_j$, $(\nabla\boldsymbol{a})_{ij} = \partial_{x_i} a_j$. Moreover, for any order 2 tensors $\mathbb{g}$, $\mathbb{h}$, with entries $\{g_{ij}\}_{i,j=1,\dots,d}$, $\{h_{ij}\}_{i,j=1,\dots,d}$, we define $(\nabla{\cdot}\mathbb{g})_j = \partial_{x_i} g_{ij}$, $\boldsymbol{a}{\cdot}\nabla = a_i\partial_{x_i}$, $(\mathbb{g}{\cdot}\boldsymbol{a})_i = g_{ij}a_j$, $\mathbb{g}{:}\mathbb{h} = g_{ij}\mathbb{h}_{ij}$ where repeated indices are summed from 1 to d.

The equation of state is assumed to derive from a specific entropy, $s(\rho, e)$, through the thermodynamics identity: $T\,\mathrm{d}s := \mathrm{d}e + p\,\mathrm{d}\tau$, where $\tau := \rho^{-1}$, $e := \rho^{-1}E - \frac{1}{2}\boldsymbol{u}^2$ is the specific internal energy, $\boldsymbol{u} := \rho^{-1}\boldsymbol{m}$ is the velocity of the fluid particles. For instance it is usual to take $s = \log(e^{\frac{1}{\gamma-1}}\rho^{-1})$ for a polytropic ideal gas. Using the notation $s_e := \frac{\partial s}{\partial e}$ and $s_\rho := \frac{\partial s}{\partial \rho}$, this definition implies that $s_e := T^{-1}$, $s_\rho := -pT^{-1}\rho^{-2}$. The equation of state takes the form $p := -\rho^2 s_\rho s_e^{-1}$. The key structural assumption is that $-s$ is strictly convex with respect to $\tau := \rho^{-1}$ and e. Upon introducing $\sigma(\tau, e) := s(\rho, e)$, the convexity hypothesis is equivalent to assuming that $\sigma_{\tau\tau} \le 0$, $\sigma_{ee} \le 0$, and $\sigma_{\tau\tau}\sigma_{ee} - \sigma_{\tau e}^2 \le 0$. This in turn implies that $\partial_\rho(\rho^2 s_\rho) < 0$, $s_{ee} < 0$, $0 < \partial_\rho(\rho^2 s_\rho)s_{ee} - \rho^2 s_{\rho e}^2$, or equivalently that the following matrix

$$\Sigma := \begin{pmatrix} \rho^{-1}\partial_\rho(\rho^2 s_\rho) & \rho s_{\rho e} \\ \rho s_{\rho e} & \rho s_{ee} \end{pmatrix}, \tag{2}$$

is negative definite. In the rest of the note we assume that the entropy is strictly convex and the temperature is positive, i.e., $0 < s_e$.

A physical way to regularize the Euler system (1) consists of considering this system as the limit of the Navier-Stokes equations. We claim that the Navier-Stokes regularization is not appropriate for numerical purposes. The first problem that we identify is that the minimum entropy principle cannot be satisfied for general initial data if the thermal dissipation is not zero. More precisely, assuming that the thermal diffusivity is nonzero, for any $r \in \mathbb{R}$, there exist initial data so that the set $\{s \ge r\}$ is not positively invariant, where s is the specific entropy, see e.g., [12, Thm 8.2.3]. Another argument often invoked against the presence of thermal dissipation is that it is incompatible with symmetrization of the Navier-Stokes system when using the generalized entropies of [7] for polytropic ideal gases. The function $\rho f(s)$ is said to be a generalized entropy if $f'(\gamma-1)\gamma^{-1} - f'' > 0$, $f' > 0$ and $f \in \mathcal{C}^2(\mathbb{R};\mathbb{R})$. It is proved in [9] that the only generalized entropy that symmetrizes the Navier-Stokes system is the trivial one ρs when the thermal diffusivity is nonzero, see also [15, (2.11) and Remark 2, page 460]. Although symmetrization of the viscous fluxes is not necessary to establish entropy dissipation (see e.g., [13, §1.1]), it is nevertheless true that the Navier-Stokes system violates generalized entropy inequalities if $f''(s) \neq 0$.

The objective of this note is to introduce a regularization of (1) that is compatible with thermodynamics and can be used for numerical approximations.

3. **General regularization.**

We investigate in this section the properties of following general regularization for the Euler system:

$$\partial_t \boldsymbol{U} + \nabla\cdot \boldsymbol{F}(\boldsymbol{U}) = \nabla\cdot \boldsymbol{T}, \qquad \boldsymbol{U}(\boldsymbol{x},0) = (\rho_0(x), \boldsymbol{m}_0(\boldsymbol{x}), E(\boldsymbol{x}))^T, \tag{3}$$

where $\boldsymbol{T} = (\boldsymbol{f}, \mathbb{g}, \boldsymbol{h} + \mathbb{g}\cdot\boldsymbol{u})^T$ and the fluxes $\boldsymbol{f}$, $\mathbb{g}$, and $\boldsymbol{h}$ are as general as possible. A regularization theory for general nonlinear hyperbolic system has been developed in [13] and [12, Chap 6]. Our objective in this note is more restrictive. We want to construct the fluxes $\boldsymbol{f}$, $\mathbb{g}$, and $\boldsymbol{h}$ so that (3) gives a positive density, gives a minimum principle on the specific entropy, and is compatible with a large class of entropies. It is assumed in the rest of the note that (3) has a smooth solution.

3.1. **Positivity of the density.**

Modulo mild regularity assumptions on the velocity, the theory of second-order elliptic equations implies that $\boldsymbol{f} = a(\rho, e)\nabla\rho$ is appropriate to guaranty the positivity of the density, where $a(\rho, e)$ is a smooth positive function. The following is established in [5]

Lemma 3.1 (Positive Density Principle)**.** *Let $\boldsymbol{f} = a(\rho, e)\nabla\rho$ in (3), with $a \in L^\infty(\mathbb{R}^2;\mathbb{R})$ and $\inf_{(\xi,\eta)\in\mathbb{R}^2} a(\xi,\eta) > 0$. Assume that $\boldsymbol{u}$ and $\nabla\cdot\boldsymbol{u} \in L^\infty(\mathbb{R}^d\times\mathbb{R}_+;\mathbb{R})$. Assume also that there are constant states at infinity ρ^∞, $\boldsymbol{u}^\infty$, so that the supports of $\rho(\cdot,\cdot) - \rho^\infty$ and $\boldsymbol{u}(\cdot,\cdot) - \boldsymbol{u}^\infty$ are compact in $\mathbb{R}^d\times(0,t)$, for any $t > 0$. Assume finally that $\rho_0 - \rho_\infty \in L^2(\mathbb{R}^d;\mathbb{R})$. Then the solution of (3) is such that*

$$\operatorname*{ess\,inf}_{\boldsymbol{x}\in\mathbb{R}^d} \rho(\boldsymbol{x},t) \geq 0, \qquad \forall t \geq 0. \tag{4}$$

3.2. **Minimum entropy principle.**

Since physically admissible weak solutions of the Euler equations satisfy the following inequality $\partial_t s + \boldsymbol{u}\cdot\nabla s \geq 0$, they also satisfy a minimum entropy principle, i.e., the set $\{s \geq s_0\}$, where s_0 is the infimum of the specific entropy of the initial data, is positively invariant. The importance of the minimum entropy principle has been established by [14].

Requesting that the triple $\boldsymbol{f}$, $\boldsymbol{h}$ and $\mathbb{G}$ be such that the solution of (3) satisfies a minimum entropy principle narrows down the choices that can be made for the viscous fluxes. It is shown in [5] that the following structure is sufficient for this purpose:

$$\boldsymbol{f} = a(\rho, e)\nabla\rho \qquad\qquad a(\rho, e) \geq 0, \tag{5}$$

$$\mathbb{g} = \mathbb{G}(\nabla^s\boldsymbol{u}) + \boldsymbol{f}\otimes\boldsymbol{u}, \qquad\qquad \mathbb{G}(\nabla^s\boldsymbol{u}){:}\nabla\boldsymbol{u} \geq 0, \tag{6}$$

$$\boldsymbol{h} = \boldsymbol{l} - \tfrac{1}{2}\boldsymbol{u}^2\boldsymbol{f}, \quad \boldsymbol{l} = (a-d)(p\rho^{-1} + e)\nabla\rho + d\nabla(\rho e) \qquad d(\rho, e) \geq 0. \tag{7}$$

Theorem 3.2 (Minimum Entropy Principle)**.** *Assume that ρ_0 and e_0 are constant outside some compact set. Assume also that (5)-(6)-(7) hold. Assume that the solution to (3) is smooth, then the minimum entropy principle holds,*

$$\operatorname*{ess\,inf}_{\boldsymbol{x}\in\mathbb{R}^d} s(\boldsymbol{x},t) \geq \operatorname*{ess\,inf}_{\boldsymbol{x}\in\mathbb{R}^d} s_0(\boldsymbol{x}), \qquad \forall t \geq 0.$$

3.3. **Generalized entropies.** We investigate in this section whether the regularization of the Euler equations (3) is compatible with some or all generalized entropy inequalities identified in [8]. A function $\rho f(s)$ is called a generalized entropy if f is twice differentiable and

$$f'(s) > 0, \qquad f'(s)c_p^{-1} - f''(s) > 0, \qquad \forall(\rho, e) \in \mathbb{R}^2_+, \tag{8}$$

where $c_p(\rho, e) = T\partial_T s(p, T)$ is the specific heat at constant pressure. It is shown in [8] that $-\rho f(s)$ is strictly convex with respect to ρ^{-1} and e if and only if (8) holds, i.e., (8) characterizes the maximal set of admissible entropies for the compressible Euler equations that are of the form $\rho f(s)$. The following result is proved in [5]:

Theorem 3.3 (Entropy Inequalities)**.** *Assume that* (6)-(5)-(7) *hold. Any weak solution to the regularized system* (3) *satisfies the entropy inequality*

$$\partial_t(\rho f(s)) + \nabla\cdot\big(\boldsymbol{u}\rho f(s) - d\rho\nabla f(s) - af(s)\nabla\rho\big) \geq 0, \tag{9}$$

for all generalized entropies $\rho f(s)$ if and only if $a = d$.

Corollary 1. *Any weak solution to the regularized system* (3) *satisfies the entropy inequality* (9) *for the physical entropy ρs (i.e., $f(s) = s$) if* $2\Gamma - 2\Delta^{\frac{1}{2}} < 1 - \frac{a}{d} < 2\Gamma + 2\Delta^{\frac{1}{2}}$ *where $\Gamma = det(\Sigma)\rho^2 s_e^{-2} p_e^{-2}$ and $\Delta = \Gamma(1 + \Gamma)$.*

In the case of a polytropic ideal gases, i.e., $s = \log(e^{\frac{1}{\gamma-1}}\rho^{-1})$ with $\gamma > 1$, we have $c_p = \gamma(\gamma-1)^{-1}$, $\det(\Sigma) = (\gamma-1)^{-1}e^{-2}$, $\boldsymbol{f} = a\nabla\rho$, and $\boldsymbol{l} = \gamma de(\frac{a}{d}-1+\frac{1}{\gamma})\nabla\rho + d\rho\nabla e$. The range for the ratio ad^{-1} for Corollary 1 to hold is

$$\frac{2}{\gamma-1}(1-\sqrt{\gamma}) < 1 - \frac{a}{d} < \frac{2}{\gamma-1}(1+\sqrt{\gamma}). \tag{10}$$

In particular the choice $1 - \frac{a}{d} = \frac{1}{\gamma}$ is clearly in the admissible range. For this choice $\boldsymbol{l} = d\rho\nabla e$ and $\boldsymbol{f} = d\frac{\gamma-1}{\gamma}\nabla\rho$, i.e., $\boldsymbol{l}$ does not involve any mass dissipation.

4. **Conclusions.**

We show in this section that the regularization proposed above reconciles the Navier-Stokes and the parabolic regularization points of view.

4.1. **Parabolic regularization.** One natural question that comes to mind is how different is the general regularization (3) from the simple parabolic regularization:

$$\partial_t\boldsymbol{U} + \nabla\cdot\boldsymbol{F}(\boldsymbol{U}) = \epsilon\Delta\boldsymbol{U}, \qquad \boldsymbol{U}(\boldsymbol{x}, 0) = \boldsymbol{U}_0(\boldsymbol{x}), \tag{11}$$

where $\boldsymbol{U} = (\rho, \boldsymbol{m}, E)^T$, $\boldsymbol{F}(\boldsymbol{U}) = (\boldsymbol{m}, \boldsymbol{u}\otimes\boldsymbol{m} + p\mathbb{I}, \boldsymbol{u}(E+p))^T$. The answer is given by the following, somewhat a priori frustrating result:

Proposition 1 (Parabolic regularization)**.** *The parabolic regularization* (11) *is identical to* (3) *with* (6)–(7) *where $a = d = \epsilon$, $\mathbb{G} = \epsilon\rho\nabla\boldsymbol{u}$.*

Even when $a = d$, one important interest of the class of regularization (3), when compared to the monolithic parabolic regularization (11), is that it decouples the regularization on the velocity from that on the density and internal energy. In particular the regularization on the velocity can be made rotation invariant by making the tensor $\mathbb{G}$ a function of the symmetric gradient $\nabla^s\boldsymbol{u}$. This decoupling was not a priori evident when looking at (11).

4.2. **Connection with phenomenological models.** Using the assumptions (6)–(7) in the balance equation (3) we obtain the following system:

$$\partial_t \rho + \nabla\cdot\boldsymbol{m} - \nabla\cdot\boldsymbol{f} = 0, \tag{12}$$

$$\partial_t \boldsymbol{m} + \nabla\cdot(\boldsymbol{u}\otimes\boldsymbol{m}) + \nabla p - \nabla\cdot(\mathbb{G}(\nabla^s\boldsymbol{u}) + \boldsymbol{f}\otimes\boldsymbol{u}) = 0, \tag{13}$$

$$\partial_t E + \nabla\cdot(\boldsymbol{u}(E+p)) - \nabla\cdot(\boldsymbol{l} + \tfrac{1}{2}\boldsymbol{u}^2\boldsymbol{f} + \mathbb{G}(\nabla^s\boldsymbol{u})\cdot\boldsymbol{u}) = 0, \tag{14}$$

When looking at (12)–(14) it is not immediately clear how this system can be reconciled either with the Navier-Stokes regularization or with any phenomenological modeling of dissipation. It is remarkable that this exercise can actually been done by introducing the quantity $\boldsymbol{u}_m = \boldsymbol{u} - \rho^{-1}\boldsymbol{f}$. The above conservation equations then become

$$\partial_t \rho + \nabla\cdot(\boldsymbol{u}_m\rho) = 0, \tag{15}$$

$$\partial_t \boldsymbol{m} + \nabla\cdot(\boldsymbol{u}_m\otimes\boldsymbol{m}) + \nabla p - \nabla\cdot(\mathbb{G}(\nabla^s\boldsymbol{u})) = 0, \tag{16}$$

$$\partial_t E + \nabla\cdot(\boldsymbol{u}_m E) - \nabla\cdot(\boldsymbol{l} - e\boldsymbol{f}) + \nabla\cdot\big((p\mathbb{I} - \mathbb{G}(\nabla^s\boldsymbol{u}))\cdot\boldsymbol{u}\big) = 0. \tag{17}$$

This system resembles the Navier-Stokes regularization with two velocities. If one sets $a = d$, the term $\boldsymbol{l} - e\boldsymbol{f}$ becomes $d\rho\nabla e$, which upon assuming $\mathrm{d}e = c_v\,\mathrm{d}T$, reduces to $d(\rho,e)\rho c_v\nabla T$, i.e., one obtains Fourier's law: $\boldsymbol{l} - e\boldsymbol{f} = d(\rho,e)\rho c_v\nabla T$.

The system (15)–(17) resembles, at least formally, a model of fluid dynamics of [2] (see e.g., equations (1) to (5) in [2]). The author has derived the above system of conservation equations (up to some non-essential disagreement on the term $\boldsymbol{l} - e\boldsymbol{f}$) by invoking phenomenological considerations. The mathematical properties of this system have been investigated by [3]. Brenner has been defending for years the idea that it makes phenomenological sense to distinguish the so-called mass velocity, $\boldsymbol{u}_m$, from the so-called volume velocity, $\boldsymbol{u}$. This idea seems to be supported by our mathematical derivation which did not invoke any had oc phenomenological assumption. Recall that our primal motivation in this project is to find a regularization of the compressible Euler equations that can serve as a good numerical device, and by being good we mean that the model must give positive density, positive internal energy, a minimum entropy principle and be compatible with a large class of entropy inequalities.

4.3. **Concluding remarks.** Let us finally rephrase our findings. In its most general form, the regularized system (15)–(17) can be re-written as follows:

$$\partial_t \rho + \nabla\cdot(\boldsymbol{u}_m\rho) = 0, \tag{18}$$

$$\partial_t \boldsymbol{m} + \nabla\cdot(\boldsymbol{u}_m\otimes\boldsymbol{m}) + \nabla p - \nabla\cdot(G(\nabla^s\boldsymbol{u})) = 0, \tag{19}$$

$$\partial_t E + \nabla\cdot(\boldsymbol{u}_m E) - \nabla\cdot\boldsymbol{q} + \nabla\cdot\big((p\mathbb{I} - G(\nabla^s\boldsymbol{u}))\cdot\boldsymbol{u}\big) = 0 \tag{20}$$

$$\boldsymbol{u}_m = \boldsymbol{u} - a(\rho,e)\nabla\log\rho \tag{21}$$

$$\boldsymbol{q} = (a-d)p\nabla\log\rho + d\rho\nabla e, \qquad a(\rho,e)\geq 0,\ \ d(\rho,e)\geq 0. \tag{22}$$

It is established in Lemma 3.1 that the definition of $\boldsymbol{f} = a(\rho,e)\nabla\rho$ is compatible with the positive density principle. The particular form of $\boldsymbol{q}$ in (22) results from the definition of $\boldsymbol{l}$, see (7), which is required for the minimum entropy principle to hold, as established in Theorem 3.2. It is finally proved in Theorem 3.3 that the most robust regularization, i.e., that which is compatible with all the generalized entropy à la [8], corresponds to the choice $a = d$. A relaxation of the constraint $a = d$ is described in Corollary 1. As observed in §4.1, the parabolic regularization can be put into the form (18)–(22) with the particular choice $\mathbb{G} = a\nabla \boldsymbol{u}$, which is not rotation invariant and uses the same viscosity coefficient for all fields.

REFERENCES

[1] A. Bonito, J.-L. Guermond, and B. Popov. Stability analysis of explicit entropy viscosity methods for nonlinear scalar conservation equations. *Math. Comp.*, 2013.

[2] H. Brenner. Fluid mechanics revisited. *Physica A: Statistical Mechanics and its Applications*, 370(2):190 – 224, 2006.

[3] E. Feireisl and A. Vasseur. New perspectives in fluid dynamics: mathematical analysis of a model proposed by Howard Brenner. In *New directions in mathematical fluid mechanics*, Adv. Math. Fluid Mech., pages 153–179. Birkhäuser Verlag, Basel, 2010.

[4] J.-L. Guermond and R. Pasquetti. Entropy-based nonlinear viscosity for fourier approximations of conservation laws. *C. R. Math. Acad. Sci. Paris*, 346:801–806, 2008.

[5] J.-L. Guermond and B. Popov. Viscous regularization of the Euler equations and entropy principles. *SIAM, Journal on Applied Mathematics*, 2012. In review.

[6] J.-L. Guermond, R. Pasquetti, and B. Popov. Entropy viscosity method for nonlinear conservation laws. *Journal of Computational Physics*, 230:4248–4267, 2011.

[7] A. Harten. On the symmetric form of systems of conservation laws with entropy. *J. Comput. Phys.*, 49(1):151–164, 1983.

[8] A. Harten, P. D. Lax, C. D. Levermore, and W. J. Morokoff. Convex entropies and hyperbolicity for general Euler equations. *SIAM J. Numer. Anal.*, 35(6):2117–2127 (electronic), 1998.

[9] T. J. R. Hughes, L. P. Franca, and M. Mallet. A new finite element formulation for computational fluid dynamics. I. Symmetric forms of the compressible Euler and Navier-Stokes equations and the second law of thermodynamics. *Comput. Methods Appl. Mech. Engrg.*, 54 (2):223–234, 1986.

[10] M. Nazarov. Convergence of a residual based artificial viscosity finite element method. *Computers & Mathematics with Applications*, (0):–, 2012.

[11] G. Puppo. Numerical entropy production for central schemes. *SIAM Journal on Scientific Computing*, 25(4):1382–1415, 2003.

[12] D. Serre. *Systèmes de lois de conservation I: hyperbolicité, entropies, ondes de choc.* Diderot Editeur, Paris, 1996.

[13] D. Serre. Viscous system of conservation laws: singular limits. In *Nonlinear conservation laws and applications*, volume 153 of *IMA Vol. Math. Appl.*, pages 433–445. Springer, New York, 2011.

[14] E. Tadmor. A minimum entropy principle in the gas dynamics equations. *Appl. Numer. Math.*, 2(3-5):211–219, 1986.

[15] E. Tadmor. Entropy stability theory for difference approximations of nonlinear conservation laws and related time-dependent problems. *Acta Numer.*, 12:451–512, 2003.

[16] V. Zingan, J.-L. Guermond, J. Morel, and B. Popov. Implementation of the entropy viscosity method with the discontinuous galerkin method. *Computer Methods in Applied Mechanics and Engineering*, 253(0):479 – 490, 2013.

E-mail address: guermond@math.tamu.edu
E-mail address: popov@math.tamu.edu

DYNAMIC INSTABILITY OF THE VLASOV-POISSON-BOLTZMANN SYSTEM IN HIGH DIMENSIONS

Soohyun Bae

Faculty of Liberal Arts and Sciences
Hanbat National University
Daejeon 305-719, Korea

Sun-Ho Choi

Department of Mathematics
National University of Singapore
Singapore 117543, Singapore

Seung-Yeal Ha

Department of Mathematical Sciences
Seoul National University
Seoul 151-747, Korea

Abstract. We present the existence and dynamic instability of stationary radial solutions to the attractive Vlasov–Poisson–Boltzmann system. We show that all stationary radial solutions are local Maxwellians and for the instability of the stationary radial solution, we explicitly construct a one-parameter family of perturbed solutions via the Galilean boost method. Initially, these perturbed solutions can be close to the given stationary radial solution as much as possible in any L^p-norm, $p \in (\frac{n}{2}, \infty],\ n > 2$ where n is the spatial dimension. The perturbed solutions have the same local mass density profile as a stationary radial solution but a different bulk velocity profile. At the macroscopic level, these perturbations correspond to traveling waves.

1. **Introduction.** The purpose of this paper is to construct a radial stationary solution $f = f(x, v)$ to the attractive Vlasov–Poisson–Boltzmann system in the absence of external force and background density and to study its dynamic instability in terms of L^p-topology. Consider an ensemble of particles interacting through a self-consistent Newtonian attractive force and undergoing collisions between particles. In this situation, the kinetic description for the thermodynamic states of the ensemble is effectively described by the one-particle distribution function $f = f(x, v, t)$ at a phase position $(x, v) \in \mathbb{R}^n \times \mathbb{R}^n$ at time $t \in \mathbb{R}_+$. The dynamics of the distribution

2000 *Mathematics Subject Classification.* Primary: 35Q83; Secondary: 35B35, 35B40.

Key words and phrases. Vlasov–Poisson–Boltzmann, instability, radial solutions.

The work of S. Bae is supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012-0002056). The work of S.-Y. Ha is partially supported by a National Research Foundation of Korea Grant funded by the Korean Government(2009-0093137) and a research grant of the College of Natural Sciences in SNU. .

function f is governed by the self-consistent Vlasov–Poisson–Boltzmann (V-P-B) system:

$$\begin{aligned} &\partial_t f + v\cdot\nabla_x f - \nabla_x\varphi\cdot\nabla_v f = Q(f,f), \quad x, v\in\mathbb{R}^n,\ t>0,\\ &\Delta\varphi = \rho, \quad \rho = \int_{\mathbb{R}^n} f dv, \end{aligned} \tag{1}$$

where φ is the self-consistent force potential, and $Q(f,f)$ is the collision operator registering binary collisions between particles. Its explicit form reads as

$$Q(f,f)(x,v,t) \equiv \frac{1}{Kn}\int_{\mathbb{R}^n\times\mathbf{S}_+^{n-1}} B(|v-v_*|,\theta)(f(v')f(v_*') - f(v)f(v_*))dv_*d\omega. \tag{2}$$

Here v' and v_*' denote post-collided velocities resulting from the pre-collided velocities v, v_*:

$$v' = v - [(v-v_*)\cdot\omega]\omega, \quad v_*' = v_* + [(v-v_*)\cdot\omega]\omega, \qquad \omega\in\mathbb{S}_+^{n-2}. \tag{3}$$

Moreover, we used the simplified notation:

$$f(v)\equiv f(x,v,t), \quad f(v_*)\equiv f(x,v_*,t), \quad f(v')\equiv f(x,v',t)$$
$$\text{and} \quad f(v_*')\equiv f(x,v_*',t).$$

In the formal infinite Knudsen limit(Kn → ∞), system (1) becomes the collisionless Vlasov–Poisson(V-P) system with attractive force:

$$\begin{aligned} &\partial_t f + v\cdot\nabla_x f - \nabla_x\varphi\cdot\nabla_v f = 0, \quad x, v\in\mathbb{R}^n,\ t>0,\\ &\Delta\varphi = \rho, \quad \rho = \int_{\mathbb{R}^n} f dv. \end{aligned} \tag{4}$$

The existence theory for (1) has been studied in two distinct regimes: near Maxwellian and near vacuum. In [7, 14, 15, 22, 23], the global existence and time-asymptotic behavior of solutions have been studied in the near-Maxwellian regime. In contrast, there have been few near-vacuum results [16] available for the soft and Maxwellian potentials in the framework of Bardos and Degond [2]. For other related works for the V-P-B system, we refer to [4, 8, 21, 24, 25]

In this paper, we are interested in whether the solution to the V-P-B system is $L^p(\mathbb{R}^{2n})$-stable or not, in particular for three dimensions $n=3$ and $p>3/2$. When the electric field $E=-\nabla_x\varphi$ is turned off, the V-P-B system becomes the Boltzmann equation. In this case, it is well-known [1, 17, 18] that the Boltzmann equation near vacuum is uniformly L^1-stable in the sense that

$$\sup_{t\geq 0}||f(t)-g(t)||_{L^1(\mathbb{R}^6)} \leq G||f^{in}-g^{in}||_{L^1(\mathbb{R}^6)},$$

where f and g are continuous mild solutions to the Boltzmann equation corresponding to small initial data f^{in} and g^{in} respectively, and $G(\geq 1)$ is a generic positive constant independent of t. In contrast, when the electric field is turned on, there is no such $L^p(\mathbb{R}^6)$-stability result yet even for an interesting spatial dimension $n=3$. Recently the authors [6] showed that nonlinear Vlasov equations with attractive forces such as the attractive V-P system are $L^p(\mathbb{R}^6)$-unstable by constructing nonlinear perturbations of a stationary solution. Of course, aforementioned instability result does not exclude the possibility of uniform $L^p(\mathbb{R}^6)$-stability for small solutions. In the absence of collisions, the V-P system is uniformly $L^1(\mathbb{R}^{2n})$-stable for small and decaying solutions in high dimensions $n\geq 4$ [3] (see the corresponding result for the Vlasov-Yukawa system [5, 19]). Hence the stability issue of the V-P-B system might be very sensitive to the size of solutions and dimensions of the spatial domain.

We next briefly delineate our strategy and results for the instability of system (1). Our scenario is to show that the regular stationary radial solutions are L^p-unstable for some $p \in (n/2, \infty]$. To launch our scenario, we first need to have stationary radial solutions. This is why we restrict our force to be attractive. It turns out that stationary radial solutions are local Maxwellian type and the logarithm of the corresponding local mass density satisfies a nonlinear second-order ordinary differential equation (ODE). We then derive an explicit asymptotic estimate by using a phase-portrait method. The detailed asymptotic behavior of entire stationary radial solution $f_0 = f_0(|x|, |v|)$ implies that

$$f_0 \in L^p(\mathbb{R}^{2n}), \quad p \in \Big(\frac{n}{2}, \infty\Big].$$

For the instability estimate of the stationary radial solutions, we use the method of Galilean boost introduced in [6]. Due to the Galilean boost invariance of the V-P-B system, the Galilean boost of the given stationary radial solution plays the role of unstable perturbed solution. Hence we can conclude that the V-P-B system (1) is $L^p(\mathbb{R}^{2n})$-unstable, $p > n/2$ for general initial data. More precisely, we state our main result in the following theorem.

Theorem 1.1. *Let $f_0 = f_0(x, v)$ be an entire stationary radial solution to* (1). *Then, for any $\varepsilon > 0$, there exists a perturbation f^{in} of f_0 and $T > 0$ such that, for any solution $f = f(t)$ with initial datum f^{in}, we have*

$$||f^{in} - f_0||_{L^p} < \varepsilon \quad \textit{and} \quad ||f(t) - f_0||_{L^p} \geq ||f_0||_{L^p}, \quad t \geq T = T(\varepsilon),\ p \in \Big(\frac{n}{2}, \infty\Big).$$

Remark 1. 1. For the existence of stationary radial solution, the sign of the force (attractive force) is crucial, because for the repulsive case, the stationary radial solutions do not exist unless there are external confining forces. Therefore the method of Galilean boost cannot be used in this case. It is still an interesting open problem whether the V-P-B system with attractive forces is L^p-stable or not for small solutions.
2. Similar instability result was also true for the V-P system with attractive force (see [6]). In this case the instability result is valid for $p \in [1, \infty]$. The dynamic instabilities of the V-P-B and V-P systems are essentially due to the *nonlinear interactions* between particles and field. For the Boltzmann equation with a linear external force, the uniform L^p-stability is valid as the Boltzmann equation (see [11, 12]).

The rest of the paper is divided into five sections. In Section 2, we study the structure of stationary radial solutions to the V-P-B system and derive an elliptic equation related to the stationary radial solution. In Section 3, we derive the asymptotic rate of solution to the elliptic equation. Additionally, we obtain the $L^p(\mathbb{R}^{2n})$-regularity of the stationary radial solution to the V-P-B system by this asymptotic rate. In Section 4, we construct a one-parameter family of stationary radial solutions via the method of Galilean boost, and prove that the stationary radial solution is dynamically L^p-unstable by the one-parameter family. Finally Section 5 is devoted to the summary of main results. In Appendix A, we present the long straightforward proof of Proposition 2.

Notation. Throughout this paper, we will use simplified notation as follows:

$$\begin{aligned} ||f||_{L^p} &:= \left(\int_{\mathbb{R}^3}\int_{\mathbb{R}^n} |f(x,v)|^p dvdx\right)^{1/p}, \quad 1 \le p < \infty, \\ E[f](x,t) &:= -\nabla_x \varphi(x,t) = -\nabla_x\left(\frac{1}{|x|^{n-2}} *_x \rho(x,t)\right), \quad n > 2, \end{aligned}$$

where a measurable function $f = f(x,v,t)$ with $\rho(x,t) = \int_{\mathbb{R}^n} f(x,v)dv$.

2. **Preliminaries.**

In this section, we briefly review the previous results on the existence and stability of the V-P-B system, and present the structural results on stationary radial solutions to the V-P-B system (1).

2.1. **Brief review of previous results.**

For the existence of stationary solutions to the V-P-B system, several authors [9, 10] have been interested in the relationship between the external force or background density ρ_0 and the stationary solution $f(x,v)$. In particular, Duan, Yang, and Zhu [10] verified that there are stationary solutions to the V-P-B system, when the background density ρ_0 tends to a positive constant as $|x| \to \infty$, i.e.,

$$|\rho_0 - 1| \le C(\ln(e + |x|))^{-\alpha},$$

where C and α are some positive constants. Then, the V-P-B system has the stationary solution

$$f(x,v) = \frac{1}{(2\pi)^{3/2}} \exp\left(\phi(x) - \frac{|v|^2}{2}\right)$$

with

$$-\Delta\phi + \exp(\phi) = \rho_0.$$

In [9], Duan and Yang also obtained stationary solutions to the V-P-B system, when the background density is a small perturbation of a positive constant in the following sense:

$$||\overline{\rho} - 1||_{W_k^{m,\infty}} = \sup_{x\in\mathbb{R}^3} (1 + |x|)^k \sum_{|\alpha|\le m} |\partial_x^\alpha(\overline{\rho} - 1)|,$$

where $k, m \ge 0$ are integers.

For the stability of the stationary solution to the V-P-B system, previous authors [9, 23] have considered the convergence of a perturbed near-Maxwellian solution to the global Maxwellian or stability in some functional sense. Yang, Yu, and Zhao [23] proved that for a smooth near-Maxwellian perturbation, there is a unique global classical solution to the V-P-B system with a background charge density, when either the mean free path is small or the background charge density is large. Furthermore, this solution converges to the global Maxwellian, as time goes to infinity in the following sense:

$$\lim_{t\to\infty} \sup_{x\in\mathbb{R}^3} \sum_{|\alpha|\le N-4} \int_{\mathbb{R}^3} \frac{|\partial_x^\alpha(f(t,x,\xi) - \overline{M}(\xi))|^2}{M_-(\xi)} d\xi = 0,$$

where $\overline{M}$ and M_- are suitably chosen Maxwellians. Duan and Yang [9] assume that $||\overline{\rho} - 1||_{W_2^{N+1,\infty}}$ and $[[u_0]]$ are small enough, where

$$f_0(x,\xi) = e^{\phi}M + \sqrt{M}u_0(x,\xi) \ge 0.$$

Then, the nonlinear stability of solutions near the stationary state holds in the following sense:

$$[[u(t)]]^2 + \lambda_0 \int_0^t [[u(t)]]_\nu^2 ds \le C_0 [[u_0]]^2 m$$

where $f(t,x,\xi) = e^{\phi} M + \sqrt{M} u(t,x,\xi) \ge 0$ is a solution to the V-P-B system and

$$[[u(t)]]_\nu^2 := \sum_{|\alpha|+|\beta|\le N} ||\partial_x^\alpha \partial_\xi^\beta \{\mathbf{I}-\mathbf{P}\}u||_\nu^2 + \sum_{|\alpha|\le N-1} (||\partial_x^\alpha \nabla_x (a,b,x)||^2 + ||\partial_x^\alpha (a+3c)||^2),$$

where $\mathbf{L}$ is a linearized collision operator, $\mathbf{P}$ denotes the projection operator from $L^2(R_\xi^3)$ to ker$\mathbf{L}$, and a, b, and c are the coefficients of the macroscopic component $\mathbf{P}u$.

2.2. **Structure of stationary radial solutions.** Let f_0 be a stationary radial solution, i.e.,

$$f_0(x,v) = \bar{f}_0(r,p), \quad r := |x|, \; p := |v|. \tag{1}$$

For any $v \in \mathbb{R}^n$, we set

$$v^\perp := \{y \in \mathbb{R}^n \; : \; y \cdot v = 0\}.$$

Lemma 2.1. *Let G and H be two measurable functions defined on $\mathbb{R}^2$ and $\mathbb{R} \times \mathbb{R}^n$, respectively and they satisfy relation:*

$$\frac{x \cdot v}{|x|} G(|x|,|v|) = H(|x|,v), \quad a.e. \quad x,v \in \mathbb{R}^n. \tag{2}$$

Then the function H is 0 a.e. $(x,v) \in \mathbb{R}^{2n}$.

Proof. It suffices to consider the case where relation (2) holds for all $x,v \in \mathbb{R}^n$. Let (x,v) be any point in $\mathbb{R}^{2n}$. Then we can choose $\bar{x} \in v^\perp$ with $|x| = |\bar{x}|$. For such $\bar{x}$, we have

$$H(|x|,v) = H(|\bar{x}|,v) = \frac{\bar{x} \cdot v}{|x|} G(|x|,|v|) = 0.$$

Hence $H \equiv 0$. □

Lemma 2.2. *Let f_0 be the stationary radial solution to the V-P-B system (1). Then it is also the stationary radial solution to the V-P system (4).*

Proof. Since f_0 is stationary, it satisfies

$$v \cdot \nabla_x f_0 - \nabla_x \varphi \cdot \nabla_v f_0 = Q(f_0,f_0). \tag{3}$$

We now rewrite relation (3) in terms of $\bar{f}_0$ in (1) to derive relation (2).

- (L.H.S. of (3)): In this case, we use

$$\nabla_x \varphi = \frac{x}{r} \partial_r \varphi = \alpha_n \frac{x}{r^n} \int_0^r r_*^{n-1} \rho(r_*) dr_* \tag{4}$$

to obtain

$$\begin{aligned} v \cdot \nabla_x f_0 - \nabla_x \varphi \cdot \nabla_v f_0 &= \frac{x \cdot v}{r} \partial_r \bar{f}_0 - \frac{\nabla_x \varphi \cdot v}{p} \partial_p \bar{f}_0 \\ &= \frac{x \cdot v}{r} \Big(\partial_r \bar{f}_0 - \alpha_n \frac{\partial_p \bar{f}_0}{r^{n-1} p} \int_0^r r_*^{n-1} \rho(r_*) dr_* \Big), \end{aligned} \tag{5}$$

where constant α_n is the volume of an n-dimensional unit sphere.

- (R.H.S. of (3)): We use representation (2) for $Q(f,f)$:

$$\begin{aligned}
&Q(f_0,f_0)(x,v)\\
=&\iint_{S^2_+\times\mathbb{R}^n} B(v-v_*,\omega)(f_0(x,v'_*)f_0(x,v')-f_0(x,v)f_0(x,v_*))dv_*d\omega\\
=&\iint_{S^2_+\times\mathbb{R}^n} B(v-v_*,\omega)(\bar f_0(|x|,|v'_*|)\bar f_0(|x|,|v'|)-\bar f_0(|x|,|v|)\bar f_0(|x|,|v_*|))dv_*d\omega\\
=&B(\bar f_0)(|x|,v).
\end{aligned}\tag{6}$$

Hence we have

$$\frac{x\cdot v}{r}\underbrace{\left(\partial_r\bar f_0-\alpha_n\frac{\partial_p\bar f_0}{r^{n-1}p}\int_0^r r_*^{n-1}\rho(r_*)dr_*\right)}_{\text{function of }|x|\text{ and }|v|}=B(\bar f_0)(|x|,v).$$

We now apply Lemma 2.1 to get

$$B=0,\quad\text{or equivalently}\quad Q(f_0,f_0)=0\quad\text{a.e. }(x,v)\in\mathbb{R}^{2n},$$

i.e., f_0 satisfies

$$v\cdot\nabla_x f_0-\nabla_x\varphi\cdot\nabla_v f_0=0,\quad\text{a.e. }(x,v)\in\mathbb{R}^{2n}.$$

□

Proposition 1. *Let f_0 be a stationary radial solution to the V-P-B system (1). Then f_0 is separable in r and p; more precisely, we have*

$$f_0(x,v)=R(r)e^{-c_0p^2},\qquad a.e.\ (x,v)\in\mathbb{R}^{2n},$$

where c_0 is a positive constant.

Proof. Let f_0 be a stationary radial solution. Then it follows from Lemma 2.2 that

$$Q(f_0,f_0)(x,v)=0\quad\text{a.e. }(x,v)\in\mathbb{R}^{2n}.$$

Then it is well known [13] that f is a local maxwellian:

$$f_0(x,v)=e^{a(r)+b(r)\cdot v-c(r)|v|^2},$$

where $a,c:\mathbb{R}\longrightarrow\mathbb{R}$, $b:\mathbb{R}\to\mathbb{R}^n$. Since f is radial,

$$b\equiv 0.$$

Hence f becomes

$$f_0(x,v)=e^{a(r)-c(r)|v|^2}=:R(r)e^{-c(r)|v|^2},\quad c(r)>0.$$

We next claim that

$$c(r)=c_0\ :\ \text{constant}.$$

The proof of the claim: By elementary calculations,

$$\begin{aligned}
0&=v\cdot\nabla_x f_0-\nabla_x\varphi\cdot\nabla_v f_0\\
&=v\cdot\nabla_x(Re^{-c(r)|v|^2})-\nabla_x\varphi\cdot\nabla_v(Re^{-c(r)|v|^2})\\
&=v\cdot(\nabla_x R)e^{-c(r)|v|^2}+v\cdot Re^{-c(r)|v|^2}(-\nabla_x c(r)|v|^2)\\
&\quad-\nabla_x\varphi\cdot Re^{-c(r)|v|^2}(-\nabla_v(c(r)|v|^2)).
\end{aligned}$$

We divide the above relation by $e^{-c(r)|v|^2}$ and use $\nabla_v |v|^2 = 2v$ to find

$$\begin{aligned} 0 &= v\cdot\nabla_x R - R|v|^2 v\cdot\nabla_x c(r) + 2Rc(r)\nabla_x\varphi\cdot v \\ &= (v\cdot x)\Big[\frac{R'(r)}{r} - \frac{R(r)c'(r)}{r}|v|^2 + \frac{2R(r)c(r)}{r^n}\Big(\int_0^r \rho(r_*) r_*^{n-1}dr_*\Big)\Big], \end{aligned} \tag{7}$$

where we used (4). For the case $x \notin v^\perp$, it follows from (7) that

$$\frac{R(r)c'(r)}{r}|v|^2 = \frac{R'(r)}{r} - \frac{2R(r)c(r)}{r^n}\int_0^r \rho(r_*) r_*^{n-1}dr_*.$$

Since the R.H.S. of the above relation is a function of r, we have

$$\frac{R(r)c'(r)}{r}|v|^2 = 0, \quad \text{a.e.,} \qquad \text{i.e.} \quad c'(r) = 0, \quad \text{a.e. } x\in\mathbb{R}^n.$$

This completes the proof of the claim. □

Remark 2. Note that Proposition 1 says that all stationary radial solutions to the V-P-B system (1) are local maxwellians.

2.3. **Derivation of the equation for R.** In this part, we derive the equation for R. Let f_0 be a stationary radial solution to the V-P-B system. Then it follows from Proposition 1 that

$$f_0(x,v) = R(r)e^{-c_0|v|^2}, \quad \text{for suitable } R(r). \tag{8}$$

We next derive a defining equation for R. We substitute ansatz (8) to system (1) to get

$$v\cdot\nabla_x(R(r)e^{-c_0|v|^2}) + E[f]\cdot\nabla_v(R(r)e^{-c_0|v|^2}) = 0. \tag{9}$$

If we divide Equation (9) by $e^{-c_0|v|^2}$, we have

$$v\cdot\Big(\nabla_x R(r) - 2c_0 E[f]R(r)\Big) = 0. \tag{10}$$

Since $\nabla_x R(r) - 2c_0E[f]R(r)$ is independent of v in (10), we conclude that

$$\nabla_x R(r) - 2c_0 E[f]R(r) = 0. \tag{11}$$

However, note that

$$\rho = \int_{\mathbb{R}^n} R(r)e^{-c_0|v|^2}dv = R(r)\int_{\mathbb{R}^n} e^{-c_0|v|^2}dv = \Big(\frac{\pi}{c_0}\Big)^{\frac{n}{2}} R(r).$$

Since we are looking for a nontrivial solution, Equation (11) yields

$$\frac{\nabla_x R(r)}{R(r)} = 2c_0E[f] \quad \text{equivalently} \quad \nabla_x \log R(r) = 2c_0 E[f]. \tag{12}$$

We take the divergence in Equation (12) to find

$$\Delta \log R(r) = 2c_0\nabla\cdot E[f] = 2c_0\nabla\cdot(-\nabla_x\varphi) = -2c_0\rho = -2c_0\Big(\frac{\pi}{c_0}\Big)^{\frac{n}{2}}R(r).$$

We now set

$$u := \log R, \quad \text{i.e.} \quad R = e^u.$$

Then the function u satisfies

$$\Delta u + \kappa_n e^u = 0, \tag{13}$$

where κ_n is a positive constant defined by

$$\kappa_n := 2c_0\Big(\frac{\pi}{c_0}\Big)^{\frac{n}{2}}.$$

Once we solve the above elliptic equation, we can find a stationary radial solution. In the following section, we study the existence and asymptotic behavior of global radial solutions to the elliptic equation (13) at $r = \infty$.

3. Asymptotic behavior of regular solutions to $\Delta u + \kappa_n e^u = 0$.

In this section, we study the asymptotic behavior of entire radial solutions to the following n-dimensional elliptic problem:

$$\begin{aligned} &\Delta u + \kappa_n e^u = 0, \qquad x \in \mathbb{R}^n, \\ &(u, \nabla u)(x) = (\underline{u}, 0), \quad x = 0, \end{aligned} \tag{1}$$

where $\underline{u}$ is a constant. i.e., radial solution $u = u(r)$ satisfies

$$\begin{aligned} u'' + \frac{(n-1)u'}{r} + \kappa_n e^u = 0, \quad r > 0, \\ (u(0), u'(0)) = (\underline{u}, 0), \quad r = 0. \end{aligned} \tag{2}$$

The asymptotic behavior for (2) was studied in Joseph and Lundgren's classical work (see Lemma 7, page 265 [20]). However their motivation to study (2) is different from ours. Joseph and Lundgren focused on Equation (2) with two-point boundary conditions at $r = 0, 1$ and replaced κ_n by λ. They analyzed the multiplicity of radial solutions depending on the parameter λ. However, our main interest in Equation (2) is the asymptotic behavior of the general solutions equipped with initial data at $r = 0$. Of course, some brief and compact arguments in Section 3.2 can be found in [20]. For convenience, we derive an alternative, direct proof of the asymptotic behavior to (2) in Section 3.2.

3.1. Existence of entire regular solutions.

In this part, we briefly sketch the local and global existence of regular solutions to Equation (1).

For a proof of local existence, we multiply each side of Equation (2) by r^{n-1} and then integrate to find

$$t^{n-1}u'(t) = t_*^{n-1}u'(t_*) - \kappa_n \int_{t_*}^t \tau^{n-1} e^{u(\tau)} d\tau.$$

We divide this equation by r^{n-1} and integrate one more time, then we have

$$u(r) = u(r_*) + \int_{r_*}^r \frac{t_*^{n-1}}{t^{n-1}} u'(t_*) dt - \kappa_n \int_{r^*}^r \frac{1}{t^{n-1}} \int_{t_*}^t \tau^{n-1} e^{u(\tau)} d\tau dt. \tag{3}$$

Since we have the boundary condition $(u(0), u'(0)) = (\underline{u}, 0)$, we can apply standard contraction mapping and fixed-point arguments to Equation (3) with the following function-valued operator $\mathfrak{F}$:

$$\mathfrak{F}(u)(r) := \underline{u} - \kappa_n \int_0^r \frac{1}{t^{n-1}} \int_0^t \tau^{n-1} e^{u(\tau)} d\tau dt.$$

to derive a local existence of the $\mathcal{C}^2$ solution to (2). We omit the detailed straightforward arguments.

On the other hand, for a global existence, it suffices to show that u' is uniformly bounded by the continuation principle. For this, we multiply Equation (2) by u' to get

$$\begin{aligned} \frac{1}{2}|u'(r)|^2 + (n-1)\int_0^r \frac{|u'(\zeta)|^2}{\zeta} d\zeta &= -\kappa_n \int_0^r u'(\zeta) e^{u(\zeta)} d\zeta = -\int_{\underline{u}}^{u(r)} e^{\zeta} d\zeta \\ &= e^{u(0)} - e^{u(r)} < e^{u(0)}. \end{aligned}$$

Therefore, a priori we can obtain a uniform bound of $|u'(r)|$ and we also have

$$|u(r)| \leq |\underline{u}| + \int_0^r u'(t)dt \leq |\underline{u}| + e^{\underline{u}}r.$$

Hence the local solution can be continued to an arbitrary interval to yield a global solution.

Remark 3. (i) By direct calculation, it is easy to see that Equation (2) has a scale invariance, i.e., for any solution $u = u(r)$ to (2) with initial data $(u(0), u'(0)) = (\underline{u}, 0)$ and for any real number $\alpha \in \mathbb{R}$,

$$u_\alpha(r) = \alpha + u(e^{\frac{\alpha}{2}}r)$$

is also a solution to the ODE (2) with translated initial data $(u(0), u'(0)) = (\underline{u} + \alpha, 0)$. This property might suggest the existence of singular solutions with a scale invariance, which are given by

$$u_s(r; n) := -2\log r - \log \kappa_n + \log 2(n-2) = \log \frac{2(n-2)}{\kappa_n r^2}, \quad n > 2. \tag{4}$$

Note that the $u_s(r; n)$ is in fact a singular solution to the ODE (2) satisfying

$$u'' + \frac{(n-1)u'}{r} + \kappa_n e^u = 0, \qquad \lim_{r\to 0+} u(r) = \infty.$$

(iii) For low dimensions $n = 1, 2$, Equation (2) with $u(0) = -\log \kappa_n$ has explicit solutions:

$$u(r) = \begin{cases} -2\log\left(\cosh \dfrac{r}{\sqrt{2}}\right) - \log \kappa_n, & n = 1, \\ -2\log\left(1 + \dfrac{r^2}{8}\right) - \log \kappa_n, & n = 2. \end{cases}$$

3.2. Asymptotic behavior of entire regular solutions.

In this part, we present the asymptotic behavior of an entire regular solution $u = u(r)$ at $r = \infty$ for high dimension $n > 2$.

Note that it follows from Proposition 1 that the stationary global radial solution to the stationary V-P-B system is of the form $f(x, v) = \exp(u(r))\mu(p)$, where $\mu(p)$ is a Maxwellian. Therefore if we want to check whether the regular radial solution f to the stationary V-P-B system belongs to $L^p(\mathbb{R}^{2n})$ for some $p \geq 1$, we need to investigate the asymptotic behaviors of the $u(r)$ as $r \to \infty$. We first state the main theorem of this section and postpone its proof to the latter part of this section.

Theorem 3.1. (Asymptotic behavior at infinity) *For $n > 2$, let $u = u(r)$ be a global regular solution to the ODE* (2). *Then for any initial value $\underline{u}$, $u = u(r)$ approaches the singular solution* (4) *as $r \to \infty$, i.e.,*

$$\lim_{r\to\infty} |u(r) - u_s(r; n)| = 0.$$

Proof. We postpone the proof to the last part of this subsection. □

Remark 4. Theorem 3.1 says that the entire regular solution to (2) behaves like the explicit singular solution $u(r; n)$ whose explicit form is given by the formula (4) as $r \to \infty$.

To avoid r-dependent coefficient in (2), we introduce a new independent variable t:

$$t = \log r, \quad \text{or} \quad e^t = r,$$

and set

$$V(t) := u(r) - u_s(r;n) = u(e^t) + 2t + \log \kappa_n - \log 2(n-2).$$

We next derive an ODE for V. By direct calculation, we have

$$V'(t) = e^t u'(e^t) + 2, \quad V''(t) = e^t u'(e^t) + e^{2t} u''(e^t), \tag{5}$$

and

$$\begin{aligned} V'' + (n-2)V' &= e^t u'(e^t) + e^{2t} u''(e^t) + (n-2)e^t u'(e^t) + 2(n-2) \\ &= e^{2t} u''(e^t) + (n-1)e^t u'(e^t) + 2(n-2) \\ &= e^{2t}\left[u''(e^t) + \frac{(n-1)}{e^t} u'(e^t)\right] + 2(n-2) \\ &= -\kappa_n e^{u(e^t)+2t} + 2(n-2) \\ &= -2(n-2)e^{u(e^t)+2t+\log\kappa_n - \log 2(n-2)} + 2(n-2) \\ &= 2(n-2)(1-e^V), \end{aligned}$$

Hence we obtain the autonomous ODE for $V(t)$:

$$\begin{aligned} &V'' + (n-2)V' = 2(n-2)(1-e^V), \quad t \in \mathbb{R}, \\ &\lim_{t\to-\infty} V(t) = -\infty, \quad \lim_{t\to-\infty} V'(t) = 2. \end{aligned} \tag{6}$$

Equation (6) was also obtained in Joseph and Lundgren's work [20] (see equation XI.1, page 262) by two successive changes of variables. We also rewrite Equation (2) as a system of first-order ODEs:

$$\begin{aligned} &V' = Q, \quad t \in \mathbb{R}, \\ &Q' = -(n-2)Q + 2(n-2)(1-e^V). \end{aligned} \tag{7}$$

Note that this two-dimensional system (7) has the unique equilibrium solution

$$(V, Q) = (0, 0).$$

Recall that our purpose is to verify

$$\lim_{t\to\infty} V(t) = 0.$$

3.2.1. *Preparatory lemmas.* Below, we present three preparatory lemmas for the proof of Theorem 3.1.

Lemma 3.2. *Let $u = u(r)$ and $u_s(r;n)$ be a general regular solution to* (2) *and the singular solution defined by* (4)*, respectively. Then $Q = Q(t)$ satisfies*

$$Q(t) < 2, \quad t \in \mathbb{R}.$$

Proof. Note that $u = u(r)$ satisfies

$$(r^{n-1}u'(r))' + r^{n-1}\kappa_n e^{u(r)} = 0, \quad r \in \mathbb{R}_+.$$

We integrate this equation with respect to r to get

$$u'(r) = -\frac{1}{r^{n-1}} \int_0^r \tau^{n-1}\kappa_n e^{u(\tau)} d\tau.$$

Therefore, we have

$$u'(r) < 0 \quad r \in \mathbb{R}_+.$$

Relation (5) and the above relation imply

$$V'(t) = e^t u'(e^t) + 2 < 2, \quad t \in \mathbb{R}.$$

□

Lemma 3.3. *Let V be the solution to the ODE (6). Then $(V, Q = V')$ satisfies*

$$Q(t^*)e^{(n-2)t^*} - Q(t_*)e^{(n-2)t_*} = 2(n-2)\int_{t_*}^{t^*}(1-e^V)e^{(n-2)t}dt, \tag{8}$$

$$\begin{aligned}&\frac{1}{2}\Big(Q(t^*)^2 - Q(t_*)^2\Big)\\ &\qquad = 2(n-2)\int_{V(t_*)}^{V(t^*)}(1-e^V)dV - (n-2)\int_{t_*}^{t^*}Q^2(t)dt,\end{aligned} \tag{9}$$

where t_* and t^* are extended real numbers in $\mathbb{R}\cup\{\pm\infty\}$.

Proof. Note that Equation (6) can be rewritten as

$$Q' + (n-2)Q = 2(n-2)(1-e^V). \tag{10}$$

(i) We multiply (10) by the integration factor $e^{(n-2)t}$ to find

$$\Big(e^{(n-2)t}Q\Big)' = 2(n-2)(1-e^V)e^{(n-2)t}.$$

We now integrate the above relation from $t = t_*$ to $t = t^*$ to get the desired result.

(ii) We multiply Equation (10) by Q to find

$$\left(\frac{Q^2}{2}\right)' + (n-2)Q^2 = 2(n-2)Q(1-e^V).$$

We integrate the above equation in t to obtain the desired result. □

We set $\mathcal{H}$ by the half-space in (V, Q) space:

$$\begin{aligned}\mathcal{H} \quad &:= \{(V,Q)\in\mathbb{R}^2 \ : \ -\infty\le V\le\infty,\ Q<2\}\\ &= \mathcal{H}^+\cup\mathcal{H}^0\cup\mathcal{H}^-,\\ \mathcal{H}^+ \quad &:= \{(V,Q)\in\mathbb{R}^2 \ : \ -\infty\le V\le\infty,\ 0<Q<2\},\\ \mathcal{H}^0 \quad &:= \{(V,Q)\in\mathbb{R}^2 \ : \ -\infty\le V\le\infty,\ Q=0\},\\ \mathcal{H}^- \quad &:= \{(V,Q)\in\mathbb{R}^2 \ : \ -\infty\le V\le\infty,\ Q<0\}.\end{aligned} \tag{11}$$

Lemma 3.4. *Let $(V(t), Q(t))$ be the solution to system (7), and assume that*

$$\lim_{t\to\infty}V(t) = V^\infty, \quad \lim_{t\to\infty}Q(t) = Q^\infty.$$

Then we have

$$(V^\infty, Q^\infty)\in\mathcal{H}^0.$$

Proof. It follows from Lemma 3.2 that $(V^\infty, Q^\infty)\in\mathcal{H}$. Hence it suffices to show that

$$(V^\infty, Q^\infty)\notin\mathcal{H}^+\cup\mathcal{H}^-.$$

Suppose not, i.e.,

$$(V^\infty, Q^\infty)\in\mathcal{H}^+\cup\mathcal{H}^-.$$

This implies

$$|Q^\infty| > 0. \tag{12}$$

We now take $t_* = 0$ and $t^* = \infty$ in Equation (9) of Lemma 3.3 to get

$$\frac{1}{2}\Big[(Q^\infty)^2 - Q(0)^2\Big] = 2(n-2)\int_{V(0)}^{V^\infty}(1-e^V)dV - (n-2)\int_0^\infty Q^2(t)dt. \tag{13}$$

Then it is easy to see that

$$\left|\frac{1}{2}\Big[(Q^\infty)^2 - Q(0)^2\Big]\right| < \infty, \quad 2(n-2)\left|\int_{V(0)}^{V^\infty} (1-e^V)dV\right| < \infty.$$

However, (12) yields

$$\left|(n-2)\int_0^\infty Q^2(t)dt\right| = \infty.$$

Hence we have a contradiction in (13). Therefore if the trajectory $(V(t), Q(t))$ converges, then the limit point should be on the horizontal line $\mathcal{H}^0$. □

Recall that (V, Q) satisfies

$$\begin{aligned} V' &= Q, \quad t \in \mathbb{R}, \\ Q' &= -(n-2)Q + 2(n-2)(1-e^V). \end{aligned}$$

Note that as long as, $Q > 0$, V is strictly increasing, and, in contrast, as long as $Q < 0$, V is strictly decreasing. Let $(V(t), Q(t))$ be the phase trajectory of the dynamics (7) satisfying

$$V(-\infty) = -\infty \quad \text{and} \quad Q(-\infty) = 2.$$

Then there are several possibilities for the asymptotic behavior of (V, Q) as $t \to \infty$.

Proposition 2. *Let (V, Q) be the global solution to system (7). Then there are two different asymptotic patterns depending on the spatial dimension n.*

1. *For $2 < n < 10$, the trajectory $(V(t), Q(t))$ swirls around the unique equilibrium point $(0, 0)$ as $t \to \infty$.*
2. *For $n \geq 10$, the trajectory $(V(t), Q(t))$ goes to the equilibrium point $(0, 0)$ directly without touching the line $V = 0$ as $t \to \infty$.*

Proof. Since the proof is rather long and tedious to follow, we leave its proof in Appendix A. □

3.2.2. *The proof of Theorem 3.1.* Recall that it suffices to check that for any global solution (V, Q) to the system (7)

$$\lim_{t\to\infty} (V(t), Q(t)) = (0, 0).$$

• Case A ($2 < n < 10$): It follows from Proposition 2 that the trajectory of (V, Q) is a spiral. Therefore we can assume that (V, Q) is confined to the bounded set after $t \geq T$. Then it follows from the uniqueness of two-dimensional flows that the flow (V, Q) converges to either a limit cycle or an equilibrium point. Once we can show that the trajectory cannot approach some limit cycle, then we are done. Suppose this trajectory converges to the limit cycle. Then we can choose four sequences $\{T_i\}$,$\{t_i\}$,$\{S_i\}$ and $\{s_i\}$ satisfying

$$\begin{aligned} &Q(T_i) = Q(t_i) = 0, \quad V(T_i) > 0, \quad V(t_i) < 0 \quad \text{and} \\ &\qquad -\infty < T_1 < t_1 < T_2 < t_2 \cdots < T_i < t_i < \cdots < \infty, \\ &V(S_i) = V(s_i) = 0, \quad Q(S_i) > 0, \quad Q(s_i) < 0 \quad \text{and} \\ &\qquad -\infty < S_1 < s_1 < S_2 < s_2 \cdots < T_i < t_i < \cdots < \infty. \end{aligned}$$

Then by the identity (9) in Lemma 3.3, we have

$$\begin{aligned}
0 &= \lim_{i\to\infty} \frac{1}{2}(Q^2(T_{i+1}) - Q^2(T_i)) \\
&= \lim_{i\to\infty} 2(n-2)\int_{V(T_i)}^{V(T_{i+1})} (1-e^V)dV - \lim_{i\to\infty}(n-2)\int_{T_i}^{T_{i+1}} Q^2 dt \\
&= -\lim_{i\to\infty}(n-2)\int_{T_i}^{T_{i+1}} Q^2 dt \\
&= -\lim_{i\to\infty}(n-2)\left[\int_{T_i}^{t_i} + \int_{t_i}^{T_{i+1}}\right] Q^2 dt,
\end{aligned} \tag{14}$$

where we used the fact that since $V(T_i)$ converges to the limit cycle,

$$\lim_{i\to\infty} V(T_i) = V_\infty > 0. \tag{15}$$

This implies that

$$\begin{aligned}
0 &= \lim_{i\to\infty}\int_{T_i}^{t_i} Q^2 dt = \lim_{i\to\infty}\int_{V(T_i)}^{V(t_i)} Q dV, \\
0 &= \lim_{i\to\infty}\int_{t_i}^{T_{i+1}} Q^2 dt = \lim_{i\to\infty}\int_{V(t_i)}^{V(T_{i+1})} Q dV.
\end{aligned}$$

Therefore, we have

$$\lim_{t\to\infty} Q(t) = 0 \tag{16}$$

Again we use the identity (9) in Lemma 3.3 and (16) to find

$$\begin{aligned}
0 &= \lim_{i\to\infty}\frac{1}{2}\Big(Q^2(T_i) - Q^2(S_i)\Big)^2 \\
&= \lim_{i\to\infty} 2(n-2)\int_{V(S_i)}^{V(T_i)}(1-e^V)dV - \lim_{i\to\infty}(n-2)\int_{S_i}^{T_i} Q^2 dt \\
&= \lim_{i\to\infty} 2(n-2)\int_{V(S_i)}^{V(T_i)}(1-e^V)dV \quad \text{by (14)} \\
&= \lim_{i\to\infty} 2(n-2)\int_{0}^{V(T_i)}(1-e^V)dV \\
&= 2(n-2)\int_0^{V_\infty}(1-e^V)dV \\
&= 2(n-2)\Big(V_\infty - (e^{V_\infty} - 1)\Big)
\end{aligned}$$

This yields

$$V_\infty = 0,$$

which is a contradiction to (15). Similarly, we can derive

$$\lim_{i\to\infty} V(t_i) = 0.$$

- Case B ($n \geq 10$): We have already proven this case in Proposition 2. Therefore (V, Q) converges to the unique equilibrium point $(0, 0)$. This completes the proof.

4. **Nonlinear instability of stationary radial solutions to the V-P-B system.** In this section, we present the proof of Theorem 1.1. For this, we first recall the definition of the Galilean boost of measurable functions.

Definition 4.1. Let $f = f(x, v, t)$ be a measurable function. Then we define the Galilean boost f_β of f by

$$f_\beta(x, v, t) := f(x + t\beta, v + \beta, t), \quad \beta \in \mathbb{R}^n.$$

In the following lemma, we show that the Galilean boost of solutions to the V-P-B system is still a solution of it.

Lemma 4.2. *Let $f \in C^1(\mathbb{R}^{2n} \times \mathbb{R})$ be a continuously differentiable function. Then we have*

$$\begin{aligned}&(\partial_t f_\beta + v \cdot \nabla_x f_\beta + E[f_\beta] \cdot \nabla_v f_\beta - Q(f_\beta, f_\beta))(x, v, t)\\ &\quad = (\partial_t f + v \cdot \nabla_x f + E[f] \cdot \nabla_v f - Q(f, f))(x + t\beta, v + \beta, t).\end{aligned}$$

Proof. We separate its estimate into two parts (Vlasov and collision parts).

We claim the following: For $\beta \in \mathbb{R}^n$,

$$\begin{aligned}&(i)\ (\partial_t f_\beta + v \cdot \nabla_x f_\beta + E[f_\beta] \cdot \nabla_v f_\beta)(x, v, t)\\ &\qquad\qquad = (\partial_t f + v \cdot \nabla_x f + E[f] \cdot \nabla_v f)(x + t\beta, v + \beta, t).\\ &(ii)\ Q(f_\beta, f_\beta)(x, v, t) = Q(f, f)(x + \beta t, v + \beta, t).\end{aligned}$$

The proof of the claim: (i) By direct calculation, the first partial derivatives of f_β are given by

$$\begin{aligned}\partial_t f_\beta(x, v, t) &= \lim_{h \to 0} \frac{f_\beta(x, v, t + h) - f_\beta(x, v, t)}{h}\\ &= \lim_{h \to 0} \frac{f(x + (t + h)\beta, v + \beta, t + h) - f(x + t\beta, v + \beta, t)}{h}\\ &= \Big(\partial_t f + \beta \cdot \nabla_x f\Big)(x + t\beta, v + \beta, t),\\ \nabla_x f_\beta(x, v, t) &= \nabla_x f(x + t\beta, v + \beta, t),\\ \nabla_v f_\beta(x, v, t) &= \nabla_v f(x + t\beta, v + \beta, t).\end{aligned}$$

Finally, we combine the above relations and Lemma 3.1 in our previous paper [6] to get

$$\begin{aligned}&(\partial_t f_\beta + v \cdot \nabla_x f_\beta + E[f_\beta] \cdot \nabla_v f_\beta)(x, v, t)\\ &\qquad = \ \partial_t f(x + t\beta, v + \beta, t) + (v + \beta) \cdot \nabla_x f(x + t\beta, v + \beta, t)\\ &\qquad + \ E[f](x + t\beta, t) \cdot \nabla_v f(x + t\beta, v + \beta, t)\\ &\qquad = \ (\partial_t f + v \cdot \nabla_x f + E[f] \cdot \nabla_v f)(x + t\beta, v + \beta, t).\end{aligned}$$

(ii) We first recall the collision transformation:

$$v' + v'_* = v + v_*, \qquad |v'|^2 + |v'_*|^2 = |v|^2 + |v_*|^2.$$

This yields

$$\begin{aligned}(v' + \beta) + (v'_* + \beta) &= (v + \beta) + (v_* + \beta),\\ |v'_* + \beta|^2 + |v'_* + \beta|^2 &= |v + \beta|^2 + |v_* + \beta|^2.\end{aligned} \tag{1}$$

Hence it follows from (1) that

$$(v + \beta)' = v' + \beta, \qquad (v_* + \beta)' = v'_* + \beta. \tag{2}$$

We use (2) to get

$$
\begin{aligned}
&Q(f_\beta, f_\beta)(x,v)\\
&= \iint_{\mathbb{R}^n\times\mathbb{S}^{n-1}_+} B(v-v_*,\omega)\Big(f_\beta(v'_*)f_\beta(v') - f_\beta(v)f_\beta(v_*)\Big)dv_*d\omega\\
&= \iint_{\mathbb{R}^n\times\mathbb{S}^{n-1}_+} B(v-v_*,\omega)\Big(f(x+\beta t, v'_*+\beta)f(x+\beta t, v'+\beta)\\
&\qquad\qquad -f(x+\beta t, v+\beta)f_\beta(x+\beta t, v_*+\beta)\Big)dv_*d\omega\\
&= \iint_{\mathbb{R}^n\times\mathbb{S}^{n-1}_+} B((v+\beta)-(v_*+\beta),\omega)\Big(f_\beta(x+\beta t,(v_*+\beta)')f(x+\beta t,(v+\beta)')\\
&\qquad\qquad -f(x+\beta t, v+\beta)f(x+\beta t, v_*+\beta)\Big)dv_*d\omega\\
&= Q(f,f)(x+\beta t, v+\beta).
\end{aligned}
$$

□

Remark 5. The result of Lemma 4.2 holds for weak stationary solutions; i.e., if f_0 is a weak stationary solution, then $f_\beta := f_{0\beta}$ is also a weak solution.

We now return to the proof of Theorem 1.1.

Step A (Construction of perturbations): Let f_0 be a stationary radial solution whose existence and asymptotic behavior are guaranteed by Theorem 3.1, and let $p \in (\frac{n}{2}, \infty)$. To generate the desired perturbed solutions of the given stationary radial solution f_0, we take the following family of perturbed data: For $\beta \in \mathbb{R}^n$,

$$f^{in}_\beta(x,v) := f_0(x, v+\beta).$$

Note that f^{in}_β and f_0 have the same local mass density but different velocity:

$$\rho[f^{in}_\beta] = \rho[F], \quad u[f^{in}_\beta] = u[f_0] - \beta.$$

However, it follows from Lemma 4.2 that the solution to Equation (1) with initial datum f^{in}_β is given by

$$f_\beta(x,v,t) = f_0(x+t\beta, v+\beta).$$

Thus corresponding macroscopic quantities are

$$\rho[f_\beta](x,t) = \rho[f_0](x+\beta t), \quad u[f_\beta](x,t) = u[f_0](x) - \beta.$$

Step B (Instability estimate): We claim the following:

(i) $\lim\limits_{|\beta|\to 0} ||f^{in}_\beta - f_0||_{L^p} = 0.$

(ii) For given $\varepsilon > 0$, $\exists\, T = T(\varepsilon) > 0$ such that for $t > T(\varepsilon)$, $||f(t) - f_0||_p \geq ||f_0||_p$,

The proof of the claim: (i) By definition, we have

$$
\begin{aligned}
||f^{in}_\beta - f_0||^p_{L^p} &= \int_{\mathbb{R}^{2n}} |f^{in}_\beta(x,v) - f_0(x,v)|^p dvdx\\
&= \int_{\mathbb{R}^{2n}} |f_0(x,v+\beta) - f_0(x,v)|^p dvdx.
\end{aligned}
$$

Note that the integrand is bounded by an integrable function, i.e.,

$$|f_0(x,v+\beta) - f_0(x,v)|^p \leq 2^{p-1}\Big(|f_0(x,v+\beta)|^p + |f_0(x,v)|^p\Big) : \text{ integrable},$$

Thus by the Lebesgue dominated convergence theorem, we have

$$\lim_{|\beta|\to 0} ||f_\beta^{in} - f_0||_{L^p}^p = \int_{\mathbb{R}^{2n}} \lim_{|\beta|\to 0} |f_0(x, v+\beta) - f_0(x,v)|^p dvdx = 0.$$

Therefore for a given $\varepsilon > 0$, there exists β_0 such that, for $|\beta| \le \beta_0 = \beta_0(\varepsilon)$,

$$||f_\beta^{in} - f_0||_{L^p} < \varepsilon.$$

(ii) We set

$$z_\beta := (-\beta t, -\beta) \in \mathbb{R}^{2n},$$

and let $B_R(z_\beta)$ be the R ball with a center z_β. For a given small positive number $\eta > 0$, we choose sufficiently large R such that

$$\Big(\int_{\mathbb{R}^{2n}\setminus B_R(z_0)} f_0(x,v)^p dxdv\Big)^{1/p} = ||f_0 - f_0\chi_{B_R(z_0)}||_{L^p} < \eta. \tag{3}$$

By the triangle inequality, we have

$$\begin{aligned} &||f(t)\chi_{B_R(z_\beta)} - f_0\chi_{B_R(z_0)}||_{L^p} - ||f(t) - f_0||_{L^p} \\ &\quad \le ||f(t)\chi_{B_R(z_\beta)} - f_0\chi_{B_R(z_0)} - (f(t) - f_0)||_{L^p} \\ &\quad = ||(f(t)\chi_{B_R(z_\beta)} - f(t)) - (f_0\chi_{B_R(z_0)} - f_0)||_{L^p} \\ &\quad \le ||(f(t)\chi_{B_R(z_\beta)} - f(t))||_{L^p} + ||f_0\chi_{B_R(z_0)} - f_0||_{L^p} \\ &\quad \le 2\eta. \end{aligned}$$

Therefore, we can obtain the following inequality:

$$||f(t)\chi_{B_R(z_\beta)} - f_0\chi_{B_R(z_0)}||_{L^p} - 2\eta \le ||f(t) - f_0||_{L^p}. \tag{4}$$

For a small $\varepsilon > 0$, we choose $T_0 = T_0(\varepsilon)$ to satisfy

$$|\beta| T_0 > R.$$

Then it easy to see that

$$\mathrm{dist}(B_R(z_0), B_R(z_\beta)) \ge |\beta| T_0, \quad t \ge T := 4T_0.$$

Hence for $t \ge T$ and $\frac{n}{2} < p < \infty$, we have

$$\begin{aligned} & ||f(t)\chi_{B_R(z_\beta)} - f_0\chi_{B_R(z_0)}||_{L^p}^p \\ = \quad & \int_{\mathbb{R}^{2n}} |f(x,v,t)\chi_{B_R(z_\beta)} - f_0(x,v)\chi_{B_R(z_0)}|^p dxdv \\ = \quad & \int_{B_R(z_0)\cup B_R(z_\beta)} |f(x,v,t)\chi_{B_R(z_\beta)} - f_0(x,v)\chi_{B_R(z_0)}|^p dxdv \\ = \quad & 2^{p-1}\Big(\int_{B_R(z_\beta)} |f(x,v,t)\chi_{B_R(z_\beta)}|^p dxdv + \int_{B_R(z_0)} |f_0(x,v)\chi_{B_R(z_0)}|^p dxdv\Big) \\ = \quad & 2^p \int_{B_R(z_0)} |f_0(x,v)\chi_{B_R(z_0)}|^p dxdv. \end{aligned}$$

This again yields

$$||f(t)\chi_{B_R(z_\beta)} - f_0\chi_{B_R(z_0)}||_{L^p} = 2||f_0\chi_{B_R(z_0)}||_{L^p}. \tag{5}$$

By the triangle inequality,

$$\begin{aligned} ||f_0||_{L^p} &= ||f_0 - f_0\chi_{B_R(z_0)} + f_0\chi_{B_R(z_0)}||_{L^p} \\ &\le \Big(||f_0 - f_0\chi_{B_R(z_0)}||_{L^p} + ||f_0\chi_{B_R(z_0)}||_{L^p}\Big). \end{aligned}$$

Thus we have

$$||f_0||_{L^p} - ||f_0 - f_0\chi_{B_R(z_0)}||_{L^p} \le ||f_0\chi_{B_R(z_0)}||_{L^p}. \tag{6}$$

For $t > T$, we have

$$\begin{array}{rcll} ||f(t) - f_0||_{L^p} & \geq & ||f(t)\chi_{B_R(z_\beta)} - f_0\chi_{B_R(z_0)}||_{L^p} - 2\eta & \text{by (4)} \\ & = & 2||f_0\chi_{B_R(z_0)}||_{L^p} - 2\eta & \text{by (5)} \\ & \geq & 2\Big(||f_0||_{L^p} - ||f_0 - f_0\chi_{B_R(z_0)}||_{L^p}\Big) - 2\eta & \text{by (6)} \\ & \geq & 2||f_0||_{L^p} - 4\eta & \text{by (3)} \end{array}$$

If we take $\eta > 0$ sufficiently small so that

$$\eta \leq \frac{1}{4}||f_0||_{L^p}.$$

then for such η we have

$$||f(t) - f_0||_{L^p} \geq ||f_0||_{L^p}.$$

This completes the proof.

Remark 6. For the $p = \infty$ case, we can obtain a similar unstable result as for $p < \infty$ by using the radially monotonic property of the stationary solution $f = f(x, v)$, which is constructed in the previous section.

5. **Conclusion.** In this paper, we established the existence and instability of stationary radial solutions to the Vlasov–Poisson–Boltzmann system with attractive self-consistent forces in the absence of external forcing and background density. The stationary radial solutions are in fact the corresponding solutions to the Vlasov–Poisson system, which are of the form of local Maxwellians:

$$f(x, v) = e^{u(|x|)}e^{-c_0|v|^2},$$

where $u = u(|x|)$ are $\mathcal{C}^2$ functions satisfying

$$\Delta_x u + \kappa_n e^u = 0, \qquad \kappa_n = 2c_0\big(\pi/c_0\big)^{n/2}.$$

By the phase-portrait analysis, we make a rigorous argument for the asymptotic behavior of u, which was first obtained by Joseph and Lungren [20]:

$$\lim_{r\to\infty}\left(u(r) - \log\frac{2(n-2)}{\kappa_n r^2}\right) = 0.$$

Therefore, our constructed radial solution $f = f(x, v)$ belongs to $L^p(\mathbb{R}^{2n})$, $\quad p > \frac{n}{2}, \quad n > 2$. For the instability of the stationary radial solutions, we employed the Galilean boost method. This Galilean boost generates a one-parameter family of time-dependent perturbed solutions whose supports are *almost* separated from that of the given stationary radial solution as $t \to \infty$. By this simple observation, the Vlasov–Poisson–Boltzmann system with an attractive force is dynamically unstable for the $L^p(\mathbb{R}^{2n})$ norm with $p > \frac{n}{2}$.

Appendix A. **The proof of Proposition 2.** In this appendix, we present the detailed proof of Proposition 2.

It follows from Lemma 3.4 that the trajectory cannot stay within the region $\mathcal{H}^+$ and $\mathcal{H}^-$ forever. Hence we have the following possibilities(see Figure 1):

S-A: $(V(t), Q(t))$ does touch the line $\{(V, Q) \in \mathbb{R}^2 : V < 0,\ Q = 0\}$ at a finite time or infinite time.

S-B: $(V(t), Q(t))$ tends to $(0, 0)$ without touching the line $\{(V, Q) \in \mathbb{R}^2 : V < 0,\ Q = 0\}$ and $\{(V, Q) \in \mathbb{R}^2 : V = 0,\ Q > 0\}$.

S-C: $(V(t), Q(t))$ approaches asymptotically to a point on the line $\{(V, Q) \in \mathbb{R}^2 : V > 0,\ Q = 0\}$ at $T = \infty$.

S-D: $(V(t),Q(t))$ touches the line $\{(V,Q)\in\mathbb{R}^2 : V>0,\ Q=0\}$ at a finite time T, and it approaches the unique fixed point $(0,0)$ in a spiraling trajectory.

S-E: $(V(t),Q(t))$ does not touch the line $\{(V,Q)\in\mathbb{R}^2 :\ Q=0\}$ but $V(t)$ goes to infinity.

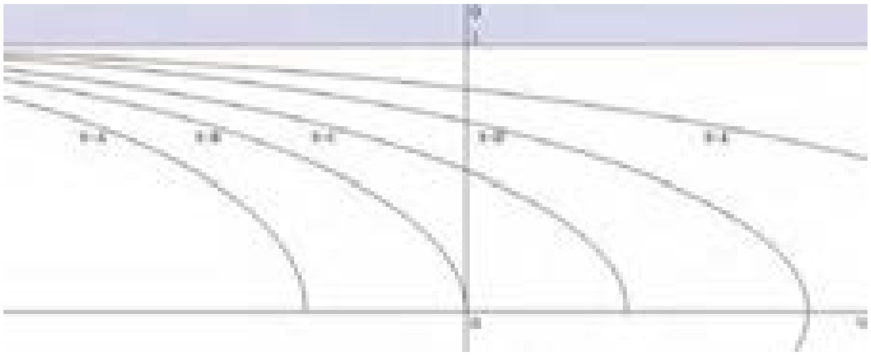


Figure 1.

In the following, we will show that the S-D case only occurs for $n<10$ and the S-B case only occurs for $n\geq 10$.

Case A-1: Suppose $(V(t),Q(t))$ touches the line $\{(V,Q)\in\mathbb{R}^2 : V<0,\ Q=0\}$ at a finite time $t=T<\infty$. Then we set $t_*\to-\infty$ and $t^*=T$ in (8) and use $Q(T)=0,\quad Q(-\infty)=2$ to get

$$0=2(n-2)\int_{-\infty}^{T}(1-e^{V})e^{(n-2)t}dt>0,$$

where we used the fact that $V(t)<0,\quad t\in(-\infty,T]$. This gives a contradiction.

Case A-2: Suppose $(V(t),Q(t))$ touches the line $\{(V,Q)\in\mathbb{R}^2 : V<0,\ Q=0\}$ at $t=\infty$. In this case,

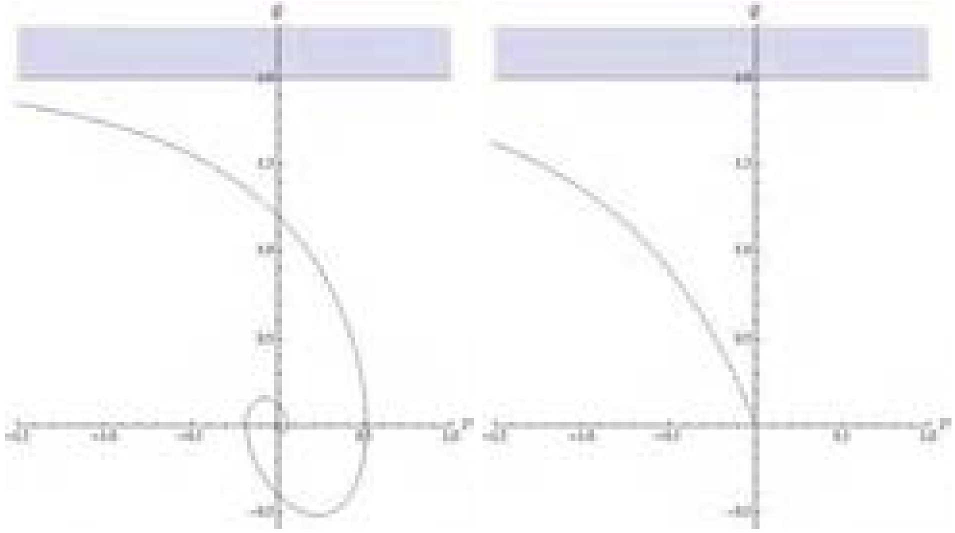
$$\begin{aligned}0=\lim_{T\to\infty}Q(T) &= \lim_{T\to\infty}\frac{2(n-2)\int_{-\infty}^{T}(1-e^{V})e^{(n-2)t}dt}{e^{(n-2)T}}\\ &= \lim_{T\to\infty}\frac{2(n-2)(1-e^{V(T)})e^{(n-2)T}}{(n-2)e^{(n-2)T}}\\ &= \lim_{T\to\infty}2(1-e^{V(T)})>0.\end{aligned}$$

We combine Case A-1 and Case A-2 to conclude that Case A cannot occur for the dynamics of (7) regardless of dimension n.

Now, we can prove the second statement of this proposition. Assume that for $n\geq 10$, Case C or D or E occurs. Then, by Lemma 3.4, there is a finite number T such that

$$T=\sup\{\tau : V(t)<0 \text{ on } (-\infty,\tau]\}.$$

Then we have $V(T)=0$ and $Q(T)>0$. For arbitrary $\mu>0$, there might be several intersection points between the line $Q=-\mu V$ and the trajectory $(V(t),Q(t))$.



(a) $2 < n < 10$

(b) $n \geq 10$

FIGURE 2.

Among them, we set (V_μ, Q_μ) to be the intersection point with the largest V values. By direct calculation, we have

$$\begin{aligned}
-\mu \leq \left.\frac{dQ}{dV}\right|_{V=V_\mu} &= -(n-2) + 2(n-2)\left.\frac{1-e^V}{Q}\right|_{V=V_\mu} && \text{by (6)} \\
&= -(n-2) + 2(n-2)\frac{1-e^{V_\mu}}{-\mu V_\mu} \\
&= -(n-2) + \frac{2(n-2)}{\mu}\frac{e^{V_\mu}-1}{V_\mu} \\
&< -(n-2) + \frac{2(n-2)}{\mu} && \text{by } V_\mu < 0.
\end{aligned} \tag{1}$$

By (1), we can obtain

$$\mu^2 - (n-2)\mu + 2(n-2) > 0, \quad \text{for all } \mu > 0.$$

This implies $(n-2)^2 - 8(n-2) < 0$ but we already assume that $n \geq 10$. This is a contradiction. Therefore, for spatial dimension $n \geq 10$, the trajectory $(V(t), Q(t))$ goes to the equilibrium point $(0,0)$ directly without touching the line $V = 0$ as $t \to \infty$.

Below, we consider the case where $2 < n < 10$.

Case B-1: Suppose Scenario B occurs in a finite time $t = T$. Then by the same argument with $t_* = -\infty$ and $t^* = T$ as in Case A-1, we obtain a contradiction.

Case B-2: Suppose (V, Q) converges to $(0,0)$ as $t \to \infty$. In this case, by the monotonicity of V, the trajectory satisfies

$$(V(t), Q(t)) \to (0,0) \quad \text{as} \quad t \to \infty \quad \text{and} \quad V(t) < 0, \quad 0 < Q(t) < 2, \quad t \geq 0.$$

Consider a nontrivial test function $q(t)$ that is a solution to the following ODE:

$$q'' + (n-2)q' + bq = 0, \quad (q, q')(0) = (0, 1), \tag{2}$$

where $b > \frac{(n-2)^2}{4}$ is a positive constant to be determined later. By direct calculation, we have an oscillatory solution:

$$q(t) = \frac{2e^{-(n-2)t/2}}{\sqrt{-(n-2)^2 + 4b}} \sin\Big(\frac{t}{2}\sqrt{-(n-2)^2 + 4b}\Big).$$

Since $\lim_{t\to\infty} V(t) = 0$, there exists a T such that

$$\frac{e^V - 1}{V} \approx 1, \quad t \geq T.$$

Then for such T, we can choose b to satisfy

$$\frac{(n-2)^2}{4} < b < 2(n-2)\frac{e^V - 1}{V} \quad \text{on } [T, \infty). \tag{3}$$

However, V satisfies the ODE (6):

$$V'' + (n-2)V' + 2(n-2)\frac{(e^V - 1)}{V}V = 0. \tag{4}$$

We now consider

$$V \times (2) - q \times (4).$$

This yields

$$(Vq' - V'q)' + (n-2)(Vq' - qV') + \Big(b - 2(n-2)\frac{(e^V - 1)}{V}\Big)Vq = 0. \tag{5}$$

Since $q(t)$ is a oscillatory solution, we can choose a $T < t_0 < t_1$ satisfying the following:

$$q(t) \geq 0 \text{ on } [t_0, t_1], \quad q(t_0) = q(t_1) = 0 \text{ and } q'(t_0) > 0 > q'(t_1). \tag{6}$$

We multiply (5) by the integrating factor $e^{(n-2)t}$ to find

$$\Big(e^{(n-2)s}(Vq' - V'q)\Big)' = -\Big(b - 2(n-2)\frac{(e^{V(s)} - 1)}{V(s)}\Big)V(s)q(s)e^{(n-2)s}.$$

We integrate this relation from t_0 to t_1 to obtain

$$\begin{aligned} &e^{(n-2)t_1}V(t_1)q'(t_1) - e^{(n-2)t_0}V(t_0)q'(t_0) \\ &\quad = -\int_{t_0}^{t_1}\Big(b - 2(n-2)\frac{(e^{V(s)} - 1)}{V(s)}\Big)V(s)q(s)e^{(n-2)s}ds, \end{aligned} \tag{7}$$

where we used

$$q(t_0) = q(t_1) = 0.$$

Note that relation (6) implies that

$$\text{L.H.S. of (7)} > 0.$$

In contrast, the R.H.S. of (7) is negative because

$$\Big(b - 2(n-2)\frac{(e^{V(s)} - 1)}{V(s)}\Big) < 0, \quad V(s)q(s)e^{(n-2)s} < 0, \quad t \in (t_0, t_1).$$

Hence Scenario B does not occur for the $2 < n < 10$ case.

Case C: Suppose Scenario C occurs. In this case, we must have

$$\lim_{t\to\infty} V'(t) = 0, \quad \lim_{t\to\infty} Q'(t) = 0.$$

However, it follows from (7) that this forces (V, Q) to approach $(0, 0)$ as $t \to \infty$, which is contradictory to Scenario C.
Case E: Suppose Scenario E occurs. In this case, we can find t_* that satisfies $V(t_*) = 0$. For such t_*, we also choose $t^* = \infty$ in (9) to find

$$\frac{1}{2}\Big(Q(\infty)^2 - Q(t_*)^2\Big) = -\Big[2(n-2)\int_0^\infty (e^V - 1)dV + (n-2)\int_{t_*}^\infty Q^2(t)dt\Big].$$

Then it is easy to see that the R.H.S. of the above relation is $-\infty$, whereas the L.H.S. is bounded. Hence we have a contradiction.

Therefore for the $2 < n < 10$ case, $(V(t), Q(t))$ touches the line $\{(V, Q) \in \mathbb{R}^2 : V > 0,\ Q = 0\}$ at a finite time T, i.e.,

$$V(T) = V_*, \quad Q(T) = 0.$$

Then it follows from (7) that

$$V'(T) = 0, \quad Q'(T) = 2(n-2)(1 - e_*^V) < 0.$$

Hence for $t \in (T, T+\varepsilon)$, for some $\varepsilon > 0$,

$$Q(t) < Q(T) = 0, \quad t \in (T, T+\varepsilon).$$

This also implies that V is strictly decreasing on $(T, T+\varepsilon)$. We repeat the same arguments in this proposition to show that the trajectory (V, Q) should touch the line $\{(V, Q) \in \mathbb{R}^2 : V < 0,\ Q = 0\}$ in a finite time $T_* > T$ again. By repeating this argument, we can see that the trajectory (V, Q) swirls around the unique fixed point $(0, 0)$ in the two-dimensional phase plane.

REFERENCES

[1] R. J. Alonso and I. M. Gamba *Distributional and classical solutions to the Cauchy-Boltzmann problem for soft potentials with integrable angular cross section*, J. Stat. Phys., **137** (2009), 1147–1165.

[2] C. Bardos and P. Degond *Global existence for the Vlasov–Poisson equation in three space variables with small initial data*, Ann. Inst. Henri Póincare C, **2** (1985), 101–118.

[3] M. Chae and S.-Y. Ha *New Lyapunov functionals of the Vlasov-Poisson system*, SIAM J. Math. Anal., **37** (2006), 1709–1731.

[4] M. Chae, S.-Y. Ha and H. Hwang *Time-asymptotic behavior of the Vlasov–Poisson–Boltzmann system near vacuum*, J. Differential Equations, **230** (2006), 71–85.

[5] S.-H. Choi, S.-Y. Ha and H. Lee *Dispersion estimates for the two-dimensional Vlasov.Yukawa system with small data*, J. Differential Equations, **250**(1) (2011), 515–550.

[6] S.-H. Choi and S.-Y. Ha *Dynamic instability of stationary solutions to the nonlinear Vlasov equations*, Int. J. Numer. Anal. Model. Ser. B, **2** (2011), 415–421.

[7] L. Desvillettes and J. Dolbeault *On long time asymptotics of the Vlasov–Poisson–Boltzmann equation*, Commun. Partial Diffential Equations, **16** (1991), 451–489.

[8] R.-J. Duan and R.M. Strain *Optimal time decay of the Vlasov–Poisson–Boltzmann system in $\mathbb{R}^3$*, Arch. Rat. Mech. Anal., **199** (2011), 291–328.

[9] R.-J. Duan and T. Yang *Stability of the one-species Vlasov–Poisson–Boltzmann system*, SIAM J. Math. Anal., **41**(6) (2010), 2353–2387.

[10] R.-J. Duan, T. Yang and C. Zhu *Existence of stationary solutions to the Vlasov–Poisson–Boltzmann system*, J. Math. Anal. Appl., **327** (2007), 425–434.

[11] R.-J. Duan, T. Yang and C. Zhu *L^1 and BV-type stability of the Boltzmann equatiion with external forces*, J. Differential Equations, **227** (2006), 1–28.

[12] R.-J. Duan, M. Zhang and C. Zhu *L^1-stability for the Vlasov-Poisson-Boltzmann system around vacuum*, Mathematical models and Methods in Applied Sciences, **16** (2006), 1505–1526.

[13] R. Glassey The Cauchy Problem in Kinetic Theory Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. 1996

[14] R. Glassey and W.A. Strauss *Perturbation of essential spectra of evolution operators and the Vlasov–Poisson–Boltzmann system*, Discrete Contin. Dynam. Systems, **5** (1999), 457–472.

[15] Y. Guo *The Vlasov–Poisson–Boltzmann system near Maxwellians*, Commun. Pure Appl. Math., **55** (2002), 1104–1135.

[16] Y. Guo *The Vlasov–Poisson–Boltzmann system near vacuum*, Commun. Math. Phys., **218** (2001), 293–313.

[17] S.-Y. Ha *Nonlinear functionals of the Boltzmann equation and uniform stability estimates*, J. Differential Equations, **215** (2005), 178–205.

[18] S.-Y. Ha *L^1-stability of the Boltzmann equation for the hard-sphere Model*, Arch. Rat. Mech. Anal., **173**(2) (2004), 279–296.

[19] S.-Y. Ha and H. Lee *Global well posedness of the relativistic Vlasov–Yukawa system with small data*, J. Math. Phys., **48** (2007), 123508.

[20] D.D. Joseph and T.S. Lundgren *Quasilinear Dirichlet problems driven by positive sources*, Arch. Rat. Mech. Anal., **49**(4) (1973), 241–269.

[21] P. L. Lions and B. Perthame *Propagation of moments and regularity for the 3-dimensional Vlasov–Poisson system*, Invent. Math., **105**(1) (1991), 415–430.

[22] S. Mischler *On the initial boundary value problem for the Vlasov–Poisson–Boltzmann system*, Commun. Math. Phys., **210** (2000), 447–466.

[23] T. Yang, H.-J. Yu and H.-J. Zhao *Cauchy problem for the Vlasov–Poisson–Boltzmann system*, Arch. Rat. Mech. Anal., **182**(3) (2006), 415–470.

[24] T. Yang and H.-J. Yu *Optimal convergence rates of classical solutions for Vlasov–Poisson–Boltzmann system*, Commun. Math. Phys., **301** (2011), 319–355.

[25] M. Zhang *Stability of the Vlasov–Poisson–Boltzmann system in* $\mathbb{R}^3$, J. Differential Equations, **247** (2009), 2027–2073.

E-mail address: shbae@hanbat.ac.kr
E-mail address: matcs@nus.edu.sg
E-mail address: syha@snu.ac.kr
URL: http://www.math.snu.ac.kr/~syha

NODAL CONDITIONS FOR HYPERBOLIC SYSTEMS OF BALANCE LAWS

Rinaldo M. Colombo

Università degli studi di Brescia
Via Branze 38
25123 Brescia, Italy

Michael Herty

RWTH Aachen University
Templergraben 55
52056 Aachen, Germany

Abstract. In recent years, research on nodal conditions for one dimensional systems of hyperbolic balance laws has been developed by both the mathematical and the engineering community. The result is a theory consisting of a variety of specific (often unrelated) applications, comprising for instance vehicular traffic dynamics, data flows in telecommunication networks, supply chains as well as physical systems like water canals or gas networks. In many of these cases, *ad hoc* well–posedness results were obtained, allowing to rigorously state and consider problems of deep applicative interest, such as the optimal control or the controllability of these equations. The present work aims to summarise several recent contributions to this field from a unified point of view.

1. **Introduction.** Consider a system of balance laws in one space dimension of the form

$$\partial_t u + \partial_x f(u) = g(t, x, u)\,, \tag{1}$$

where $t \in \mathbb{R}^+$ is time, $x \in \mathbb{R}$ is the space variable, f is the flow and g the source term. A wide variety of situations fit into (1) equipped with a further algebraic condition imposed at a *nodal* point x_*, which can be fixed, leading to

$$\Psi\left(t, u(t, x_*+)\right) = 0. \tag{2}$$

or a function $x_* = x_*(t)$ implement a *coupling* with ordinary differential equations, such as

$$\Psi\left(t, x_*(t), \dot{x}_*(t), \ddot{x}_*(t), u\left(t, x_*(t)+\right)\right) = 0\,. \tag{3}$$

This presentation reviews various examples of problems leading to (1)–(2) or (1) and (3), showing the analogies and the differences among the many available recent results. A key role is played by the nodal condition (2) or (3). Indeed, the function Ψ may either stem from physical considerations of the specific real situation being modeled, or play the role of an external control to be chosen according to suitable optimality criteria. In the case of phase transitions, for instance, the nodal condition

2000 *Mathematics Subject Classification.* Primary: 35L65; Secondary: 49J20.

Key words and phrases. Hyperbolic Balance Laws, Coupling Conditions for Conservation Laws, Conservation Laws on Networks.

The work has been supported by DFG HE5386/8-1, EXC128, and by the Vigoni project *Non-Local Transport Processes Modeling, Analysis, Numerics and Optimal Control.*

is often referred to as *kinetic condition*, since it relates the flow of changing phase fluid to the states on the sides of the phase boundary.

Below, we mostly consider systems with a single nodal point, although the finite propagation speed typical of (1) easily allows the extension to the case of multiple nodal points and thereby to networks, see also [45].

The current literature offers several well posedness theorems for (1)–(2) or (1)–(3). Mostly, these results are local both in time and in the u variable, in the sense that they often require the initial datum to be a perturbation with sufficiently small total variation of a stable constant state. A widely used analytical technique to obtain these results is *wave front tracking*, originated in [33] and since then extensively refined and exploited, see for instance [15, 23, 25, 28, 32, 34, 44, 50].

2. Fixed Nodal Points.

In the case of a fixed nodal point, one usually considers n half-lines with common origin at $x_* = 0$. System (1) then consists of n independent balance laws $\partial_t u_j + \partial_x f_j(u_j) = g_j(t, x, u_j)$, each defined for $x > 0$, coupled through the nodal condition (2). The wave front tracking procedure can be used to construct solutions to (1)–(2) as soon as a well posedness theory for the Riemann problem at the nodal point, i.e., of

$$\left\{\begin{array}{ll} \partial_t u + \partial_x f(u) = 0 & \\ \Psi\left(u(t, 0+)\right) = 0 & \begin{array}{l} t \in \mathbb{R}^+ \\ x \in \mathbb{R}^+ \end{array} \\ u(0, x) = \bar{u}\,, & \end{array}\right. \tag{4}$$

where

$$u = \left[\begin{array}{c} u_1 \\ \dots \\ u_n \end{array}\right], \qquad f(u) = \left[\begin{array}{c} f_1(u_1) \\ \dots \\ f_n(u_n) \end{array}\right], \qquad g(t, x, u) = \left[\begin{array}{c} g_1(t, x, u_1) \\ \dots \\ g_n(t, x, u_n) \end{array}\right], \tag{5}$$

with constant values $\bar{u}_j$ is obtained. Once a good well posedness theory for (1)–(2) is available, the well posedness of the Cauchy problem might also be obtained.

To this aim, one first considers the homogeneous the case $g = 0$. As in the case of the standard Cauchy problem for a system on hyperbolic conservation laws, using, for instance, the wave front tracking algorithm [33], a sequence of approximate solutions to the Cauchy problem at the nodal point, namely

$$\left\{\begin{array}{ll} \partial_t u + \partial_x f(u) = 0 & \\ \Psi\left(u(t, 0+)\right) = 0 & \begin{array}{l} t \in \mathbb{R}^+ \\ x \in \mathbb{R}^+\,, \end{array} \\ u(0, x) = u_o(x)\,, & \end{array}\right. \tag{6}$$

is constructed. Bounds on the total variation allow, through Helly Compactness Theorem, to prove the convergence of this sequence. More careful estimates, relying on pseudopolygonals [16] when $n = 2$ and on the Liu–Yang functional [17, 64] for $n \geq 3$, ensure that the approximate solutions or their limit are L^1–Lipschitz function of the initial datum. Typically, results of this type are local in the u space, in the sense that the initial datum u_o in (6) is required to satisfy

$$u_o \in u^* + L^1 \quad \text{and} \quad \mathrm{TV}(u_o - u^*) < \delta \tag{7}$$

for a sufficiently small δ and a given state u^* required to satisfy typically $\Psi(u^*) = 0$.

For all this to hold, classical assumptions on f satisfied in the various cases discussed below, are that Jacobian $Df(u^*)$ is strictly hyperbolic and that each characteristic field is either genuinely nonlinear or linearly degenerate. More precisely

f needs to be at least $C^4(\mathbb{R}^N;\mathbb{R}^N)$ where N the overall dimension of the system (1) and $Df(u^*)$ is required to admit N real distinct eigenvalues

$$\lambda_1(u^*) < \lambda_2(u^*) < \ldots < \lambda_N(u^*)$$

such that, together with the corresponding eigenvectors $r_1, r_2, \ldots, r_N$, satisfy

$$\text{for } i = 1, \ldots, N \quad \left\{ \begin{array}{rl} \text{either} & \nabla\lambda_i(u) \cdot r_i(u) \neq 0 \text{ at } u = u^*\,, \\ \text{or} & \nabla\lambda_i(u) \cdot r_i(u) \equiv 0 \text{ in a neighborhood of } u^*\,. \end{array} \right.$$

We refer, for instance, to [15, 34] for more details in the standard case of no junction. For completeness, we remark that the latter conditions above were significantly relaxed in [1], see also [34].

The presence of a nodal point, with the corresponding nodal condition (2), deeply interferes with the necessary set of estimates. A standard requirement on the nodal condition (2), besides that $\Psi \in C^1(\mathbb{R}^N;\mathbb{R}^N)$, is a *transversality* condition between the gradient of Ψ and the eigenvectors of Df at u^* of the type

$$\det\left[D_{u_1}\Psi(u^*)\, r_2(u_1^*) \quad \cdots \quad D_{u_n}\Psi(u^*)\, r_2(u_1^*)\right] \neq 0\,.$$

This condition, through an application of the Implicit Function Theorem, allows to obtain suitable estimates on interactions hitting the nodal point. The standard well posedness result for (1)–(2) when $g = 0$ thus ensures the existence of an $\mathbf{L}^1$–Lipschitz semigroup of solutions defined globally in time and locally in u, in the sense of (7). Various characterizations of the solutions, as well as its stability with respect to perturbations of f or Ψ are also available.

As in the case of no junctions, the essential role played by the Implicit Function Theorem hinders the extension of well posedness results beyond condition (7).

Once well posedness is obtained for $g = 0$, the full case

$$\left\{ \begin{array}{ll} \partial_t u + \partial_x f(u) = g(t,x,u) & \\ \Psi\left(u(t,0+)\right) = 0 & \begin{array}{l} t \in \mathbb{R}^+ \\ x \in \mathbb{R}^+\,, \end{array} \\ u(0,x) = u_o(x)\,, & \end{array} \right. \tag{8}$$

usually follows through operator splitting, along the lines initiated in [35]. The source term is then treated as an ordinary differential equation, where u depends on time t and on the space variable as parameter. As a consequence, the usual assumptions on g remind of those in the theory of ordinary differential equations: namely local Lipschitz continuity and sublinearity in u. Due to the key role played by total variation estimates, it is also necessary to ensure that the source term may not cause blow–ups in the oscillations of the approximate solutions. To this end, uniform bound on the total variation in space of g are often required.

Remark that the introduction of a source term typically makes it impossible to obtain global in time existence results. In general, without any invariance condition, the source term may cause a drift of the solution in the u space, exiting the neighborhood of u^* where the Implicit Function Theorem, and hence Riemann problems, can be solved.

A natural further extensions consists in dealing with (1)–(2) on a *network*, which is modeled as directed graph $(\mathcal{G}, \mathcal{A})$. The set of arcs $j \in \mathcal{A}$ connect various nodal points, see for instance [44] for a rigorous definition of network that applies to the present case. The analytical treatment of (1)–(2) on networks consists in a superposition of results described above for (8), one at each nodal point of the network. Clearly, the finite speed of propagation inherent to (1) easily allows to obtain results on a finite time interval. Existence and stability globally in time is much harder to obtain and further assumptions become necessary, see for instance the example [45]. Note that (8) can also be seen as a particular case of a system of balance laws in a domain with boundary, see [28, Section 4.1].

The following paragraphs are devoted to different examples of (8) in specific situations of interest in various mathematical and engineering applications.

2.1. **Gas Dynamics in Pipelines.** Historically, Euler equations of gas dynamics have been the paradigm according to which the theory of conservation (or balance) laws has been developed. Also in the case of nodal points, (high–pressure) gas pipelines provide a wealth of stimulating real problems, see for instance [8, 9, 23, 28, 29, 32, 52, 60].

In a typical real pipe the ratio between pipe length and diameter is large enough to justify the use of Euler equations in *one* space dimension, or approximations thereof. Therefore, the natural setting for gas pipeline models consists of n pipes connected at a fixed point, say $x_* = 0$, and the gas is described through the isentropic Euler equations, so that (1) consists of n copies of the p-system, i.e., in (5) we set

$$u_j = \begin{bmatrix}\rho_j\\ q_j\end{bmatrix} \qquad f_j(u) = \begin{bmatrix} q_j \\ \frac{{q_j}^2}{\rho_j} + p(\rho_j)\end{bmatrix} \qquad g_j(t,x,u) = \begin{bmatrix} 0 \\ -\nu\, \frac{q_j |q_j|}{\rho_J} - \rho_j\, \bar g\, \sin\alpha_j(x)\end{bmatrix} \tag{9}$$

where the pressure law p can be, for instance, the usual γ–law $p(\rho) = \kappa\, \rho^\gamma$. Above, g_j is a typical source term with a zero component in the mass equation, while the second component describes the effect $\bar g \sin\alpha_j(x)$ of gravity ($\alpha_j(x)$ being the slope of the j-th pipe at x) and the effect $-\nu\, q\, |q|/\rho$ of friction (ν being a constant parameter) on the balance of momentum. Remark that in the case $n = 2$ the present setting may also describe the dynamics of gas flowing in a pipe with a kink [60].

The 1D framework allows a rather simple modeling structure and, mostly, very fast numerical integrations. A model in *one* space dimension well describes the dynamics within a pipe, but hardly covers geometry effects at a junction, which is clearly an intimately 3D phenomenon. As a consequence, the literature offers several different choices for the nodal condition (2), depending on the specific needs of each particular situation. In the engineering literature, the nodal conditions are typically supplied with parameters whose values are empirically justified.

In the *subsonic* setting, condition (2) has to provide n conditions to single out a unique solution to the Riemann problem (4). One component, say the first, of Ψ

$$\Psi_1(u) = \sum_{j=1}^n a_j\, q_j \tag{10}$$

ensures the conservation of mass. Here, a_j is the surface section of the j-th pipe.

The other components of Ψ may impose, for instance, equal pressure at the node

$$\Psi_j(u) = p(\rho_j) - p(\rho_1) \qquad j = 2, \ldots, n \tag{11}$$

or the continuity of the dynamic pressure at the node

$$\Psi_j(u) = a_j \left(\frac{{q_j}^2}{\rho_j} + p(\rho_j)\right) - a_1 \left(\frac{{q_1}^2}{\rho_1} + p(\rho_1)\right) \qquad j = 2, \ldots, n\,. \tag{12}$$

We refer to [8, 9, 23, 29] and to the references therein for more details on the related modeling and computational results.

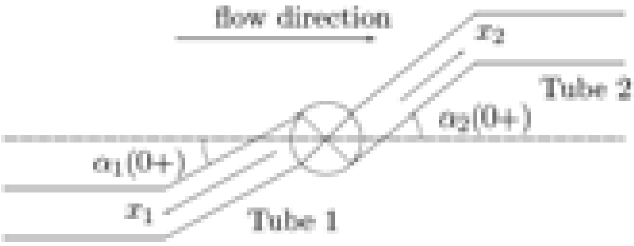


FIGURE 1. Schematic of a gas compressor station with two connected pipes and slope α_i.

On the other hand, the nodal point may well be a model for a compressor station between $n = 2$ pipes having the same section $a_1 = a_2$, see Figure 1. In this case, the nodal condition consists of (10), while (11) is replaced, for instance, by

$$\Psi_2(t,u) = q_2 \left(\left(\frac{p(\rho_2)}{p(\rho_1)} - 1 \right)^{(\gamma-1)/\gamma} \right) - \Pi(t) \tag{13}$$

see [28, Section 3.1], [65, Section 4.4, Formula (4.9)] or [67]. Here, Π is proportional to the power exerted by the compressor. This framework naturally leads to various control problems, where the open–loop control $\Pi = \Pi(t)$ has to be chosen to satisfy suitable optimality criteria, see [10, 28, 43, 51, 52]. Closed–loop control problems, where Π depends on on the traces $u_1(t, x_*+)$ and $u_2(t, x_*+)$ of the state of the fluid, were studied in the case of smooth solutions in [42, 52].

For completeness, we recall that the above has been partly extended to the case of the full 3×3 system of Euler equations in [30, 32].

A drift–flux model for a two–phase gas was considered in [7] in the isothermal case under a no–slip assumption. Here, (1)–(2) can be used, with (5) and

$$u_j = \begin{bmatrix} \rho_j^1 \\ \rho_j^2 \\ (\rho_j^1 + \rho_j^2) v_j \end{bmatrix}, \qquad f_j(u_j) = \begin{bmatrix} \rho_j^1 \, v \\ \rho_j^2 \, v \\ (\rho_j^1 + \rho_j^2) \left((v_j)^2 + \frac{a^2}{2} \right) \end{bmatrix}, \qquad g(t,x,u) = 0\,.$$

Here, ρ_j^i is the density of the i-th component in the j-th pipe and v_j is the fluid speed, common to both phases in the j-th pipe. The sound speed a is assumed to be the same for all phases, see [6] for a more general setting. At the junction, the nodal condition (2) imposes the conservation of the total mass of each phase, similarly to (10). Besides, it is usually required that all velocity flows at the junction are equal, so that

$$\Psi_j(u) = \frac{1}{2} \left(v_j^2 - v_1^2 \right) + \frac{a^2}{2} \log \frac{\rho_j^1 + \rho_j^2}{\rho_1^1 + \rho_1^2} \qquad j = 2, \ldots, n\,. \tag{14}$$

The existence of weak solutions to (1)–(2)–(10)–(14) for constant initial data with separated wave speeds was obtained in [7, Proposition 3.1].

In order to capture effects of the 3D situation numerical integrations [53, 57] of multi–dimensional formulation of gas dynamics close to a nodal point have been performed. A 2D domain $\mathcal{D}$ models the nodal point (see Figure 2 for an example) and for $(x, y) \in \mathcal{D}$ and the usual $\gamma-$law for p the isentropic Euler equations

$$\partial_t \begin{bmatrix} \rho \\ \rho\, u \\ \rho\, v \end{bmatrix} + \partial_x \begin{bmatrix} \rho\, u \\ \rho\, u^2 + p(\rho) \\ \rho\, u\, v \end{bmatrix} + \partial_y \begin{bmatrix} \rho\, v \\ \rho\, u\, v \\ \rho\, v^2 + p(\rho) \end{bmatrix} = 0. \tag{15}$$

are solved numerically. For piecewise constant initial data comparisons of flow and pressure computations by nodal conditions (10)–(11) and (10), (12)with numerical averaging of solutions $(\rho, \rho u, \rho v)$ of (15) have been studied [57]. For initial data and small velocities $v_{j,0}$ the results should qualitatively similar behavior. Further studies exist for the Euler equations [53].

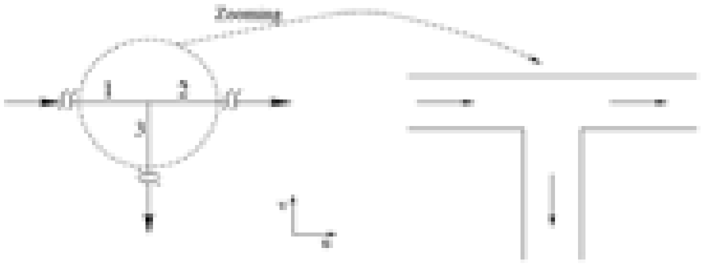


FIGURE 2. Domain $\mathcal{D}$ for numerical integration of gas dynamics equations in 2D resembling a nodal point.

2.2. **Traffic Flow on Road Networks.** The use of conservation laws in the modeling of traffic dynamics goes back to the pioneering works of Lighthill–Whitham [63] and Richards [66], introducing the (LWR) model (1) where $u \in [0, 1]$ is the averaged traffic density, $f(u) = u\, v(u)$ is the flow, $g = 0$ and the speed law $v = v(u)$ is a smooth decreasing function vanishing at the maximal density $u = 1$. The case of junctions was then considered for instance in [21, 59, 61], see also [45].

It is customary to distinguish the n roads impinging on the node located at $x_* = 0$ between l roads where vehicles move towards the node and $n-l$ roads where the cars exit. One is thus lead to a system of the type (1), namely

$$\left\{ \begin{array}{lll} \partial_t u_j + \partial_x \left(u_j\, v_j(u_j)\right) = 0 & x \in \mathbb{R}^- & j \in \{1, \ldots, l\} \\ \partial_t u_j + \partial_x \left(u_j\, v_j(u_j)\right) = 0 & x \in \mathbb{R}^+ & j \in \{l+1, \ldots, n\}\,. \end{array} \right. \tag{16}$$

The nodal condition (2) now prescribes the conservation of cars and the priority rules at the junction. The former reads

$$\Psi_1(u) = \sum_{j=1}^{l} u_j\, v_j(u_j) - \sum_{j=l+1}^{n} u_j\, v_j(u_j)\,,$$

while the latter take different forms. In an *autonomous* settings, they are chosen according to various maximality conditions. For example, it is commonly postulated that drivers tend to optimize the flow through the junction, see [21, 45]. These maximization principles remind of an entropy condition at the node, guaranteeing a well–posed Riemann problem (4). The presence of traffic light leads to a *non–autonomous* setting, where Ψ depends explicitly (typically periodically) on time, see [27]. A vanishing viscosity approach to the nodal condition (2) in traffic flow is developed in [20].

In order to have a better description of traffic flows, in particular of the dependence of the averaged velocity v_j upon the traffic density, several extensions of the LWR model were proposed. Among the so called *"second–order"*, we recall the family of Aw–Rascle–Zhang equations [5, 68], where $u_j = (\rho_j, \rho_j w_j)$ and the average vehicular velocity v_j is coupled to the traffic density ρ_j and to a property w_j by

$$w_j = v_j + p(\rho_j)\,.$$

The system (16) is thus replaced by the $(2n) \times (2n)$–system of type (1)

$$\left\{\begin{array}{llll} \partial_t \left[\begin{array}{c} \rho_j \\ \rho_j w_j \end{array}\right] + \partial_x \left[\begin{array}{c} \rho_j v_j \\ \rho_j v_j w_j \end{array}\right] = 0 & x \in \mathbb{R}^- & j \in \{1, \ldots, l\} \\ \partial_t \left[\begin{array}{c} \rho_j \\ \rho_j w_j \end{array}\right] + \partial_x \left[\begin{array}{c} \rho_j v_j \\ \rho_j v_j w_j \end{array}\right] = 0 & x \in \mathbb{R}^+ & j \in \{l+1, \ldots, n\}\,, \end{array}\right. \tag{17}$$

As in the LWR case, the nodal condition (2) ensures the conservation of vehicles, with

$$\Psi_1(u) = \sum_{j=1}^{l} \rho_j\, v_j - \sum_{j=l+1}^{n} \rho_j\, v_j\,, \tag{18}$$

and imposes further additional relations, for instance the maximization of the flow through the junction, as in [44].

Different coupling conditions for the Aw–Rascle–Zhang have been presented in [56]. Assume for now $n = 3$ and $l = 2$. Besides conservation of vehicles (18) the value of w is preserved through the node since it is a Lagrangian variable moving at velocity v. In the case $l = 2$ the conservation of w is achieved by modifying the pressure $p(\rho)$ on the exiting road ($j = 3$). In the case of Riemann data $u_{j,o}$ the new pressure $p^\dagger$ is given by the p for $x \geq t\ v_{3,o}$. For $x \leq t\ v_{3,o}$ the precise form of the function $p^\dagger$ depends on an additional condition on the mixture of cars. Assuming cars enter turn by turn from the two impinging roads ($j \in \{1, 2\}$) the following formulation is derived $p^\dagger(\rho) = \frac{1}{2}\,(w_{2,o} + w_{1,o}) - v^\dagger$. Herein, $v^\dagger$ solves $\frac{1}{\rho} = \frac{1}{2}\left(p^{-1}(\frac{1}{w_{1,o}-v}) + p^{-1}\left(\frac{1}{w_{2,o}-v}\right)\right)$ for fixed values $w_{i,o}$ [56, Section 6]. It has been shown that the coupling condition is given by (18) and

$$\Psi_2(u) = \sum_{j=1}^{l} \rho_j\, v_j\, w_j - \sum_{j=l+1}^{n} \rho_j\, v_j\, w_j.$$

A further development in traffic descriptions leads to multiphase models. It is now commonly accepted that any fundamental diagram $\rho \to \rho\, v(\rho)$ for the LWR model is reliabale at high speeds (low densities), but real data show that it becomes hardly acceptable at low speeds (high densities). From a macroscopic point of view, this leads to the introduction of different *phases*, typically the free and the congested one, see [12, 22, 24, 31]. Typically, these models display a dynamics of phases that very much resembles, for instance, that of liquid–vapor transitions. Different segments along the same roads can be in either phases, with vehicles entering and exiting these regions, crossing the moving phase boundaries. Therefore, suitable nodal conditions need to be prescribed along any phase boundary $x = x_*(t)$. These nodal conditions are rather implicit: in general, they are prescribed through the selection of a specific definition of solutions to Riemann problems at the phase boundary, rather than explicitly through a function Ψ like (3). Here, a Riemann problem at the boundary is a system of the form (4) with $n = 2$ and data in

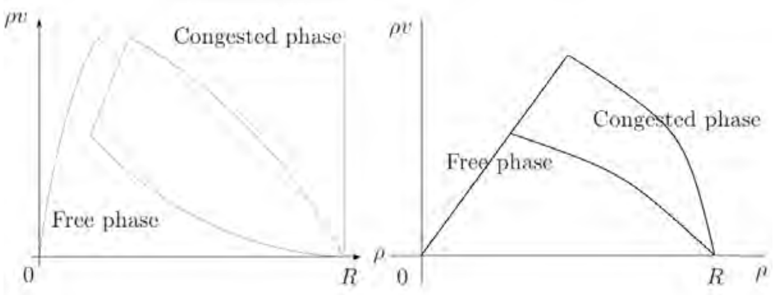


FIGURE 3. Left, the fundamental diagram $\rho \to \rho\, v$ of the model presented in [22] and, right, that of the models in [12, 31]. The maximal car density is assumed to be R.

two different phases, i.e., in the two connected component that constitute the u space, see for instance Figure 3. For a detailed treatment of the nodal conditions in phase transitions traffic model we defer to the cited literature [12, 22, 24, 31]. These models posed various new question, originating rich research directions at the strictly theoretical level, see [26], at the numerical level [2, 12, 18, 19, 58] and towards junctions or networks, see [25, 46]. In the latter case, they are supplied with further nodal conditions of the type (2) that first ensure the conservation of vehicles and then, as in the single phase case, prescribe the priorities rules at the junction.

Coupling conditions based on numerical integrations of fine–scale traffic flow models have been proposed [54]. The fine–scale traffic flow model including geometry effects at the node is given by a multi–lane description of traffic flow at a nodal point. Then, the closing or opening of a lane is used as approximation and simulations for different scenarios are performed.

2.3. **Product flow in supply chains.** A continuum description of the product density in high–volume production lines has been derived and analysed [3]. The governing equations are of the type (1), where

$$u = \begin{bmatrix} u_1 \\ \dots \\ u_n \end{bmatrix}, \qquad f(u) = \begin{bmatrix} f_1(u_1) \\ \dots \\ f_n(u_n) \end{bmatrix} \quad \text{and} \quad f_j(u_j) = \min\{v_j u_j, \mu_j\}\,. \tag{19}$$

Here, the index j refers to the supplier, u_j is the product density; v_j, respectively μ_j, is the non–negative constant production velocity, respectively capacity, see [4, 47, 48]. Similarly to traffic flow networks, we distinguish between the l suppliers delivering to the nodal point at $x_* = 0$ and the $n-l$ suppliers receiving parts from x_*. But differently from what happens in traffic modeling, due to possible differences in the total capacities of the different suppliers, the total mass is typically *not* conserved at the node. Therefore, the dynamics of the suppliers $1, \dots, l$ is coupled through *a priori* unknown new functions $q_1(t), \dots, q_l(t)$ describing the amount of parts waiting in suitable buffers when the lines they have to enter is congested. This leads to a time–dependent coupling condition of the type (2), namely

$$\Psi(t,u) = \int_{t_o}^{t} \left(\sum_{j=1}^{l} f_j\left(u_j(s,0-)\right) - \sum_{j=l+1}^{n} f_j\left(u_j(s,0+)\right) \right) ds - q_j(t) + q_j(t_o)\,. \tag{20}$$

Clearly, (20) is not sufficient to obtain a well–posed Riemann problem at the node and therefore is complemented by the $j = l+1, \ldots, n$ coupling conditions

$$\Psi_j(t,u) = q_j(t) - q_j(t_o) + \int_{t_o}^{t} \left(f_j\left(u_j(s,0+)\right) - \sum_{k=1}^{l} \alpha_{jk}(s) f_k\left(u_k(s,0-)\right) \right) ds \quad (21)$$

that prescribe how the parts are distributed among the outgoing lines. Indeed, the known parameter $\alpha_{jk}(t) \in [0,1]$ is the portion of parts flowing from the k-th line that have to enter the j-th one at time t, for $k = 1, \ldots, l$ and $j = l+1, \ldots, n$.

The well–posedness of the Riemann problem (4)–(19)–(20)–(21), as well as that of the corresponding Cauchy problem, is proved in [36, 55]. An extension of the model (19) has been proposed in [37, 40] to treat the case of supply chains with spatially and temporal depending capacities. Therein, $\mu_j = \mu_j(t,x)$ and equation (19) is replaced by $u_j = (\rho_j, \mu_j)$ and $f_j(u) = ((19), -\mu)$, respectively. Coupling conditions of the form (2) are proposed conserving total mass and the value of μ. Existence of solutions to the Cauchy problem for initial data with zero total variation in $\mu_{j,o}$ has been established [37, Theorem 3].

2.4. **Data flow on telecommunication networks.** The transport of data packages on the internet has been modeled using conservation laws on networks [38, 39]. Each transmission line corresponds to a link in the network connected at a nodal point at $x_* = 0$. A model for the data package density $\rho_j \geq 0$ and source destination information π_j for the packages is

$$u_j = \begin{bmatrix} \rho_j \\ \pi_j \end{bmatrix}, \; f_j(u) = \begin{bmatrix} \bar{f}_j(\rho) \\ \bar{f}_j(\rho)/\rho \end{bmatrix} \text{ and } \; \bar{f}_j(\rho) := \begin{cases} \bar{v}\,\rho & 0 \leq \rho \leq \sigma_j \\ \bar{v}\,\sigma_j \frac{\rho_{\max}-\rho}{\rho_{\max}-\sigma_j} & \sigma_j \leq \rho \leq \rho_{\max} \end{cases},$$

where $\rho_{\max}$ is the maximal package density on each link, $\bar{v} > 0$ a fixed transportation velocity and $\sigma_j > 0$ a parameter corresponding to the probability of package losses during transmission. A variety of coupling conditions have been discussed [39, Section 4] ensuring in particular the conservation of mass at the node similar to (2.2). This condition is not sufficient to guarantee well–posedness of the Riemann problem and as in vehicular traffic flow additional conditions have to be imposed.

2.5. **Networks of open canals.** Water flow in open canals [11, 51, 62] can be described by the St. Venant equations in on space dimension. When n canals enter or exit the same origin, say $x_* = 0$, we are lead to (1) with (5) and

$$u_j = \begin{bmatrix} a_j \\ v_j \end{bmatrix}, \qquad f_j(u) = \begin{bmatrix} a_j \\ \frac{1}{2}\,{v_j}^2 + \bar{g}\,h(a_j) \end{bmatrix}, \qquad g_j(u) = \begin{bmatrix} 0 \\ -\bar{g}\,h(a_j) \end{bmatrix}, \qquad (22)$$

where v_j is the water speed in the j-th canal, a_j is the vertical cross section occupied by the water, g is gravity, and $H = h(a)$ is the water level. Condition (2) takes here the form

$$\Psi_1(u) = \sum_{j=1}^{n} a_j\,v_j, \qquad \Psi_j(u) = \frac{1}{2}\,{v_j}^2 + \bar{g}\,h(a_j) - \frac{1}{2}\,{v_1}^2 - \bar{g}\,h(a_1) \qquad j = 2, \ldots, n$$

see [29, 62], ensuring that the conservation of water and the Bernoulli law are satisfied at the junction.

Consider now $n = 2$ canals, the first entering into the second at, say, $x_* = 0$, where an underflow gate regulates the water levels. A different choice of the

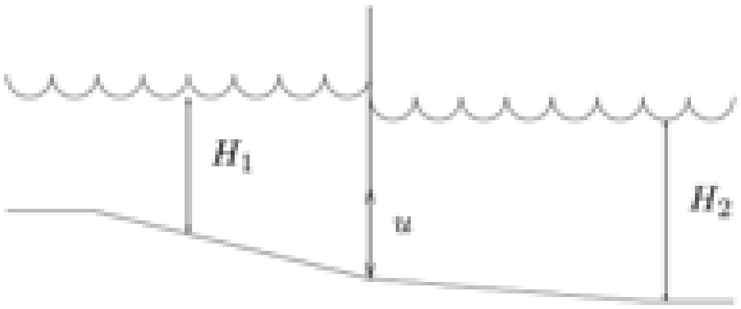


FIGURE 4. An underflow gate with opening $\mathbf{u} = \Pi$. The water height in the two connected canals is indicated by H_i.

unknown variables leads to (1)–(5) with

$$\begin{aligned} u_j &= \begin{bmatrix} H_j \\ q_j \end{bmatrix}, \qquad f_j(u) = \begin{bmatrix} q_j \\ \frac{{q_j}^2}{H_j} + \frac{1}{2}\,\bar{g}\,{H_j}^2 \end{bmatrix}, \\ g_j(t,x,u) &= \begin{bmatrix} 0 \\ -\bar{g}\,h_j\,\sin\alpha_j(x) - \nu(x)\,\frac{q_j\,|q_j|}{H_j} \end{bmatrix} \end{aligned} \tag{23}$$

for $j = 1,2$, see Figure 4. Here, with reference to the j-th canal, q_j is the flow of water, H_j the water level, $\alpha_j(x)$ the bed slope, $\bar{g}$ gravity and ν a friction coefficient, see [28, 41, 51]. In the nodal condition (2) we obtain

$$\Psi_1(t,u) = a_1\,q_1 - a_2\,q_2, \qquad \Psi_2(t,u) = \frac{{q_1}^2}{H_1 - H_2} - \Pi(t).$$

In the second component, the time varying term $\Pi = \Pi(t)$ is related to the height of the underflow gate, meaning that for $\Pi = 0$ water can not flow through the gate. The well posedness of the resulting system is proved in [28], together with the L^1–Lipshitz continuity of weak solutions with respect to variations in u. As a consequence, when an L^1–continuous cost integral is selected, the existence of optimal control problems follows. For the same system, closed–loop controls are considered in [11].

In [49] the shallow water equations (23) have been used to describe flow over a weir and pooled stepped chutes, see Figure 5. Those are found typically next to large dams in order to release overflowing water. A coupling condition based on the conservation of the overspill and based on energy losses due to change of potential energy during the spill is

$$\Psi_1(u) = a_1\,q_1 - a_2\,q_2,\ \Psi_2(u) = a_1 q_1 - C\left((h_1 - H_1)_+ - (h_2 - H_2)_+\right)^{\frac{3}{2}}. \tag{24}$$

Herein, $(z)_+ = \max(z,0)$, $C = 0.6\sqrt{\bar{g}}$, $H_{1,2}$ are the heights of the weir in the respective chutes and the other quantities are as before. Well-posedness results for the Riemann problem for large data as well as well posedness of the Cauchy problem has been established [49, Theorem 3.1].

3. **Nodal Points in Mixed ODEs–PDEs Systems.** When nodal conditions are coupled with ordinary differential equations, the conditions allowing to prove

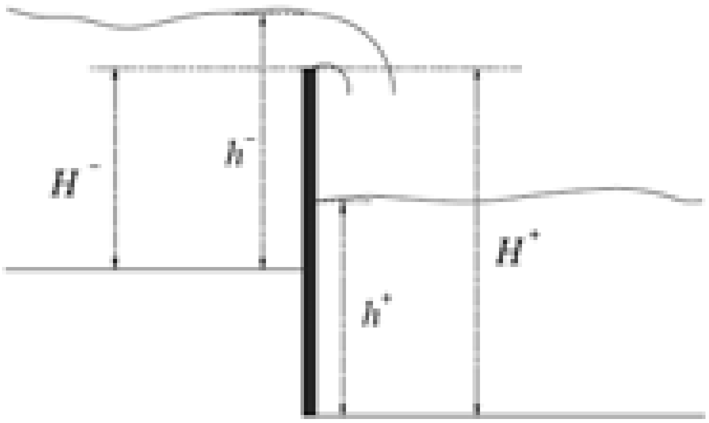


FIGURE 5. Schematic of two connected pooled steps with a weir in between, The weir has height H^+ and the upper pooled step has an elevation of $H^+ - H^-$ compared with the lower pooled step. The water heights are $h_{1,2} = h^\pm$.

well posedness are more intricate. Moreover, key questions about the existence of solutions globally in time remain mostly unanswered.

3.1. **Fluid-Solid Interaction.** An inviscid compressible fluid fills a vertical pipe with uniform section. A disc of mass m is free to move in the pipe, subject to gravity $\bar{g}$, to the interaction with the fluid and with the pipe walls. The disc prevents any flow of mass through its location. It is then natural to describe the fluid on the two sides of the disc by means of the p-system and the fluid–solid interaction through (2). More precisely, we let $n = 2$ and use (1)–(5)–(9) with

$$\Psi(x_*, \dot{x}_*, \ddot{x}_*, u_1, u_2) = \begin{bmatrix} \dot{x}_* - q_1/\rho_1 \\ \dot{x}_* - q_2/\rho_2 \\ \ddot{x}_* + \bar{g} + \frac{1}{m}\left(p(\rho_2) - p(\rho_1)\right) \end{bmatrix}. \tag{25}$$

The requirement $\Psi_1\left(x_*(t), \dot{x}_*(t), \ddot{x}_*(t), u_1\left(t, x_*(t)+\right), u_2\left(t, x_*(t)-\right)\right) = 0$, respectively $\Psi_2\left(x_*(t), \dot{x}_*(t), \ddot{x}_*(t), u_1\left(t, x_*(t)+\right), u_2\left(t, x_*(t)-\right)\right) = 0$, ensures that the speed of the disc equals that of the fluid above, respectively below, to it. The effects of gravity and of the difference in the fluid pressure between the two sides of the disc are described by $\Psi_3\left(x_*(t), \dot{x}_*(t), \ddot{x}_*(t), u_1\left(t, x_*(t)+\right), u_2\left(t, x_*(t)-\right)\right) = 0$. The resulting model (1)–(5)–(9)–(25) is able to describe the effects of a shock hitting the disc, we refer to [14] for (local in time) well posedness results and numerical integrations of this system.

3.2. **A Manhole in a Sewerage System.** Consider a vertical manhole that disposes the water its collects through two horizontal pipes at its bottom, see Figure 6. Call a_i the wet cross sectional area, $h_i = h_i(a_i)$ is the corresponding height and q_i the water flow in the i-th tube. A given function $p_i = p_i(a_i)$ describes the water pressure, possibly through the Preissman slot. The height of water in the manhole is $x_* = x_*(t)$ and A_M is th manhole sectional area. We describe the full system

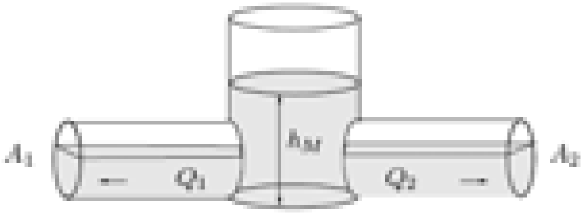


FIGURE 6. A vertical manhole with two horizontal tubes that exit from it.

setting $n = 2$ and by means of (1)–(3)–(5) with, formally similarly to (22),

$$u_j = \begin{bmatrix} a_j \\ q_j \end{bmatrix} \quad \text{and} \quad f_j(u_j) = \begin{bmatrix} q_j \\ \frac{{q_j}^2}{a_j} + p(a_j) \end{bmatrix}. \tag{26}$$

The nodal condition (3) both ensures the conservation of water while passing from the manhole to the pipes, the conservation of energy and provides a differential equation for the water level in the manhole:

$$\Psi(t, x_*, \dot{x}_*, u_1, u_2) = \begin{bmatrix} \frac{{q_1}^2}{{a_1}^2} - \frac{{q_2}^2}{{a_2}^2} + 2\bar{g}\left(h_1(a_1) - h_2(a_2)\right) \\ \frac{{q_1}^2}{{a_1}^2} + 2\bar{g}\, h_1(a_1) - \frac{(q_1+q_2)|q_1+q_2|}{{A_M}^2} \\ \dot{x}_* - \frac{1}{A_M}\left(Q(t) - q_1 - q_2\right) \end{bmatrix}$$

where Q is the water inflow in the manhole, see [13] for more details.

REFERENCES

[1] F. Ancona and A. Marson. Existence theory by front tracking for general nonlinear hyperbolic systems. *Arch. Ration. Mech. Anal.*, 185(2):287–340, 2007.

[2] B. Andreianov, P. Goatin, and N. Seguin. Finite volume schemes for locally constrained conservation laws. *Numer. Math.*, 115(4):609–645, 2010. With supplementary material available online.

[3] D. Armbruster, P. Degond, and C. Ringhofer. A model for the dynamics of large queuing networks and supply chains. *SIAM J. Applied Mathematics*, 66(3):896–920, 2006.

[4] D. Armbruster, S. Göttlich, and M. Herty. A continuous model for supply chains with finite buffers. *SIAM J. Appl. Math.*, 2011.

[5] A. Aw and M. Rascle. Resurection of second order models of traffic flow. *SIAM J. Appl. Math.*, 60:916–944, 2000.

[6] M. Banda, M. Herty, and J. Ngnotchouye. Coupling drift-flux models with uneqal sonic speeds. *Mathematical and Computational Applications*, 15(4):574–584, 2010.

[7] M. Banda, M. Herty, and J. Ngnotchouye. Toward a mathematical analysis for drift-flux multiphase flow models in networks. *SIAM Journal on Scientific Computing*, 31(6):4633–4653, 2010.

[8] M. K. Banda, M. Herty, and A. Klar. Coupling conditions for gas networks governed by the isothermal Euler equations. *Netw. Heterog. Media*, 1(2):295–314 (electronic), 2006.

[9] M. K. Banda, M. Herty, and A. Klar. Gas flow in pipeline networks. *Netw. Heterog. Media*, 1(1):41–56, 2006.

[10] G. Bastin, J. Coron, and B. d'Andrea Novel. Boundary feedback control and lyapunov stability analysis for physical networks of 2× 2 hyperbolic balance laws. In *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, pages 1454–1458. IEEE, 2008.

[11] G. Bastin, J. Coron, and B. dAndréa Novel. Using hyperbolic systems of balance laws for modeling, control and stability analysis of physical networks. In *Lecture notes for the Pre-congress workshop on complex embedded and networked control systems, 17th IFAC World Congress*, 2008.

[12] S. Blandin, D. Work, P. Goatin, B. Piccoli, and A. Bayen. A general phase transition model for vehicular traffic. *SIAM J. Appl. Math.*, 71(1):107–127, 2011.

[13] R. Borsche, R. M. Colombo, and M. Garavello. On the coupling of systems of hyperbolic conservation laws with ordinary differential equations. *Nonlinearity*, 23(11):2749–2770, 2010.

[14] R. Borsche, R. M. Colombo, and M. Garavello. Mixed systems: ODEs - balance laws. *J. Differential Equations*, 252(3):2311–2338, 2012.

[15] A. Bressan. *Hyperbolic systems of conservation laws*, volume 20 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2000. The one-dimensional Cauchy problem.

[16] A. Bressan and R. M. Colombo. The semigroup generated by 2 × 2 conservation laws. *Arch. Rational Mech. Anal.*, 133(1):1–75, 1995.

[17] A. Bressan, T.-P. Liu, and T. Yang. L^1 stability estimates for $n \times n$ conservation laws. *Arch. Ration. Mech. Anal.*, 149(1):1–22, 1999.

[18] C. Chalons and P. Goatin. Computing phase transitions arising in traffic flow modeling. In *Hyperbolic problems: theory, numerics, applications*, pages 559–566. Springer, Berlin, 2008.

[19] C. Chalons and P. Goatin. Godunov scheme and sampling technique for computing phase transitions in traffic flow modeling. *Interfaces Free Bound.*, 10(2):197–221, 2008.

[20] G. M. Coclite and M. Garavello. Vanishing viscosity for traffic on networks. *SIAM J. Math. Anal.*, 42(4):1761–1783, 2010.

[21] G. M. Coclite, M. Garavello, and B. Piccoli. Traffic flow on a road network. *SIAM J. Math. Anal.*, 36(6):1862–1886 (electronic), 2005.

[22] R. M. Colombo. Hyperbolic phase transitions in traffic flow. *SIAM J. Appl. Math.*, 63(2):708–721 (electronic), 2002.

[23] R. M. Colombo and M. Garavello. On the Cauchy problem for the p-system at a junction. *SIAM J. Math. Anal.*, 39(5):1456–1471, 2008.

[24] R. M. Colombo and P. Goatin. Traffic flow models with phase transitions. *Flow, turbulence and combustion*, 76(4):383–390, 2006.

[25] R. M. Colombo, P. Goatin, and B. Piccoli. Road networks with phase transitions. *J. Hyperbolic Differ. Equ.*, 7(1):85–106, 2010.

[26] R. M. Colombo, P. Goatin, and F. S. Priuli. Global well posedness of traffic flow models with phase transitions. *Nonlinear Analysis: Theory, Methods & Applications*, 66(11):2413–2426, 2007.

[27] R. M. Colombo, P. Goatin, and M. D. Rosini. On the modelling and management of traffic. *ESAIM Math. Model. Numer. Anal.*, 45(5):853–872, 2011.

[28] R. M. Colombo, G. Guerra, M. Herty, and V. Schleper. Optimal control in networks of pipes and canals. *SIAM J. Control Optim.*, 48(3):2032–2050, 2009.

[29] R. M. Colombo, M. Herty, and V. Sachers. On 2 × 2 conservation laws at a junction. *SIAM J. Math. Anal.*, 40(2):605–622, 2008.

[30] R. M. Colombo and F. Marcellini. Coupling conditions for the 3 × 3 Euler system. *Netw. Heterog. Media*, 5(4):675–690, 2010.

[31] R. M. Colombo, F. Marcellini, and M. Rascle. A 2-phase traffic model based on a speed bound. *SIAM J. Appl. Math.*, 70(7):2652–2666, 2010.

[32] R. M. Colombo and C. Mauri. Euler system at a junction. *Journal of Hyperbolic Differential Equations*, 5(3):547–568, 2007.

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

[34] C. M. Dafermos. *Hyperbolic conservation laws in continuum physics*, volume 325 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2005.

[35] C. M. Dafermos and L. Hsiao. Hyperbolic systems and balance laws with inhomogeneity and dissipation. *Indiana Univ. Math. J.*, 31(4):471–491, 1982.

[36] C. D'Apice, S. Göttlich, M. Herty, and B. Piccoli. *Modeling, simulation, and optimization of supply chains*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2010. A continuous approach.

[37] C. D'Apice and R. Manzo. A fluid dynamic model for supply chains. *Network and Heterogenous Media*, 3(1):379–398, 2006.

[38] C. D'Apice, R. Manzo, and B. Piccoli. Packet flow on telecommunication networks. *SIAM J. Math. Anal.*, 38(3):717–740, 2006.

[39] C. D'Apice, R. Manzo, and B. Piccoli. A fluid dynamic model for telecommunication networks with sources and destinations. *SIAM J. Appl. Math.*, 68(4):981–1003, 2008.

[40] C. D'Apice, R. Manzo, and B. Piccoli. Existence of solutions to Cauchy problems for a mixed continuum-discrete model for supply chains and networks. *J. Math. Anal. Appl.*, 362(2):374–386, 2010.

[41] J. de Halleux, C. Prieur, J. Coron, B. d'Andréa Novel, and G. Bastin. Boundary feedback control in networks of open channels. *Automatica*, 39(8):1365–1376, 2003.

[42] M. Dick, M. Gugat, and G. Leugering. Classical solutions and feedback stabilization for the gas flow in a sequence of pipes. *Netw. Heterog. Media*, 5(4):691–709, 2010.

[43] M. Dick, M. Gugat, and G. Leugering. A strict H^1-Lyapunov function and feedback stabilization for the isothermal Euler equations with friction. *Numer. Algebra Control Optim.*, 1(2):225–244, 2011.

[44] M. Garavello and B. Piccoli. Traffic flow on a road network using the Aw-Rascle model. *Comm. Partial Diff. Equ.*, 31(1-3):243–275, 2006.

[45] M. Garavello and B. Piccoli. *Traffic flow on networks*, volume 1 of *AIMS Series on Applied Mathematics*. American Institute of Mathematical Sciences (AIMS), Springfield, MO, 2006. Conservation laws models.

[46] P. Goatin. Traffic flow models with phase transitions on road networks. *Netw. Heterog. Media*, 4(2):287–301, 2009.

[47] S. Göttlich, M. Herty, and A. Klar. Network models for supply chains. *Commun. Math. Sci.*, 3(4):545–559, 2005.

[48] S. Göttlich, M. Herty, and A. Klar. Modelling and optimization of supply chains on complex networks. *Commun. Math. Sci.*, 4(2):315–330, 2006.

[49] G. Guerra, M. Herty, and F. Marcellini. Modeling and analysis of pooled stepped chutes. *Netw. Heterog. Media*, 6(4):665–679, 2011.

[50] G. Guerra, F. Marcellini, and V. Schleper. Balance laws with integrable unbounded sources. *SIAM J. Math. Anal.*, 41(3):1164–1189, 2009.

[51] M. Gugat. Nodal control of conservation laws on networks. *in Control and Boundary Analysis, Cagnol, J. and Zolesio, J.-P. (eds), Chapman & Hall/CRC, Boca Raton, FL*, 2005.

[52] M. Gugat, M. Herty, and V. Schleper. Flow control in gas networks: exact controllability to a given demand. *Math. Methods Appl. Sci.*, 34(7):745–757, 2011.

[53] M. Herty. Coupling conditions for networked systems of Euler equations. *SIAM J. Sci. Comput.*, 30(3):1596–1612, 2008.

[54] M. Herty and A. Klar. Modeling, simulation, and optimization of traffic flow networks. *SIAM J. Sci. Comput.*, 25(3):1066–1087, 2003.

[55] M. Herty, A. Klar, and B. Piccoli. Existence of solutions for supply chain models based on partial differential equations. *SIAM J. Math. Anal.*, 39(1):160–173, 2007.

[56] M. Herty and M. Rascle. Coupling conditions for a class of second-order models for traffic flow. *SIAM J. Math. Anal.*, 38(2):595–616, 2006.

[57] M. Herty and M. Seaïd. Simulation of transient gas flow at pipe-to-pipe intersections. *Internat. J. Numer. Methods Fluids*, 56(5):485–506, 2008.

[58] M. Herty, M. Seaïd, and A. K. Singh. A domain decomposition method for conservation laws with discontinuous flux function. *Appl. Numer. Math.*, 57(4):361–373, 2007.

[59] H. Holden and N. H. Risebro. A mathematical model of traffic flow on a network of unidirectional roads. *SIAM J. Math. Anal.*, 26(4):999–1017, 1995.

[60] H. Holden and N. H. Risebro. Riemann problems with a kink. *SIAM J. Math. Anal.*, 30(3):497–515 (electronic), 1999.

[61] J.-P. Lebacque. Les modeles macroscopiques du traffic. *Annales des Ponts.*, 67:24–45, 1993.

[62] G. Leugering and J. Schmidt. On the modelling and stabilization of flows in networks of open canals. *SIAM journal on control and optimization*, 41:164, 2002.

[63] M. J. Lighthill and G. B. Whitham. On kinematic waves. II. A theory of traffic flow on long crowded roads. *Proc. Roy. Soc. London. Ser. A.*, 229:317–345, 1955.

[64] T.-P. Liu and T. Yang. A new entropy functional for a scalar conservation law. *Comm. Pure Appl. Math.*, 52(11):1427–1442, 1999.

[65] E. Menon. *Gas Pipeline Hydraulics*. Taylor & Francis, Boca Raton, 2005.

[66] P. I. Richards. Shock waves on the highway. *Operations Research*, 4:42–51, 1956.

[67] M. C. Steinbach. On PDE solution in transient optimization of gas networks. *J. Comput. Appl. Math.*, 203(2):345–361, 2007.

[68] H. M. Zhang. A non-equilibrium traffic model devoid of gas-like behaviour. *Transportation Research Part B*, 36:275–298, 2002.

E-mail address: rinaldo@ing.unibs.it
E-mail address: herty@igpm.rwth-aachen.de

RELATIVE ENTROPY METHODS FOR HYPERBOLIC AND DIFFUSIVE LIMITS

Corrado Lattanzio

Dipartimento di Ingegneria e Scienze dell'Informazione e Matematica
Università degli Studi dell'Aquila
Via Vetoio, I 67010 Coppito (L'Aquila) AQ, Italy

Athanasios E. Tzavaras

Department of Applied Mathematics
University of Crete
GR 71409 Heraklion, Crete, Greece
and
Institute for Applied and Computational Mathematics
Foundation for Research and Technology
GR 70013 Heraklion, Crete, Greece

Abstract. We review the relative entropy method in the context of hyperbolic and diffusive relaxation limits of entropy solutions for various hyperbolic models. The main example consists of the convergence from multidimensional compressible Euler equations with friction to the porous medium equation [7]. With small modifications, the arguments used in that case can be adapted to the study of the diffusive limit from the Euler-Poisson system with friction to the Keller-Segel system [8]. In addition, the p–system with friction and the system of viscoelasticity with memory are then reviewed, again in the case of diffusive limits [7]. Finally, the method of relative entropy is described for the multidimensional stress relaxation model converging to elastodynamics [6, Section 3.2], one of the first examples of application of the method to hyperbolic relaxation limits.

1. **Introduction.** The relative entropy method was introduced in a context of hyperbolic systems by Dafermos and DiPerna [3, 2, 5] and serves as a mathematical tool for studying stability and limiting processes among thermomechanical theories. The method consists of a direct calculation of the relative entropy between a weak, *entropy dissipative* solution and a smooth, *entropy conservative* solution for the underlying thermomechanical processes, and leads to a striking stability formula. The same approach can be used to control hyperbolic relaxation limits [6, 9, 1] as well as diffusive relaxation [7]. The novelty in the latter case lies in the fact that, when dealing with a diffusive limit, the relative entropy method aims to compare weak, *entropy dissipative* solutions of the approximating hyperbolic system with smooth yet *entropy dissipative* solutions of the limit. Therefore, in order to prove that the relative entropy can serve as a Lyapunov–type functional for the model,

2000 *Mathematics Subject Classification.* 35L60, 35L65, 35Q31, 35B25, 76N15.

Key words and phrases. Relative entropy, relaxation limits, Euler equations, Keller-Segel systems, viscoelasticity.

Joint research partially supported by the EU FP7 REGPOT project *"Archimedes Center for Modeling, Analysis and Computation"*. AET partially supported by the Thales program of the Greek Secretariat for Research and Technology.

one has in that case to also control the dissipation of the limit diffusive equation in terms of the dissipation of the approximating system.

In the present paper we review some examples of diffusive relaxation analyzed in [7]: the case of 3–d isentropic gas dynamics with friction in Eulerian coordinates (Section 2), 1–d p–system with friction (Section 4), and 1–d viscoelasticity of the memory type converging to viscoelasticity of the rate type (Section 5.1). In addition to these cases, in Section 3 we shall apply this technique to the Euler-Poisson system with friction in a diffusive scaling, which has been treated in [4] by means of energy methods and compensated compactness tools.

Finally, in Section 5.2, we shall review a result from [6, Sec 3.2] concerning the convergence from viscoelasticity of the memory type to the equations of elastodynamics. This is one of the first examples in the literature where the technique of relative entropy has been utilized in the context of hyperbolic relaxation limits. The general framework of such singular limits has been studied in [9], while the corresponding analysis for general diffusive relaxation limits is still an open problem.

2. **Isentropic gas dynamics in Eulerian coordinates with friction.**

As an example to test the relative entropy method in the context of diffusive relaxation, let us consider the (scaled w.r.t. a diffusive scaling) system of isentropic gas dynamics with friction in three space dimensions:

$$\begin{cases} \rho_t + \dfrac{1}{\varepsilon}\operatorname{div}_x m = 0 \\ m_t + \dfrac{1}{\varepsilon}\operatorname{div}_x \dfrac{m\otimes m}{\rho} + \dfrac{1}{\varepsilon}\nabla_x p(\rho) = -\dfrac{1}{\varepsilon^2}m, \end{cases} \tag{1}$$

where $t \in \mathbb{R}$, $x \in \mathbb{R}^3$, the density $\rho \geq 0$ and the momentum $m \in \mathbb{R}^3$. At this level, the pressure $p(\rho)$ is a general function satisfying $p'(\rho) > 0$ so that (1) is hyperbolic. The usual example of pressures verifying all needed conditions is given by the γ–laws: $p(\rho) = k\rho^\gamma$ with $\gamma \geq 1$ and $k > 0$. The (formal) diffusive relaxation limit $\varepsilon \to 0$ yields the porous media equation

$$\bar\rho_t - \triangle_x p(\bar\rho) = 0\,. \tag{2}$$

In the sequel, we shall establish this limit via the relative entropy method.

An example of an entropy pair for (1) is given by the mechanical energy

$$\eta(\rho, m) = \frac{1}{2}\frac{|m|^2}{\rho} + h(\rho)$$

and the associated flux of mechanical work

$$q(\rho, m) = \frac{1}{2}m\frac{|m|^2}{\rho^2} + mh'(\rho)\,,$$

where $h(\rho) = \rho e(\rho)$ with $e(\rho)$ the internal energy of the gas:

$$e'(\rho) = \frac{p(\rho)}{\rho^2}; \quad h''(\rho) = \frac{p'(\rho)}{\rho}; \quad \rho h'(\rho) = p(\rho) + h(\rho).$$

For the particular case of γ–law gases, h takes the form

$$h(\rho) = \begin{cases} \dfrac{k}{\gamma-1}\rho^\gamma = \dfrac{1}{\gamma-1}p(\rho) & \text{for } \gamma > 1\,, \\ k\rho\log\rho & \text{for } \gamma = 1. \end{cases}$$

Smooth solutions of (1) satisfy the energy identity

$$\eta(\rho,m)_t + \frac{1}{\varepsilon}\operatorname{div}_x q(\rho,m) = -\frac{1}{\varepsilon^2}\nabla_m\eta(\rho,m)\cdot m = -\frac{1}{\varepsilon^2}\frac{|m|^2}{\rho} \leq 0\,,$$

which reveals the dissipative nature of the mechanical energy $\eta(\rho,m)$ along the process (1).

Let us now consider a weak solution (ρ,m) of (1) that satisfies the weak form of the entropy inequality,

$$\eta(\rho,m)_t + \frac{1}{\varepsilon}\operatorname{div}_x q(\rho,m) + \frac{1}{\varepsilon^2}\frac{|m|^2}{\rho} \leq 0\,, \tag{3}$$

and let $\bar\rho \geq 0$ be a smooth solution of the porous media equation (2). Clearly, $\bar\rho$ will also satisfy an energy dissipation identity of the form

$$h(\bar\rho)_t - \operatorname{div}_x\big(h'(\bar\rho)\nabla_x p(\bar\rho)\big) = -\frac{|\nabla_x p(\bar\rho)|^2}{\bar\rho} \leq 0\,.$$

Thanks to the relative entropy method, we obtain an identity that monitors the distance between ρ and $\bar\rho$. Such identities have been obtained via the relative entropy method for hyperbolic relaxation in [6, 9, 1], while the first results in a diffusive relaxation framework are derived in [7].

We recall that the relative entropy is defined as the quadratic part of the Taylor series expansion between two solutions (ρ,m) and $(\bar\rho,\bar m)$:

$$\begin{aligned}\eta(\rho,m\,|\bar\rho,\bar m\,) &:= \eta(\rho,m) - \eta(\bar\rho,\bar m) - \eta_\rho(\bar\rho,\bar m)(\rho-\bar\rho) - \nabla_m\eta(\bar\rho,\bar m)\cdot(m-\bar m)\\ &= \frac{1}{2}\rho\left|\frac{m}{\rho}-\frac{\bar m}{\bar\rho}\right|^2 + h(\rho\,|\bar\rho\,)\,,\end{aligned} \tag{4}$$

while the corresponding relative entropy-flux reads

$$\begin{aligned}q_i(\rho,m\,|\bar\rho,\bar m\,) &:= q_i(\rho,m) - q_i(\bar\rho,\bar m) - \eta_\rho(\bar\rho,\bar m)(m_i-\bar m_i)\\ &\qquad - \nabla_m\eta(\bar\rho,\bar m)\cdot(f_i(\rho,m)-f_i(\bar\rho,\bar m))\\ &= \frac{1}{2}m_i\left|\frac{m}{\rho}-\frac{\bar m}{\bar\rho}\right|^2 + \rho(h'(\rho)-h'(\bar\rho))\Big(\frac{m_i}{\rho}-\frac{\bar m_i}{\bar\rho}\Big) + \frac{\bar m_i}{\bar\rho}h(\rho\,|\bar\rho\,)\,,\end{aligned} \tag{5}$$

where $i=1,2,3$, f_i stands for the components of the (vector valued) flux in (1),

$$f_i(\rho,m) = m_i\frac{m}{\rho} + p(\rho)I_i\,,$$

and I_i is the i–th column of the 3×3 identity matrix.

As noticed in [7], the novelty in the case of diffusive relaxation lies mainly in the selection of the momentum $\bar m$ in (4) and in (5). Indeed $\bar m$ is chosen to adapt itself in the relaxation, what allows to handle a diffusive relaxation process, where both solutions that are compared are energy dissipative. More precisely, we choose $\bar m = -\varepsilon\nabla_x p(\bar\rho)$ and we rewrite (2) in the form of the system of Euler equations with relaxation, plus additional higher–order error terms:

$$\begin{aligned}\bar\rho_t + \frac{1}{\varepsilon}\partial_{x_i}\bar m_i &= 0\\ \bar m_t + \frac{1}{\varepsilon}\partial_{x_i}f_i(\bar\rho,\bar m) &= -\frac{1}{\varepsilon^2}\bar m + e(\bar\rho,\bar m)\,,\end{aligned} \tag{6}$$

where (we use the convention of summation over repeated indices and) $\bar{e}$ is given by

$$\begin{aligned}\bar{e} := e(\bar{\rho},\bar{m}) &= \frac{1}{\varepsilon}\operatorname{div}_x\left(\frac{\bar{m}\otimes\bar{m}}{\bar{\rho}}\right) - \varepsilon\partial_t\nabla_x p(\bar{\rho}) \\ &= \varepsilon\operatorname{div}_x\left(\frac{\nabla_x p(\bar{\rho})\otimes\nabla_x p(\bar{\rho})}{\bar{\rho}}\right) - \varepsilon\nabla_x(p'(\bar{\rho})\triangle_x p(\bar{\rho})) \\ &= O(\varepsilon)\,. \end{aligned}\tag{7}$$

Thanks to the aforementioned rewriting of the limiting equation (2), it is possible to analyze the relative entropy (4) and to prove the following result [7].

Proposition 2.1. *Let (ρ, m) be a weak entropy solution of (1) satisfying (3) and let $(\bar{\rho},\bar{m})$ be a smooth solution of (6). Then,*

$$\partial_t\eta(\rho,m\,|\bar{\rho},\bar{m}\,) + \frac{1}{\varepsilon}\operatorname{div}_x q(\rho,m\,|\bar{\rho},\bar{m}\,) \le -\frac{1}{\varepsilon^2}R(\rho,m\,|\bar{\rho},\bar{m}\,) - Q - E\,,\tag{8}$$

where

$$\begin{aligned} R(\rho,m\,|\bar{\rho},\bar{m}\,) &= \rho\left|\frac{m}{\rho}-\frac{\bar{m}}{\bar{\rho}}\right|^2\,, \\ Q &= \frac{1}{\varepsilon}\nabla^2_{(\rho,m)}\eta(\bar{\rho},\bar{m})\begin{pmatrix}\bar{\rho}_{x_i}\\ \bar{m}_{x_i}\end{pmatrix}\cdot\begin{pmatrix}0\\ f_i(\rho,m|\bar{\rho},\bar{m})\end{pmatrix}\,, \\ E &= e(\bar{\rho},\bar{m})\cdot\frac{\rho}{\bar{\rho}}\left(\frac{m}{\rho}-\frac{\bar{m}}{\bar{\rho}}\right)\,, \end{aligned}\tag{9}$$

and $e(\bar{\rho},\bar{m})$ is defined in (7).

Concerning the relative entropy estimate (8), we point out that the coefficient of the quadratic term Q depends only on $(\bar{\rho},\bar{m})$ and it is $O(1)$ in ε:

$$\frac{1}{\varepsilon}\Big(\eta_{\rho m_j}(\bar{\rho},\bar{m})\bar{\rho}_{x_i} + \eta_{m_k m_j}(\bar{\rho},\bar{m})\partial_{x_i}\bar{m}_k\Big) = \frac{1}{\varepsilon}\partial_{x_i}\left(\frac{\bar{m}_j}{\bar{\rho}}\right) = -\partial_{x_i x_j}h'(\bar{\rho})\,,$$

while the term E is an error term of order $O(\varepsilon)$. The term $R(\rho,m\,|\bar{\rho},\bar{m}\,)$ captures the dissipation of the relaxation system (1) relative to its diffusive scale limit (2). It turns out to be the *quadratic part of the dissipative relaxation term with respect to* $(\bar{\rho},\bar{m})$, justifying the notation in (9). Clearly, the property $R(\rho,m\,|\bar{\rho},\bar{m}\,)\ge 0$ is crucial in the stability analysis of the relaxation process.

An example of a framework in which to apply the relative entropy identity (8) is that of multidimensional periodic solutions, referred to as $(\mathbf{H_1})$:

(i) $(\rho\,,\,m) : (0,T)\times\mathbb{T}^3\to\mathbb{R}^4$ is a (periodic) *dissipative weak solution* of (1) with $\rho\ge 0$, satisfying the weak form of (1) and the integrated form of the entropy inequality (3):

$$\begin{aligned}\iint_{[0,+\infty)\times\mathbb{T}^3}&\left[\left(\frac{1}{2}\frac{|m|^2}{\rho}+h(\rho)\right)\dot{\theta}(t) - \frac{1}{\varepsilon^2}\frac{|m|^2}{\rho}\theta(t)\right]dxdt \\ &+ \int_{\mathbb{T}^3}\left(\frac{1}{2}\frac{|m|^2}{\rho}+h(\rho)\right)\Big|_{t=0}\theta(0)dx \ge 0\,,\end{aligned}$$

for any $\theta(t)$ nonnegative Lipschitz test function compactly supported in $[0,T)$. The family (ρ, m) is assumed to satisfy the (uniform in ε) bounds

$$\sup_{t\in(0,T)} \int_{\mathbb{T}^3} \rho dx \le K_1 < \infty\,,$$
$$\sup_{t\in(0,T)} \int_{\mathbb{T}^3} \left[\frac{1}{2}\frac{|m|^2}{\rho} + h(\rho)\right] dx \le K_2 < \infty\,,$$

which are natural within the given framework, and follow from corresponding uniform bounds on the initial data.

(ii) $\bar\rho$ is a smooth (C^3) periodic solution of the multidimensional porous media equation (2) that avoids vacuum, $\bar\rho \ge \bar\rho^* > 0$; $\bar m$ is defined via $\bar m = -\varepsilon\nabla p(\bar\rho)$.

Using the stability property in Proposition 2.1, one controls the distance between the relaxing sequence and the limiting solution by means of the distance function:

$$\varphi(t) = \int_{\mathbb{T}^3} \eta(\rho, m\,|\bar\rho, \bar m)dx\,.$$

The results are valid for pressure laws satisfying quite general conditions (see theorem below), and apply to γ–law pressures $p(\rho) = k\rho^\gamma$, $\gamma \ge 1$. For the proof we refer once again to [7].

Theorem 2.2. *Let $T > 0$ be fixed and assume $p(\rho)$ satisfies*

$$p''(\rho) \le A\frac{p'(\rho)}{\rho} \qquad \forall\ \rho > 0$$

and

$$p'(\rho) = k\gamma\rho^{\gamma-1} + o(\rho^{\gamma-1})\,, \qquad \textit{as } \rho \to +\infty\,.$$

Under hypothesis $(\mathbf{H_1})$*, the stability estimate*

$$\varphi(t) \le C\big(\varphi(0) + \varepsilon^4\big), \qquad t \in [0,T]\,,$$

holds, where C is a positive constant depending only on T, K_1, $\bar\rho$ and its derivatives. Moreover, if $\varphi(0) \to 0$ as $\varepsilon \to 0$, then

$$\sup_{t\in[0,T]} \varphi(t) \to 0, \ \ \textit{as } \varepsilon \to 0\,.$$

Remark 2.3. The relative entropy method can also be applied to other frameworks, such as 1–d dissipative weak solutions with different end states at ±∞ [7, Sec 2.3.2], as well as for comparing entropic measure-valued solutions of the Euler equations with friction to smooth solutions of the porous media equation [7, Sec 2.4].

3. **The diffusive limit from the Euler-Poisson system with friction to the Keller-Segel system.**

A variant of the above calculation may be used to establish convergence from the Euler–Poisson system with attractive potentials and friction to the Keller-Segel model. The Euler-Poisson system with friction is

$$\begin{cases} \rho_t + \dfrac{1}{\varepsilon}\operatorname{div}_x m = 0 \\ m_t + \dfrac{1}{\varepsilon}\operatorname{div}_x \dfrac{m\otimes m}{\rho} + \dfrac{1}{\varepsilon}\nabla_x p(\rho) = -\dfrac{1}{\varepsilon^2}m + \dfrac{1}{\varepsilon}\rho\nabla_x c \\ -\triangle_x c + \beta c = \rho, \end{cases} \tag{10}$$

where, as usual, $t\in\mathbb{R}$, $x\in\mathbb{R}^3$, $\rho \ge 0$, $c\in\mathbb{R}$, $m\in\mathbb{R}^3$, the pressure $p(\rho)$ satisfies $p'(\rho) > 0$ and β is a positive, sufficiently large constant, as we shall see in the sequel, which captures the effects of screening. In the limit $\varepsilon \to 0$, we obtain

$m = \rho\nabla_x c - \nabla_x p(\rho)$, and therefore the formal limit of (10) is given by the Keller-Segel type model:

$$\begin{cases} \rho_t + \operatorname{div}_x \big(\rho\nabla_x c - \nabla_x p(\rho)\big) = 0 \\ -\triangle_x c + \beta c = \rho. \end{cases} \tag{11}$$

We refer to [4] (and references therein) for convergence results using the compensated compactness method, and discussions of alternative scalings. Here, we focus to the convergence from (10) to (11) as a case study of the relative entropy method.

We again consider the entropy–entropy flux pair

$$\eta(\rho, m) = \frac{1}{2}\frac{|m|^2}{\rho} + h(\rho), \quad q(\rho, m) = \frac{1}{2} m \frac{|m|^2}{\rho^2} + mh'(\rho), \quad h''(\rho) = \frac{p'(\rho)}{\rho},$$

and note that an entropy weak solution of (10) satisfies the entropy inequality

$$\eta(\rho, m)_t + \frac{1}{\varepsilon}\operatorname{div}_x q(\rho, m) \le -\frac{1}{\varepsilon^2}\frac{|m|^2}{\rho} + \frac{1}{\varepsilon} m \cdot \nabla_x c\,. \tag{12}$$

On the other hand, smooth solutions of (11) satisfy the entropy identity

$$h(\rho)_t + \operatorname{div}_x \big(h'(\rho)(\rho\nabla_x c - \nabla_x p(\rho))\big) = -\frac{|\nabla_x p(\rho)|^2}{\rho} + \nabla_x p(\rho) \cdot \nabla_x c\,. \tag{13}$$

Note that (13) is indeed the equilibrium version ($\varepsilon = 0$) of the energy dissipation (12), as can be easily shown via the standard Hilbert expansion analysis.

As it is manifest, neither (12) nor (13) are indeed *dissipative*, due to the extra terms coming from the coupling with the equation for the concentration c. To take into account these extra terms, we consider the following modified entropy–entropy flux pair, again based on the mechanical energy of the system under consideration:

$$\mathcal{H}(\rho, m, c) = \eta(\rho, m) - \rho c\,,$$
$$\mathcal{Q}(\rho, m, c) = q(\rho, m) - mc\,.$$

Then the entropy inequality becomes

$$\mathcal{H}(\rho, m, c)_t + \frac{1}{\varepsilon}\operatorname{div}_x \mathcal{Q}(\rho, m, c) \le -\frac{1}{\varepsilon^2}\frac{|m|^2}{\rho} - \rho c_t\,. \tag{14}$$

Moreover, multiplying $(10)_3$ by c_t we get

$$\rho c_t = \frac{1}{2}\big(\beta c^2 + |\nabla_x c|^2\big)_t - \operatorname{div}_x \big(c_t \nabla_x c\big)\,,$$

which once added to (14) gives

$$\left(\mathcal{H}(\rho, m, c) + \frac{1}{2}\big(\beta c^2 + |\nabla_x c|^2\big)\right)_t + \frac{1}{\varepsilon}\operatorname{div}_x \big(\mathcal{Q}(\rho, m, c) - \varepsilon c_t \nabla_x c\big) \le -\frac{1}{\varepsilon^2}\frac{|m|^2}{\rho}\,. \tag{15}$$

The estimate (15) is the starting point to obtain the stability estimate in terms of the relative entropy and the corresponding analysis of the relaxation limit.

We rewrite the equilibrium system (11) in the variables $\bar\rho$, $\bar c$ and

$$\bar m = -\varepsilon\big(\nabla_x p(\bar\rho) - \bar\rho\nabla_x \bar c\big) = -\varepsilon\bar\rho\nabla_x\big(h'(\bar\rho) - \bar c\big)$$

in the form:

$$\begin{cases} \bar\rho_t + \frac{1}{\varepsilon}\operatorname{div}_x \bar m = 0 \\ \bar m_t + \frac{1}{\varepsilon}\operatorname{div}_x \frac{\bar m \otimes \bar m}{\bar\rho} + \frac{1}{\varepsilon}\nabla_x p(\bar\rho) = -\frac{1}{\varepsilon^2}\bar m + \frac{1}{\varepsilon}\bar\rho\nabla_x \bar c + e(\bar\rho, \bar m) \\ -\triangle_x \bar c + \beta \bar c = \bar\rho, \end{cases} \tag{16}$$

where the error term $e(\bar\rho,\bar m)$ is

$$\begin{aligned}\bar e := e(\bar\rho,\bar m) &= \frac{1}{\varepsilon}\operatorname{div}_x\left(\frac{\bar m\otimes\bar m}{\bar\rho}\right) - \varepsilon\partial_t\big(\nabla_x p(\bar\rho) - \bar\rho\nabla_x\bar c\big) \qquad (17)\\ &= \varepsilon\operatorname{div}_x\Big(\bar\rho\nabla_x\big(h'(\bar\rho)-\bar c\big)\otimes\nabla_x\big(h'(\bar\rho)-\bar c\big)\Big) - \varepsilon\partial_t\big(\bar\rho\nabla_x\big(h'(\bar\rho)-\bar c\big)\big) = O(\varepsilon)\,.\end{aligned}$$

In turn, (13) is rewritten as

$$\eta(\bar\rho,\bar m)_t + \frac{1}{\varepsilon}\operatorname{div}_x q(\bar\rho,\bar m) = -\frac{1}{\varepsilon^2}\frac{|\bar m|^2}{\bar\rho} + \frac{1}{\varepsilon}\bar m\cdot\nabla_x\bar c + \nabla_m\eta(\bar\rho,\bar m)\cdot\bar e\,,$$

or, equivalently,

$$\begin{aligned}&\Big(\mathcal H(\bar\rho,\bar m,\bar c) + \frac12\big(\beta\bar c^2 + |\nabla_x\bar c|^2\big)\Big)_t + \frac{1}{\varepsilon}\operatorname{div}_x\big(\mathcal Q(\bar\rho,\bar m,\bar c) - \varepsilon\bar c_t\nabla_x\bar c\big)\\ &= -\frac{1}{\varepsilon^2}\frac{|\bar m|^2}{\bar\rho} + \nabla_m\eta(\bar\rho,\bar m)\cdot\bar e\,.\end{aligned}$$

We now define the relative entropy

$$\mathcal H(\rho,m,c\,|\,\bar\rho,\bar m,\bar c\,) = \eta(\rho,m\,|\,\bar\rho,\bar m\,) - (\rho-\bar\rho)(c-\bar c)\,,$$

with corresponding relative entropy flux

$$\mathcal Q(\rho,m,c\,|\,\bar\rho,\bar m,\bar c\,) = q(\rho,m\,|\,\bar\rho,\bar m\,) - (m-\bar m)(c-\bar c)\,,$$

and show that:

Proposition 3.1. *For any weak, entropy solution (ρ,m,c) of (10) and any smooth solution $(\bar\rho,\bar m,\bar c)$ of (16) it holds*

$$\begin{aligned}&\partial_t\Big(\mathcal H(\rho,m,c\,|\,\bar\rho,\bar m,\bar c\,) + \frac12\big(\beta(c-\bar c)^2 + |\nabla_x(c-\bar c)|^2\big)\Big)\\ &\quad + \frac{1}{\varepsilon}\operatorname{div}_x\Big(\mathcal Q(\rho,m,c\,|\,\bar\rho,\bar m,\bar c\,) - \varepsilon(c-\bar c)_t\nabla_x(c-\bar c)\Big)\\ &\le -\frac{1}{\varepsilon^2}R(\rho,m\,|\,\bar\rho,\bar m\,) - Q - P - E\,, \qquad (18)\end{aligned}$$

where R, Q and E are defined in Proposition 2.1, but with $\bar e$ defined in (17), and

$$P = \frac{1}{\varepsilon}\frac{\bar m}{\bar\rho}(\rho-\bar\rho)\cdot\nabla_x(c-\bar c)\,.$$

Proof. We sketch the proof of (18) starting from Proposition 2.1 and analyzing the extra terms coming from the coupling with the elliptic equation involving the variable c. The estimate for $\eta(\rho,m\,|\,\bar\rho,\bar m\,)$ becomes in this case

$$\begin{aligned}&\eta(\rho,m\,|\,\bar\rho,\bar m\,)_t + \frac{1}{\varepsilon}\operatorname{div}_x q(\rho,m\,|\,\bar\rho,\bar m\,)\\ &\le -\frac{1}{\varepsilon^2}R(\rho,m\,|\,\bar\rho,\bar m\,) - Q - E + \frac{1}{\varepsilon}\rho\left(\frac{m}{\rho} - \frac{\bar m}{\bar\rho}\right)\cdot\nabla_x(c-\bar c)\,.\end{aligned}$$

Then, we multiply

$$-\triangle_x(c-\bar c) + \beta(c-\bar c) = \rho-\bar\rho$$

by $(c-\bar c)_t$ to conclude

$$(\rho-\bar\rho)(c-\bar c)_t = \frac12\big(\beta(c-\bar c)^2 + |\nabla_x(c-\bar c)|^2\big)_t - \operatorname{div}_x\big((c-\bar c)_t\nabla_x(c-\bar c)\big)\,.$$

Putting all relations together, we end up with

$$\begin{aligned}
&\partial_t \left(\mathcal{H}(\rho, m, c \,|\, \bar\rho, \bar m, \bar c\,) + \frac{1}{2}\big(\beta(c-\bar c)^2 + |\nabla_x (c - \bar c)|^2\big) \right) \\
&\qquad + \frac{1}{\varepsilon} \operatorname{div}_x \big(\mathcal{Q}(\rho, m, c \,|\, \bar\rho, \bar m, \bar c\,) - \varepsilon (c - \bar c)_t \nabla_x (c - \bar c) \big) \\
&\quad \leq -\frac{1}{\varepsilon^2} R(\rho, m \,|\, \bar\rho, \bar m\,) - Q - E \\
&\qquad + \frac{1}{\varepsilon} \rho \left(\frac{m}{\rho} - \frac{\bar m}{\bar\rho} \right) \cdot \nabla_x (c - \bar c) - \frac{1}{\varepsilon}(m - \bar m) \cdot \nabla_x (c - \bar c) \\
&\quad = -\frac{1}{\varepsilon^2} R(\rho, m \,|\, \bar\rho, \bar m\,) - Q - E - \frac{1}{\varepsilon} \frac{\bar m}{\bar\rho} (\rho - \bar\rho) \cdot \nabla_x (c - \bar c)\,,
\end{aligned}$$

which is exactly (18). □

We conclude our analysis by using the inequality (18) in the particular case $p(\rho) = h(\rho) = k\rho^2$ for which the limit system becomes

$$\begin{cases} \rho_t + \operatorname{div}_x \big(\rho \nabla_x (c - 2\rho)\big) = 0 \\ -\triangle_x c + \beta c = \rho. \end{cases} \tag{19}$$

Indeed, in that case, if $\beta > \frac{1}{2k}$, the relative entropy gives directly the L^2 control of the relaxation process, and in particular of the difference $\rho - \bar\rho$, and this is exactly what is needed in the estimate of the extra term P obtained above. Clearly, other frameworks of applications can be considered, for instance γ–laws for $\gamma \geq 2$ and ρ, $\bar\rho \geq \rho_* > 0$, for which we have in particular $h(\rho \,|\, \bar\rho) \geq C(\rho - \bar\rho)^2$.

Therefore, we denote

$$\begin{aligned}
&\psi(t) \\
&= \int_{\mathbb{T}^3} \left(\frac{1}{2}\rho \left| \frac{m}{\rho} - \frac{\bar m}{\bar\rho} \right|^2 + \frac{1}{2}\big(\beta(c - \bar c)^2 + |\nabla_x (c - \bar c)|^2\big) + k(\rho - \bar\rho)^2 - (\rho - \bar\rho)(c - \bar c) \right) dx
\end{aligned}$$

and we observe

$$\psi(t) \geq C \left(\int_{\mathbb{T}^3} \frac{1}{2}\rho \left| \frac{m}{\rho} - \frac{\bar m}{\bar\rho} \right|^2 dx + \|c - \bar c\|^2_{L^2} + \|\nabla_x (c - \bar c)\|^2_{L^2} + \|\rho - \bar\rho\|^2_{L^2} \right)$$

for β as above.

We again consider dissipative weak solutions of (10), for which in particular an integrated version of the relative entropy estimate (18) can be rigorously derived; more specific results in this direction are under investigation in [8]. We place the following hypotheses, referred to as $(\mathbf{H_2})$:

(i) $(\rho\,,\, m\,, c) : (0, T) \times \mathbb{T}^3 \to \mathbb{R}^5$ is a (periodic) *dissipative weak solution* of (10) with $p(\rho) = k\rho^2$, $\beta > \frac{1}{2k}$, with $\rho \geq 0$, satisfying the weak form of (10) and the integrated form of the relative entropy inequality (18):

$$\begin{aligned}
&\iint_{[0,+\infty)\times\mathbb{T}^3} \Bigg[\left(\mathcal{H}(\rho, m, c \,|\, \bar\rho, \bar m, \bar c\,) + \frac{1}{2}\big(\beta(c - \bar c)^2 + |\nabla_x (c - \bar c)|^2\big) \right) \dot\theta(t) \\
&\qquad - \left(\frac{1}{\varepsilon^2} R(\rho, m \,|\, \bar\rho, \bar m\,) - Q - P - E \right) \theta(t) \Bigg]\, dx dt \\
&\qquad + \int_{\mathbb{T}^3} \left(\mathcal{H}(\rho, m, c \,|\, \bar\rho, \bar m, \bar c\,) + \frac{1}{2}\big(\beta(c - \bar c)^2 + |\nabla_x (c - \bar c)|^2\big) \right)\Bigg|_{t=0} \theta(0) dx \geq 0\,,
\end{aligned}$$

for any $\theta(t)$ nonnegative Lipschitz test function compactly supported in $[0,T)$. The family (ρ, m, c) is assumed to satisfy the (uniform in ε) bounds

$$\sup_{t\in(0,T)} \int_{\mathbb{T}^3} \rho dx \le K_1 < \infty\,,$$

$$\int_{\mathbb{T}^3} \Big(\mathcal{H}(\rho, m, c\,|\bar\rho, \bar m, \bar c) + \frac{1}{2}\big(\beta(c-\bar c)^2 + |\nabla_x(c-\bar c)|^2\big)\Big)\Big|_{t=0} dx \le K_2 < \infty\,,$$

which are natural within the given framework.

(ii) $(\bar\rho, \bar c)$ is a smooth (C^3) periodic solution of (19) such that $\bar\rho \ge \bar\rho^* > 0$; $\bar m$ is given by $\bar m = -\varepsilon\bar\rho\nabla_x(2\rho - c)$.

Then the following theorem holds.

Theorem 3.2. *Let $T > 0$ be fixed and assume that hypothesis* ($\mathbf{H_2}$) *holds. Then, the following stability estimate holds:*

$$\psi(t) \le C(\psi(0) + \varepsilon^4)\,,$$

for any $t \in [0,T]$, with C a positive constant depending only on T, K_1, $\bar\rho$, $\bar c$ and their derivatives. Moreover, if $\psi(0) \to 0$ as $\varepsilon \to 0$, then

$$\sup_{t\in[0,T]} \psi(t) \to 0, \;\; \textit{as } \varepsilon \to 0\,.$$

Proof. As usual in this context, in ($\mathbf{H_2}$) we choose the test function

$$\theta(\tau) := \begin{cases} 1, & \text{for } 0 \le \tau < t, \\ \frac{t-\tau}{\kappa} + 1, & \text{for } t \le \tau < t+\kappa, \\ 0, & \text{for } \tau \ge t+\kappa, \end{cases}$$

to get, as $\kappa \to 0$,

$$\psi(t) + \frac{1}{\varepsilon^2}\int_0^t \int_{\mathbb{T}^3} R(\rho, m\,|\bar\rho, \bar m) dx d\tau \le \psi(0) + \int_0^t \int_{\mathbb{T}^3} (|Q| + |E| + |P|) dx d\tau\,.$$

Now, the hypotheses of Theorem 2.2 are satisfied for the particular case $p(\rho) = k\rho^2$. Therefore, we can carry out here the same estimates for the terms $|Q|$ and $|E|$ as follows:

$$\int_0^t \int_{\mathbb{T}^3} |Q| dx d\tau \le C_1 \int_0^t \psi(\tau) d\tau\,,$$

where C_1 depends on $\|\partial_{x_i x_j}\bar\rho\|_{L^\infty}$. The error term E in (9) is estimated by

$$\begin{aligned}\int_0^t \int_{\mathbb{T}^3} |E| dx d\tau &\le \frac{\varepsilon^2}{2}\int_0^t \int_{\mathbb{T}^3} \left|\frac{\bar e}{\bar\rho}\right|^2 \rho dx d\tau + \frac{1}{2\varepsilon^2}\int_0^t \int_{\mathbb{T}^3} \rho \left|\frac{m}{\rho} - \frac{\bar m}{\bar\rho}\right|^2 dx d\tau \\ &\le C_2\varepsilon^4\, t + \frac{1}{2\varepsilon^2}\int_0^t \int_{\mathbb{T}^3} R(\rho, m\,|\bar\rho, \bar m) dx d\tau\,,\end{aligned}$$

where C_2 depends on K_1, T and $\bar\rho$ through the following norms of derivatives up to third order:

$$\left\|\frac{1}{\bar\rho}\,\mathrm{div}_x\Big(\bar\rho\nabla_x\big(\bar\rho - \bar c\big) \otimes \nabla_x\big(\bar\rho - \bar c\big)\Big)\right\|_{L^\infty} + \left\|\frac{1}{\bar\rho}\partial_t\big(\bar\rho\nabla_x\big(\bar\rho - \bar c\big)\big)\right\|_{L^\infty}.$$

Finally, Young's inequality implies

$$\int_0^t \int_{\mathbb{T}^3} |P| dx d\tau \le C\|\nabla_x(2\bar\rho - \bar c)\|_\infty \int_0^t \psi(\tau) d\tau$$

and the proof follows from the Gronwall Lemma. □

4. **The p–system with friction.** Another, actually easier case in which one can apply the above technique is given by the p–system with friction in one space dimension:

$$\begin{aligned} u_t - \frac{1}{\varepsilon} v_x &= 0 \\ v_t - \frac{1}{\varepsilon}\tau(u)_x &= -\frac{1}{\varepsilon^2} v\,, \end{aligned} \tag{20}$$

where τ satisfies $\tau'(u) > 0$ to guarantee strict hyperbolicity. The system (20) is a model either for elasticity with friction or for isentropic gas dynamics in Lagrangian coordinates. Then u stands for the strain (or the specific volume for gases), v for the velocity and τ for the stress.

In the limit $\varepsilon \to 0$, solutions of (20) converge towards solutions of the parabolic equation

$$u_t - \tau(u)_{xx} = 0\,. \tag{21}$$

This limit may be obtained via the relative entropy estimate as we describe below; we refer to [7] for the details.

To this aim, let us consider the mechanical energy

$$\mathcal{E}(u,v) = \frac{1}{2}v^2 + W(u)\,,$$

where

$$W(u) = \int_0^u \tau(s)ds$$

stands for the stored energy, and its associated flux

$$\mathcal{F}(u,v) = -v\tau(u)\,.$$

The corresponding entropy inequality is

$$\mathcal{E}(u,v)_t + \frac{1}{\varepsilon}\mathcal{F}(u,v)_x \leq -\frac{1}{\varepsilon^2}v^2 \leq 0\,, \tag{22}$$

and captures the dissipation of the mechanical energy for weak solutions of (20). Smooth solutions of (21) satisfy the energy dissipation identity

$$\mathcal{E}(u,0)_t + \mathcal{F}(u,\tau(u)_x)_x = -\big(\tau(u)_x\big)^2 \leq 0\,.$$

The latter is the equilibrium ($\varepsilon = 0$) limit of (22).

The relative entropy is again defined as the quadratic part of the Taylor expansion of $\mathcal{E}(u,v)$ relative to the "algebraic–differential equilibrium" $(\bar u, \bar v)$, where $\bar u$ is a smooth solution of (21) and $\bar v = \varepsilon\tau(\bar u)_x$. Namely,

$$\begin{aligned} \mathcal{E}(u,v\,|\bar u,\bar v\,) &= \mathcal{E}(u,v) - \mathcal{E}(\bar u,\bar v) - \mathcal{E}_u(\bar u,\bar v)(u-\bar u) - \mathcal{E}_v(\bar u,\bar v)(v-\bar v) \\ &= \frac{1}{2}(v-\bar v)^2 + W(u\,|\bar u\,)\,. \end{aligned}$$

As corresponding flux we shall consider

$$\begin{aligned} \mathcal{F}(u,v\,|\bar u,\bar v\,) &= \mathcal{F}(u,v) - \mathcal{F}(\bar u,\bar v) + \mathcal{E}_u(\bar u,\bar v)(v-\bar v) + \mathcal{E}_v(\bar u,\bar v)(\tau(u)-\tau(\bar u)) \\ &= -(v-\bar v)(\tau(u)-\tau(\bar u))\,. \end{aligned}$$

As in the previous sections, we rewrite the equilibrium equation (21) as a damped p–system

$$\begin{cases} \bar u_t - \dfrac{1}{\varepsilon}\bar v_x = 0 \\ \bar v_t - \dfrac{1}{\varepsilon}\tau(\bar u)_x = -\dfrac{1}{\varepsilon^2}\bar v + \bar v_t\,, \end{cases} \tag{23}$$

where the term $\bar{v}_t = \varepsilon\tau(\bar{u})_{xt}$ is an error of order ε. Then a direct computation gives the following proposition.

Proposition 4.1. *For any weak, entropy solution (u, v) of (20) and any smooth solution $(\bar{u}, \bar{v})$ of (23) it holds:*

$$\mathcal{E}(u, v \,|\bar{u}, \bar{v}\,)_t + \frac{1}{\varepsilon}\mathcal{F}(u, v \,|\bar{u}, \bar{v}\,)_x \leq -\frac{1}{\varepsilon^2}(v-\bar{v})^2 + \tau(\bar{u})_{xx}\tau(u\,|\bar{u}\,) - \varepsilon\tau(\bar{u})_{xt}(v-\bar{v})\,. \tag{24}$$

The terms in the right hand side of (24) are analogous to the terms in (8) of Proposition 2.1 for the Eulerian case: the first term is dissipative and is due to the friction in the relaxation system, the second is quadratic in the flux, while the last term is a linear error term.

Finally, using this result, one can obtain stability and convergence of the relaxation limit in terms of the quantity

$$\int_{\mathbb{R}} \mathcal{E}(u, v \,|\bar{u}, \bar{v}\,) dx\,,$$

provided $\tau(u)$ satisfies appropriate growth conditions at infinity; see [7] for details.

5. Viscoelasticity with memory.

It is well known that the system of viscoelasticity of memory type can yield in different scaling limits both the equations of viscoelasticity of the rate type, as well as the equations of elasticity. We consider such scaling limits from the perspective of the relative entropy method, hoping to indicate the remarkably wide applicability of the methodology. We start by considering a quasilinear model (1–d for simplicity) with a diffusive scaling, thus entering in the framework of diffusive relaxations [7], and then we shall review the multidimensional model of stress relaxation approximating the equations of elastodynamics considered in [6, Sec 3.2].

5.1. From viscoelasticity of the memory type to viscoelasticity of the rate type.

First, we consider a diffusive scaling limit leading to a hyperbolic – parabolic system describing the dynamics of a 1–d viscoelastic material of the rate type. To this end, consider the following 3×3, one dimensional, quasilinear system of viscoelasticity with memory effects:

$$\begin{aligned} u_t - v_x &= 0 \\ v_t - \sigma(u)_x - \frac{1}{\varepsilon} z_x &= 0 \\ z_t - \frac{\mu}{\varepsilon} v_x &= -\frac{1}{\varepsilon^2} z\,, \end{aligned} \tag{25}$$

where $\mu > 0$ and the elastic stress function σ satisfies the usual condition $\sigma'(u) > 0$ for hyperbolicity. In (25), the stress $S = \sigma(u) + \frac{1}{\varepsilon} z$ is decomposed in a purely elastic part and a viscoelastic part of the memory type (see $(25)_3$ for z), scaled so that it relaxes as $\varepsilon \to 0$ to the equations of viscoelasticity of the rate type:

$$\begin{aligned} u_t - v_x &= 0 \\ v_t - \sigma(u)_x &= \mu v_{xx}\,. \end{aligned} \tag{26}$$

Indeed, in (26) the stress is given by $\sigma(u) + \mu v_x$, that is again the same elastic part plus a Newtonian viscous stress.

The mechanical energy – energy flux couple for (25) is

$$\begin{aligned}
&\mathbb{E}(u,v,z) = \int_0^u \sigma(s)ds + \frac{1}{2}v^2 + \frac{1}{2\mu}z^2 = \Sigma(u) + \frac{1}{2}v^2 + \frac{1}{2\mu}z^2,\\
&\mathbb{F}_\varepsilon(u,v,z) = -(\varepsilon\sigma(u)v + vz)\,.
\end{aligned}$$

Hence, the dissipation of mechanical energy for weak solutions of (25) reads

$$\mathbb{E}(u,v,z)_t + \frac{1}{\varepsilon}\mathbb{F}_\varepsilon(u,v,z)_x \le -\frac{1}{\mu\varepsilon^2}z^2$$

and the corresponding relation for smooth solutions of (26) is given by

$$\mathbb{E}(u,v,0)_t + \mathbb{F}_1(u,v,\sigma(u)_x)_x = -\mu(v_x)^2 \le 0\,,$$

for

$$\mathbb{E}(u,v,0) = \Sigma(u) + \frac{1}{2}v^2, \quad \mathbb{F}_1(u,v,\sigma(u)_x) = -(\sigma(u)v + \mu v v_x)\,.$$

We rewrite the equilibrium system (26) and the corresponding stress–strain response for the variables $(\bar u, \bar v, \bar z)$ with $\bar z = \varepsilon\mu\bar v_x$ as follows:

$$\begin{cases}
\bar u_t - \bar v_x = 0\\
\bar v_t - \sigma(\bar u)_x - \frac{1}{\varepsilon}\bar z_x = 0\\
\bar z_t - \frac{\mu}{\varepsilon}\bar v_x = -\frac{1}{\varepsilon^2}\bar z + \bar z_t\,,
\end{cases} \tag{27}$$

where we shall treat the term $\bar z_t$ as an $O(\varepsilon)$ error:

$$\bar z_t = \varepsilon\mu\bar v_{xt} = \varepsilon\mu\big(\sigma(\bar u)_x + \mu\bar v_{xx}\big)_x\,.$$

Finally, the relative entropy and relative entropy flux, respectively,

$$\begin{aligned}
\mathbb{E}(u,v,z\,|\bar u,\bar v,\bar z\,) &= \mathbb{E}(u,v,z) - \mathbb{E}(\bar u,\bar v,\bar z)\\
&\quad - \mathbb{E}_u(\bar u,\bar v,\bar z)(u-\bar u) - \mathbb{E}_v(\bar u,\bar v,\bar z)(v-\bar v) - \mathbb{E}_z(\bar u,\bar v,\bar z)(z-\bar z)\,,\\
\mathbb{F}_\varepsilon(u,v,z\,|\bar u,\bar v,\bar z\,) &= \mathbb{F}_\varepsilon(u,v,z) - \mathbb{F}_\varepsilon(\bar u,\bar v,\bar z) - \mathbb{E}_u(\bar u,\bar v,\bar z)\big(-\varepsilon(v-\bar v)\big)\\
&\quad - \mathbb{E}_v(\bar u,\bar v,\bar z)\big(-\varepsilon(\sigma(u)-\sigma(\bar u)) - (z-\bar z)\big) - \mathbb{E}_z(\bar u,\bar v,\bar z)\big(v-\bar v\big)\,,
\end{aligned}$$

verify the following identity:

Proposition 5.1. *Let (u,v,z) be a weak entropy solution of (25) and let $(\bar u,\bar v,\bar z)$ be a smooth solution of (27). Then*

$$\begin{aligned}
\partial_t\mathbb{E}(u,v,z\,|\bar u,\bar v,\bar z\,) &+ \frac{1}{\varepsilon}\partial_x\mathbb{F}_\varepsilon(u,v,z\,|\bar u,\bar v,\bar z\,)\\
&\le -\frac{1}{\mu\varepsilon^2}(z-\bar z)^2 + \bar v_x\sigma(u\,|\bar u\,) - \varepsilon\bar v_{xt}(z-\bar z)\,.
\end{aligned}$$

As in the previous cases, Proposition 5.1 suggests to measure the distance between systems (25) and (26) by means of

$$\int_{\mathbb{R}} \mathbb{E}(u,v,z\,|\bar u,\bar v,\bar z\,)dx\,,$$

and this can be done under appropriate structural condition on the stress function $\sigma(u)$ [7].

5.2. **A model of stress relaxation approximating the equations of elastodynamics.** As a final application of the relative entropy method, we shall review the case of the hyperbolic–hyperbolic relaxation limit $\varepsilon \to 0$ for the model of stress relaxation

$$\begin{aligned} \partial_t F_{i\alpha} &= \partial_\alpha v_i \\ \partial_t v_i &= \partial_\alpha S_{i\alpha} \\ \partial_t (S_{i\alpha} - f_{i\alpha}(F)) &= -\frac{1}{\varepsilon}(S_{i\alpha} - T_{i\alpha}(F))\,. \end{aligned} \tag{28}$$

In (28), i, $\alpha = 1, 2, 3$, F stands for the deformation gradient, v for the velocity and the stress S is again decomposed in an elastic part and a viscoelastic part with memory effects:

$$S = f(F) + \int_{-\infty}^{t} \frac{1}{\varepsilon} e^{-\frac{1}{\varepsilon}(t-\tau)} h(F(\cdot,\tau))\, d\tau\,.$$

In turn, the equilibrium stress is accordingly decomposed as $T(F) = f(F) + h(F)$. Following [6], we shall derive a relative entropy relation for smooth solutions (v, F, S) of (28) and smooth solutions $(\bar{v}, \bar{F})$ of its limit, that is the elasticity system

$$\begin{aligned} \partial_t \bar{F}_{i\alpha} &= \partial_\alpha \bar{v}_i \\ \partial_t \bar{v}_i &= \partial_\alpha T_{i\alpha}(\bar{F})\,, \end{aligned} \tag{29}$$

even if with nowadays technologies the same relation can be rigorously justified for dissipative weak solutions of (28). To this end, let us consider the framework

$$\begin{aligned} T(F) = \nabla_F W(F) &= f(F) + h(F)\,, \\ f(F) = \nabla_F W_I(F), &\quad h(F) = -\nabla_F W_v(F) \end{aligned} \tag{a}$$

and $W_v = W_I - W$ is convex. Under these structural hypotheses, the dissipation of the mechanical energy reads:

$$\begin{aligned} \partial_t \left(\frac{1}{2}|v|^2 + \Psi(F, S - f(F)) \right) &- \partial_\alpha (v_i S_{i\alpha}) \\ &+ \frac{1}{\varepsilon}(F_{i\alpha} - h^{-1}_{i\alpha}(S - f(F)))(S_{i\alpha} - T_{i\alpha}(F)) = 0\,. \end{aligned} \tag{30}$$

In (30), the free energy function Ψ is of the form

$$\Psi(F, A) = W_I(F) + A \cdot F + G(A)\,,$$

where G is a convex function such that $\nabla_A G = -h^{-1}$. Indeed, the condition that the inverse of h is a gradient is equivalent to the existence of a free energy function for (28) compatible with the Clausius-Duhem inequality. In (30) this is expressed by the positivity of the last term, revealing the dissipation arising from of the viscoelastic stresses [6].

At this point, we define the relative energy $\mathcal{E}_r(v, F, S \,|\, \bar{v}, \bar{F}, h(\bar{F}))$ generated by the mechanical energy relative to an equilibrium as follows:

$$\begin{aligned} \mathcal{E}_r := \frac{1}{2}|v - \bar{v}|^2 &+ \Psi(F, S - f(F)) - \Psi(\bar{F}, h(\bar{F})) \\ &- \nabla_F \Psi(\bar{F}, h(\bar{F})) \cdot (F - \bar{F}) - \nabla_A \Psi(\bar{F}, h(\bar{F})) \cdot (S - f(F) - h(\bar{F})) \\ = \frac{1}{2}|v - \bar{v}|^2 &+ \Psi(F, S - f(F)) - W(\bar{F}) - \nabla_F W(\bar{F}) \cdot (F - \bar{F})\,, \end{aligned}$$

by selecting an appropriate normalization so that $\Psi(F, h(F)) = W(F)$. The associated relative fluxes are then given by

$$\mathcal{F}_r^\alpha = (v_i - \bar{v}_i)(S_{i\alpha} - T_{i\alpha}(\bar{F}))\,.$$

The relative entropy computation is performed as follows: observe that (v, F, S) satisfies (30) and that the smooth solution $(\bar{v}, \bar{F})$ satisfies the energy identity

$$\partial_t \frac{1}{2}\Big(|\bar{v}|^2 + W(\bar{F})\Big) - \partial_\alpha\Big(\bar{v}_i T_{i\alpha}(\bar{F})\Big) = 0\,. \tag{31}$$

From

$$\begin{aligned} \partial_t(F_{i\alpha} - \bar{F}_{i\alpha}) &= \partial_\alpha(v_i - \bar{v}_i) \\ \partial_t(v_i - \bar{v}_i) &= \partial_\alpha(S_{i\alpha} - T_{i\alpha}(\bar{F})) \end{aligned}$$

and (29) we conclude

$$\begin{aligned} &\partial_t\Big(\frac{\partial W}{\partial F_{i\alpha}}(\bar{F})(F_{i\alpha} - \bar{F}_{i\alpha}) + \bar{v}_i(v_i - \bar{v}_i)\Big) \\ &\qquad - \partial_\alpha\Big(T_{i\alpha}(\bar{F})(v_i - \bar{v}_i) + \bar{v}_i(S_{i\alpha} - T_{i\alpha}(\bar{F}))\Big) \\ &\quad = \partial_t\Big(\frac{\partial W}{\partial F_{i\alpha}}(\bar{F})\Big)(F_{i\alpha} - \bar{F}_{i\alpha}) + (\partial_t \bar{v}_i)(v_i - \bar{v}_i) \\ &\qquad - (\partial_\alpha T_{i\alpha}(\bar{F}))(v_i - \bar{v}_i) - (\partial_\alpha \bar{v}_i)(S_{i\alpha} - T_{i\alpha}(\bar{F})) \\ &\quad = -(\partial_\alpha \bar{v}_i)\Big(S_{i\alpha} - T_{i\alpha}(\bar{F}) - \frac{\partial^2 W}{\partial F_{i\alpha}\partial F_{j\beta}}(\bar{F})(F_{j\beta} - \bar{F}_{j\beta})\Big)\,. \end{aligned} \tag{32}$$

Then, (30), (31) and (32) imply

$$\begin{aligned} &\partial_t \mathcal{E}_r - \partial_\alpha\Big((v_i - \bar{v}_i)(S_{i\alpha} - T_{i\alpha}(\bar{F}))\Big) \\ &\qquad + \frac{1}{\varepsilon}(F_{i\alpha} - h^{-1}_{i\alpha}(S - f(F)))(S_{i\alpha} - T_{i\alpha}(F)) \\ &\quad = (\partial_\alpha \bar{v}_i)\Big(T_{i\alpha}(F) - T_{i\alpha}(\bar{F}) - \frac{\partial^2 W}{\partial F_{i\alpha}\partial F_{j\beta}}(\bar{F})(F_{j\beta} - \bar{F}_{j\beta})\Big) + (\partial_\alpha \bar{v}_i)(S_{i\alpha} - T_{i\alpha}(F))\,. \end{aligned} \tag{33}$$

This relative entropy identity can be used to obtain stability and convergence of the relaxation system (28) as long as the solution of (29) remains smooth. Indeed, under appropriate conditions for the potentials W and W_I, namely that there exist positive constants $\gamma_I > \gamma_v > 0$ and $M > 0$ such that

$$\nabla_F^2 W_I(F) \ge \gamma_I I > \gamma_v I \ge \nabla_F^2(W_I - W)(F) > 0\,, \tag{b}$$

$$|\nabla_F^2 W_I(F)| \le M\,. \quad |\nabla^3 W(F)| \le M\,, \quad \forall F\,, \tag{c}$$

we get that $\Psi(F, A)$ is uniformly convex and therefore

$$\mathcal{E}_r \ge c\big(|v - \bar{v}|^2 + |F - \bar{F}|^2 + |A - h(\bar{F})|^2\big)$$

for a positive $c > 0$. Condition (b) is roughly equivalent to what is called subcharacteristic condition in the theory of relaxation. In addition, uniform convexity of $G(A)$ leads to

$$\nabla_A^2 G(A) = \big(-\nabla_F h\big)^{-1} = \big(\nabla_F^2(W_I - W)\big)^{-1} \ge \frac{1}{\gamma_v} I$$

so that

$$\begin{aligned}(F - h^{-1}(S - f(F))) \cdot (S - T(F)) &= (\nabla_A G(A) - \nabla_A G(h(F))) \cdot (A - h(F)) \\ &\geq \frac{1}{\gamma_v}|A - h(F)|^2 = \frac{1}{\gamma_v}|S - T(F)|^2 \,,\end{aligned}$$

giving the dissipation property of the relaxation term. Moreover, the first term on the right hand side of (33) is quadratic in $F - \bar{F}$, while the last term can be controlled by the dissipative relaxation term plus an $O(\varepsilon)$ error term. Hence, the following result holds (we refer to [6] for the technical details and the proof).

Theorem 5.2. *Let $(v^\varepsilon, F^\varepsilon, S^\varepsilon)$ be smooth solutions of (28) and $(\bar{v}, \bar{F})$ be a smooth solution of* (29) *defined on $\mathbb{R}^3 \times [0, T]$ and emanating from smooth data $(v_0^\varepsilon, F_0^\varepsilon, S_0^\varepsilon)$ and $(\bar{v}_0, \bar{F}_0)$. Then, under hypotheses* (a), (b), (c)*, the relative energy $\mathcal{E}_r$ satisfies (33), and, for $R > 0$, there exist constants s and $C > 0$ independent of ε such that*

$$\int_{|x|<R} \mathcal{E}_r(x,t)dx \leq C \left(\int_{|x|<R+st} \mathcal{E}_r(x,0)dx + \varepsilon \right) .$$

In particular, if the data satisfy

$$\int_{|x|<R+sT} \mathcal{E}_r(x,0)dx \longrightarrow 0 \,, \quad \text{as } \varepsilon \to 0 \,,$$

then

$$\sup_{t\in[0,T]} \int_{|x|<R} \left(|v^\varepsilon - \widehat{v}|^2 + |F^\varepsilon - \widehat{F}|^2 + |A^\varepsilon - h(\widehat{F})|^2 \right) dx \longrightarrow 0 \,.$$

REFERENCES

[1] F. Berthelin and A. Vasseur, *From kinetic equations to multidimensional isentropic gas dynamics before shocks*, SIAM J. Math. Anal., **36** (2005), 1807–1835.

[2] C.M. Dafermos, *The second law of thermodynamics and stability*, Arch. Rational Mech. Anal. **70** (1979), 167–179.

[3] C.M. Dafermos, *Stability of motions of thermoelastic fluids*, J. Thermal Stresses **2** (1979), 127–134.

[4] M. Di Francesco and D. Donatelli, *Singular convergence of nonlinear hyperbolic chemotaxis systems to Keller-Segel type models*, Discrete Contin. Dynam. Systems **13** (2010), no. 1, 79–100.

[5] R.J. DiPerna, *Uniqueness of solutions to hyperbolic conservation laws*, Indiana Univ. Math. J. **28** (1979), 137–188.

[6] C. Lattanzio and A.E. Tzavaras, *Structural properties of stress relaxation and convergence from viscoelasticity to polyconvex elastodynamics*, Arch. Rational Mech. Anal., **180** (2006), 449–492.

[7] C. Lattanzio and A.E. Tzavaras, *Relative entropy in diffusive relaxation* SIAM J. Math. Anal., **45** (2013), no. 3, 1563–1584.

[8] C. Lattanzio and A.E. Tzavaras, in preparation.

[9] A.E. Tzavaras, *Relative entropy in hyperbolic relaxation*, Commun. Math. Sci. **3** (2005), no. 2, 119–132.

E-mail address: corrado@univaq.it
E-mail address: tzavaras@tem.uoc.gr

A NOTE ON PHASE TRANSITIONS FOR THE SMOLUCHOWSKI EQUATION WITH DIPOLAR POTENTIAL

Pierre Degond

Université de Toulouse; UPS, INSA, UT1, UTM
Institut de Mathématiques de Toulouse
F-31062 Toulouse, France
and
CNRS; Institut de Mathématiques de Toulouse UMR 5219
F-31062 Toulouse, France

Amic Frouvelle

CEREMADE — UMR CNRS 7534
Université de PARIS – DAUPHINE
75775 PARIS CEDEX 16, France

Jian-Guo Liu

Department of Physics and Department of Mathematics
Duke University
Durham, NC 27708, USA

Abstract. In this note, we study the phase transitions arising in a modified Smoluchowski equation on the sphere with dipolar potential. This equation models the competition between alignment and diffusion, and the modification consists in taking the strength of alignment and the intensity of the diffusion as functions of the order parameter.

We characterize the stable and unstable equilibrium states. For stable equilibria, we provide the exponential rate of convergence. We detail special cases, giving rise to second order and first order phase transitions, respectively. We study the hysteresis diagram, and provide numerical illustrations of this phenomena.

1. **Introduction.** In this short note, we study the following modified Smoluchowski equation (also called Fokker–Planck equation), for an orientation distribution $f(\omega,t)$ defined for a time $t \geqslant 0$ and a direction $\omega \in \mathbb{S}$ (the unit sphere $\mathbb{S}$ of $\mathbb{R}^n$) as follows:

$$\partial_t f = -\nu(|J_f|)\nabla_\omega \cdot (f\, \nabla_\omega(\omega \cdot \Omega_f)) + \tau(|J_f|)\Delta_\omega f =: Q(f), \tag{1}$$

$$\Omega_f = \frac{J_f}{|J_f|}, \quad J_f(t) = \int_{\upsilon \in \mathbb{S}} \upsilon\, f(\upsilon,t)\,\mathrm{d}\upsilon. \tag{2}$$

where Δ_ω, $\nabla_\omega\cdot$, and ∇_ω are the Laplace–Beltrami, divergence, and gradient operators on the sphere. The vector $J_f \in \mathbb{R}^n$ is the first moment associated to f (the

2000 *Mathematics Subject Classification.* Primary: 35Q84, 82C05, 82C22, 82C70; Secondary: 92D50.

Key words and phrases. Spontaneous symmetry breaking, von Mises–Fisher distribution, critical density, critical exponent, rate of convergence, self-propelled particles, alignment interaction, Vicsek model.

measure on the sphere is the uniform measure such that $\int_{\mathbb{S}} \mathrm{d}\upsilon = 1$), and $\Omega_f \in \mathbb{S}$ represents the mean direction of the distribution f.

The first term of the right-hand side of (1) is an alignment term towards Ω_f, and the function ν represents the strength of this alignment. The function τ is the intensity of the diffusion on the sphere.

When τ is a constant and $\nu(|J_f|) = |J_f|$, we can recover the standard Smoluchowski equation on the sphere, with dipolar potential [11]. Indeed, the dipolar potential is given by $K(\omega, \bar{\omega}) = -\omega \cdot \bar{\omega}$, and the equation can be recast as:

$$\partial_t f = \nabla_\omega \cdot \Big(f \, \nabla_\omega \big(\int_{\mathbb{S}} K(\omega \cdot \bar{\omega}) f(\bar{\omega}, t) \mathrm{d}\bar{\omega} \big) \Big) + \tau \Delta_\omega f. \tag{3}$$

Other classical kernels [8, 14, 15] for the study of semi-dilute and concentrated suspensions of polymers are the so-called Maier–Saupe potential $K(\omega, \bar{\omega}) = -(\omega \cdot \bar{\omega})^2$, or the original Onsager potential $K(\omega, \bar{\omega}) = |\omega \times \bar{\omega}|$. In these cases, there are a lot of studies regarding the phase transition phenomenon for equilibrium states [3, 4, 9, 10, 12, 13, 16–20], and in particular a hysteresis phenomenon occurs for the Maier–Saupe potential in dimension 3. The case of the dipolar potential has also been studied precisely [11], with an analysis of the rates of convergence to the equilibrium as time goes to infinity. In this case, there exists a so-called continuous phase transition for a critical threshold τ_c: when $\tau < \tau_c$ the solution converges exponentially fast to a non-isotropic equilibrium; when $\tau > \tau_c$ it converges exponentially fast to the uniform distribution. At the critical value $\tau = \tau_c$, the solution converges to the uniform distribution at a rate $t^{-1/2}$.

Here we study the modifications arising when ν and τ depend on $|J_f|$, which can be motivated by some biological modeling (see [7] and references therein). In that case, we cannot use τ as a bifurcation parameter anymore. Instead, we will use the initial mass ρ (a conserved quantity) as the key parameter to study the phase transition. We will assume that:

Hypothesis 1.1.

(i) *The functions ν and τ are C^1, with $\nu(0) = 0$ and $\tau > 0$.*

(ii) *The function $|J| \mapsto h(|J|) = \frac{\nu(|J|)}{\tau(|J|)}$ is an increasing function. We denote by σ its inverse, i.e.*

$$\kappa = h(|J|) \Leftrightarrow |J| = \sigma(\kappa).$$

The first part of this hypothesis implies that we do not have any singularity of Q in (1) as $|J_f| \to 0$: if $|J_f| = 0$, we simply have $Q(f) = \tau(0)\Delta_\omega f$. The second part is made for the sake of simplicity. It leaves enough flexibility to to reveal key behaviors in terms of phase transitions. It would be easy to remove it at the price of an increased technicality. Additionally, it means that when f is more concentrated in the direction of Ω_f, the relative strength of the alignment force compared to diffusion is increased as well. This can be biologically motivated by the existence of some social reinforcement mechanism.

We will see that we can observe a wealth of phenomena, including hysteresis. The purpose of this note is to summarize the analytical results, as well as some numerical simulations which illustrate this phenomena. All the proofs are detailed in [6], where (1) arises as the spatially homogeneous version of a space-dependent kinetic equation, obtained as the mean-field limit of a self-propelled particle system interacting through alignment. This spatially homogeneous study is crucial to determine the macroscopic behavior of this space-dependent kinetic equation.

2. **General study.**

2.1. **Existence and uniqueness.**

We first state results about existence, uniqueness, positivity and regularity of the solutions of (1). Under hypothesis 1.1, we have the following

Theorem 1. *Given an initial finite non-negative measure f_0 in $H^s(\mathbb{S})$, there exists a unique weak solution f of (1) such that $f(0) = f_0$. This solution is global in time.*

Moreover, $f \in C^\infty((0,+\infty)\times\mathbb{S})$, with $f(t,\omega) > 0$ for all positive t, and we have the following instantaneous regularity and uniform boundedness estimates (for $m \in \mathbb{N}$, the constant C being independent of f), for all $t > 0$:

$$\|f(t)\|^2_{H^{s+m}} \leqslant C\left(1+\frac{1}{t^m}\right)\|f_0\|^2_{H^s}.$$

For later usage, we define $\Phi(|J|)$ as an anti-derivative of h: $\frac{\mathrm{d}\Phi}{\mathrm{d}|J|} = h(|J|)$. In this case, the dynamics of (1) corresponds to the gradient flow of the following free energy functional:

$$\mathcal{F}(f) = \int_{\mathbb{S}} f \ln f \,\mathrm{d}\omega - \Phi(|J_f|).$$

Indeed, if we define the dissipation term $\mathcal{D}(f)$ by

$$\mathcal{D}(f) = \tau(|J_f|)\int_{\mathbb{S}} f\,|\nabla_\omega(\ln f - h(|J_f|)\,\omega\cdot\Omega_f)|^2\,\mathrm{d}\omega,$$

we get the following conservation relation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(f) = -\mathcal{D}(f) \leqslant 0. \tag{4}$$

2.2. **Equilibria.**

We now define the von Mises distribution which provides the general shape of the non-isotropic equilibria of Q.

Definition 2.1. The von Mises distribution of orientation $\Omega \in \mathbb{S}$ and concentration parameter $\kappa \geqslant 0$ is given by:

$$M_{\kappa\Omega}(\omega) = \frac{e^{\kappa\,\omega\cdot\Omega}}{\int_{\mathbb{S}} e^{\kappa\,\upsilon\cdot\Omega}\,\mathrm{d}\upsilon}. \tag{5}$$

The order parameter $c(\kappa)$ is defined by the relation

$$J_{M_{\kappa\Omega}} = c(\kappa)\Omega,$$

and has expression:

$$c(\kappa) = \frac{\int_0^\pi \cos\theta\, e^{\kappa\cos\theta}\sin^{n-2}\theta\,\mathrm{d}\theta}{\int_0^\pi e^{\kappa\cos\theta}\sin^{n-2}\theta\,\mathrm{d}\theta}. \tag{6}$$

The concentration parameter $c(\kappa)$ defines a one-to-one correspondence $\kappa \in [0,\infty) \mapsto c(\kappa) \in [0,1)$. The case $\kappa = c(\kappa) = 0$ corresponds to the uniform distribution, while when κ is large (or $c(\kappa)$ is close to 1), the von Mises distribution is closed to a Dirac delta mass at the point Ω.

Some comments are necessary about the interval of definition of σ. First note that, under hypothesis 1.1, h is defined from $[0,+\infty)$, with values in an interval $[0,\kappa_{max})$, where we may have $\kappa_{max} = +\infty$. So σ is an increasing function from $[0,\kappa_{max})$ onto $\mathbb{R}_+$. Moreover, for later usage, we can define

$$\tau_0 = \tau(0) > 0, \quad \text{and} \quad \rho_c = \lim_{|J|\to 0}\frac{n|J|}{h(|J|)} = \frac{n\tau_0}{\nu'(0)} \tag{7}$$

where $\rho_c > 0$ may be equal to $+\infty$, and where we recall that n denotes the dimension.

The equilibria are given by the following proposition:

Proposition 2.1. *The following statements are equivalent:*

(i) *$f \in C^2(\mathbb{S})$ and $Q(f) = 0$.*
(ii) *$f \in C^1(\mathbb{S})$ and $\mathcal{D}(f) = 0$.*
(iii) *There exists $\rho \geqslant 0$ and $\Omega \in \mathbb{S}$ such that $f = \rho M_{\kappa\Omega}$, where $\kappa \geqslant 0$ satisfies the compatibility equation:*

$$\sigma(\kappa) = \rho c(\kappa). \tag{8}$$

Let us first remark that the uniform distribution, corresponding to $\kappa = 0$ is always an equilibrium. Indeed, we have $c(0) = \sigma(0) = 0$ and (8) is satisfied. However, Proposition 2.1 does not provide any information about the number of the non-isotropic equilibria. Indeed, equation (8) can be recast into:

$$\frac{c(\kappa)}{\sigma(\kappa)} = \frac{1}{\rho}, \tag{9}$$

which is valid as long as $\sigma \neq 0$. We know that σ is an increasing unbounded function from its interval of definition $[0, \kappa_{max})$ onto $[0, +\infty)$, and thanks to hypothesis 1.1 and to (7), we know that $\sigma(\kappa) \sim \frac{\rho_c}{n}\kappa$ as $\kappa \to 0$ (if $\rho_c < +\infty$). So since $c(\kappa) \sim \frac{1}{n}\kappa$ as $\kappa \to 0$ (see for instance [11]), we have the two following results (also valid in the case $\rho_c = +\infty$):

$$\frac{c(\kappa)}{\sigma(\kappa)} \to \frac{1}{\rho_c} \text{ as } \kappa \to 0 \quad \text{ and } \quad \frac{c(\kappa)}{\sigma(\kappa)} \to 0 \text{ as } \kappa \to \kappa_{max}. \tag{10}$$

We deduce that this function reaches its maximum, and we define

$$\rho_* = \min_{\kappa \in \mathbb{R}_+} \frac{\sigma(\kappa)}{c(\kappa)}. \tag{11}$$

For $\rho < \rho_*$, the only solution to the compatibility condition is $\kappa = 0$, and the only equilibrium is the uniform distribution $f = \rho$. Except from these facts, we have no further direct information of this function $\kappa \mapsto c(\kappa)/\sigma(\kappa)$, since c and σ are both increasing. Figure 1 depicts some examples of the possible shapes of the function $\kappa \mapsto c(\kappa)/\sigma(\kappa)$.

We see that depending on the value of ρ, the number of families of non-isotropic equilibria, given by the number of positive solutions of the equation (9), can be zero, one, two or even more. We now turn to the study of the stability of these equilibria, through the study of the rates of convergence.

2.3. Rates of convergence to equilibrium.

The main tool to prove convergence of the solution to a steady state is LaSalle's principle, that we recall here (the proof follows exactly the lines of [11]). By the conservation relation (4), we know that the free energy $\mathcal{F}$ is decreasing in time (and bounded from below since $|J|$ is bounded). LaSalle's principle states that the limiting value of $\mathcal{F}$ corresponds to an ω-limit set of equilibria:

Proposition 2.2. *LaSalle's invariance principle.*

Let f_0 be a positive measure on the sphere $\mathbb{S}$. We denote by $\mathcal{F}_\infty$ the limit of $\mathcal{F}(f(t))$ as $t \to \infty$, where f is the solution to the modified Smoluchowski equation (1) with initial condition f_0.

Then the set $\mathcal{E}_\infty = \{f \in C^\infty(\mathbb{S})$ s.t. $\mathcal{D}(f) = 0$ and $\mathcal{F}(f) = \mathcal{F}_\infty\}$ is not empty.

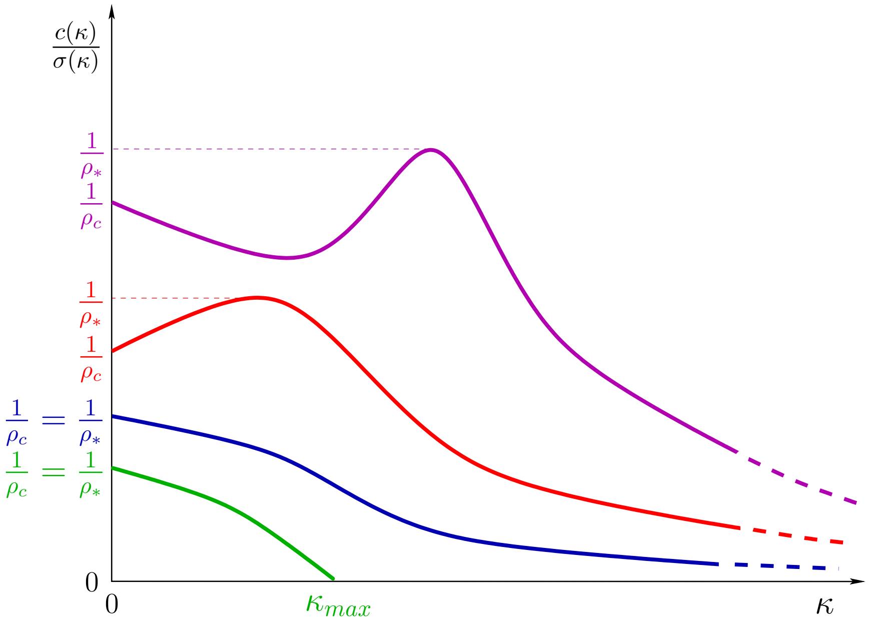


FIGURE 1. The green, blue, red and purple curves correspond to various possible profiles for the function $\kappa \mapsto \frac{c(\kappa)}{\sigma(\kappa)}$.

Furthermore $f(t)$ converges in any H^s norm to this set of equilibria (in the following sense):

$$\lim_{t\to\infty} \inf_{g\in\mathcal{E}_\infty} \|f(t) - g\|_{H^s} = 0.$$

Since we know the types of equilibria, we can refine this principle to adapt it to our problem:

Proposition 2.3. *Let f_0 be a positive measure on the sphere $\mathbb{S}$, with initial mass ρ.*

If no open interval is included in the set $\{\kappa, \rho c(\kappa) = \sigma(\kappa)\}$, then there exists a solution κ_∞ to the compatibility solution (8) such that we have:

$$\lim_{t\to\infty} |J_f(t)| = \rho c(\kappa_\infty)$$
$$\forall s \in \mathbb{R}, \lim_{t\to\infty} \|f(t) - \rho M_{\kappa_\infty \Omega_{f(t)}}\|_{H^s} = 0.$$

This last proposition helps us to characterize the ω-limit set by studying the single compatibility equation (8).

When $\kappa = 0$ is the unique solution, then this gives us that f converges to the uniform distribution. Otherwise, two cases are possible, either $\kappa_\infty = 0$, and f converges to the uniform distribution, or $\kappa_\infty \neq 0$, and the only unknown behavior is the one of $\Omega_{f(t)}$. If we are able to prove that it converges to $\Omega_\infty \in \mathbb{S}$, then f converges to a fixed non-isotropic steady-state $\rho M_{\kappa_\infty \Omega_\infty}$.

However, Proposition 2.3 does not give information about quantitative rates of convergence of $|J_f|$ to $\rho c(\kappa_\infty)$, and of $\|f(t) - \rho M_{\kappa_\infty \Omega_{f(t)}}\|_{H^s}$ to 0, as $t \to \infty$. So we now turn to the study of the behavior of the difference between the solution f and a target equilibrium $\rho M_{\kappa_\infty \Omega_{f(t)}}$.

This study consists in two types of expansion. If we expand the solution around the uniform equilibrium, some simple energy estimates give us exponential convergence when $\rho < \rho_c$. But when we expand the solution around a non-isotropic equilibrium $\rho M_{\kappa_\infty \Omega_f(t)}$, we see that the condition of stability is related to the monotonicity of the function $\kappa \mapsto c(\kappa)/\sigma(\kappa)$. Hence we can see directly on the graph of this function (see examples on Figure 1) both the number of family of equilibria and their stability: if the function is decreasing, the family is stable. By contrast it is unstable when the function is increasing. When the difference between f and $\rho M_{\kappa_\infty \Omega_f(t)}$ converges exponentially fast to 0 (on the stable branch), we are able to control the displacement of $\Omega_f(t)$, which gives convergence to $\Omega_\infty \in \mathbb{S}$. We then have convergence of f to a given equilibrium$\rho M_{\kappa_\infty \Omega_\infty}$.

All these results are summarized in the following two theorems. In what follows, we say that a constant is a universal constant when it does not depend on the initial condition f_0 (that is to say, it depends only on ρ, n and the coefficients of the equation ν and τ, and on the exponent s of the Sobolev space H^s in which the result is stated).

Theorem 2. *We have the following instability and exponential stability results around the uniform equilibrium:*

- *Suppose that $\rho < \rho_c$. We define*

$$\lambda = (n-1)\tau_0 (1 - \frac{\rho}{\rho_c}) > 0. \tag{12}$$

There exists a universal constant C, such that if $\|f_0 - \rho\|_{H^s} < \frac{\lambda}{C}$, then for all $t \geqslant 0$, we have

$$\|f(t) - \rho\|_{H^s} \leqslant \frac{\|f_0 - \rho\|_{H^s}}{1 - \frac{C}{\lambda}\|f_0 - \rho\|_{H^s}} e^{-\lambda t}.$$

- *If $\rho > \rho_c$, and if $J_{f_0} \neq 0$, then we cannot have $\kappa_\infty = 0$ in Proposition 2.3: the solution cannot converge to the uniform equilibrium.*

To study the stability around a non-isotropic equilibrium, we fix ρ, and we denote by κ a positive solution to the compatibility equation (we will not write the dependence of c and σ on κ when there is no possible confusion). We denote by $\mathcal{F}_\kappa$ the value of $\mathcal{F}(\rho M_{\kappa\Omega})$ (independent of $\Omega \in \mathbb{S}$).

Theorem 3. *We have the following instability and exponential stability results when starting close to a non-isotropic equilibrium:*

- *Suppose $(\frac{\sigma}{c})'(\kappa) > 0$. For all $s > \frac{n-1}{2}$, there exist universal constants $\delta > 0$ and $C > 0$, such that for any initial condition f_0 satisfying $\|f_0 - \rho M_{\kappa\Omega}\|_{H^s} < \delta$ for some $\Omega \in \mathbb{S}$, there exists $\Omega_\infty \in \mathbb{S}$ such that*

$$\|f - \rho M_{\kappa\Omega_\infty}\|_{H^s} \leqslant C \|f_0 - \rho M_{\kappa\Omega_{f_0}}\|_{H^s} e^{-\lambda t},$$

where the rate is given by

$$\lambda = \frac{c\tau(\sigma)}{\sigma'} \Lambda_\kappa (\frac{\sigma}{c})'. \tag{13}$$

The constant Λ_κ is the best constant for the following weighted Poincaré inequality (see the appendix of [5] for more details on this constant, which does *not depend on Ω):*

$$\int_{\mathbb{S}} |\nabla_\omega g|^2 \, M_{\kappa\Omega} \, \mathrm{d}\omega \geqslant \Lambda_\kappa \Big[\int_{\mathbb{S}} g^2 M_{\kappa\Omega} \, \mathrm{d}\omega - \big(\int_{\mathbb{S}} g \, M_{\kappa\Omega} \, \mathrm{d}\omega \big)^2 \Big]. \tag{14}$$

- *Suppose $(\frac{\sigma}{c})'(\kappa) < 0$. Then any equilibrium of the form $\rho M_{\kappa\Omega}$ is unstable, in the following sense: in any neighborhood of $\rho M_{\kappa\Omega}$, there exists an initial condition f_0 such that $\mathcal{F}(f_0) < \mathcal{F}_\kappa$. Consequently, in that case, we cannot have $\kappa_\infty = \kappa$ in Proposition 2.3.*

3. **Second order phase transition.** Let us now focus on the case where we always have $(\frac{\sigma}{c})' > 0$ for all $\kappa > 0$ (see for example the lowest two curves of Figure 1). In this case, the compatibility equation (9) has a unique positive solution for $\rho > \rho_c$. With the results of the previous subsection about stability and rates of convergence, we obtain the behavior of the solution for any initial condition f_0 with initial mass ρ.

- If $\rho < \rho_c$, then the solution converges exponentially fast towards the uniform distribution $f_\infty = \rho$.
- If $\rho = \rho_c$, the solution converges to the uniform distribution.
- If $\rho > \rho_c$ and $J_{f_0} \neq 0$, then there exists Ω_∞ such that f converges exponentially fast to the von Mises distribution $f_\infty = \rho M_{\kappa\Omega_\infty}$, where $\kappa > 0$ is the unique positive solution to the equation $\rho c(\kappa) = \sigma(\kappa)$.

The special case where $J_{f_0} = 0$ leads to the heat equation $\partial_t f = \tau_0 \Delta_\omega f$. Its solution converges exponentially fast to the uniform distribution, but this solution is not stable under small perturbation of the initial condition. Let us remark that for some particular choice of the coefficients, as in [11], it is also possible to get an algebraic rate of convergence in the second case $\rho = \rho_c$. For example when $\sigma(\kappa) = \kappa$, we have $\|f - \rho\| \leqslant C t^{-\frac{1}{2}}$ for t sufficiently large.

So we can describe the phase transition phenomena by studying the order parameter of the asymptotic equilibrium $c = \frac{|J_{f_\infty}|}{\rho}$, as a function of the initial density ρ.

We have $c(\rho) = 0$ if $\rho \leqslant \rho_c$, and c is a positive continuous increasing function for $\rho > \rho_c$. In the common situation where $\frac{c}{\sigma} = \frac{1}{\rho_c} - a\kappa^{\frac{1}{\beta}} + o(\kappa^{\frac{1}{\beta}})$ when $\kappa \to 0$, it is easy to see, since $c(\kappa) \sim \frac{1}{n}\kappa$ when $\kappa \to 0$, that we have

$$c(\rho) \sim \widetilde{a}(\rho - \rho_c)^\beta, \text{ as } \rho \overset{>}{\to} \rho_c. \tag{15}$$

Since $\frac{c}{\sigma}$ is Lipschitz, we always have $\beta \leqslant 1$. So the first derivative of c is discontinuous at $\rho = \rho_c$. This is the case of a second order phase transition (also called continuous phase transition). The critical exponent β can take arbitrary values in $(0, 1]$, as can be seen by taking $h(|J|)$ such that $\sigma(\kappa) = c(\kappa)(1 + \kappa^{\frac{1}{\beta}})$.

In general, we have the following practical criterion, which ensures a second order phase transition.

Lemma 1. *If $\frac{h(|J|)}{|J|}$ is a non-increasing function of $|J|$, then we have $(\frac{\sigma}{c})' > 0$ for all $\kappa > 0$. In this case, the critical exponent β in (15), if it exists, can only take values in $[\frac{1}{2}, 1]$.*

4. **Hysteresis.**

4.1. **Typical example.** We now turn to a specific example, where all the features presented in the stability study can be seen. We focus on the case where $\nu(|J|) = |J|$, as in [11], but we now take $\tau(|J|) = 1/(1 + |J|)$. From the modeling point of view, this occurs in the Vicsek model with vectorial noise (also called extrinsic noise) [1,2].

In this case, we have $h(|J|) = |J| + |J|^2$, so the assumptions of Lemma 1 are not fulfilled, and the function σ is given by $\sigma(\kappa) = \frac{1}{2}(\sqrt{1 + 4\kappa} - 1)$.

Expanding $\frac{c}{\sigma}$ when κ is large or κ is close to 0, we get

$$\frac{c}{\sigma} = \begin{cases} \frac{1}{n} + \frac{1}{n}\kappa + O(\kappa^2) & \text{as } \kappa \to 0, \\ \frac{1}{\sqrt{\kappa}} + O(\kappa^{-1}) & \text{as } \kappa \to \infty. \end{cases}$$

Consequently, there exist more than one family of non-isotropic equilibria when ρ is close to $\rho_c = n$ (and $\rho > \rho_c$).

The function $\kappa \mapsto \frac{c(\kappa)}{\sigma(\kappa)}$ can be computed numerically. The results are displayed in Figure 2 in dimensions $n = 2$ and $n = 3$.

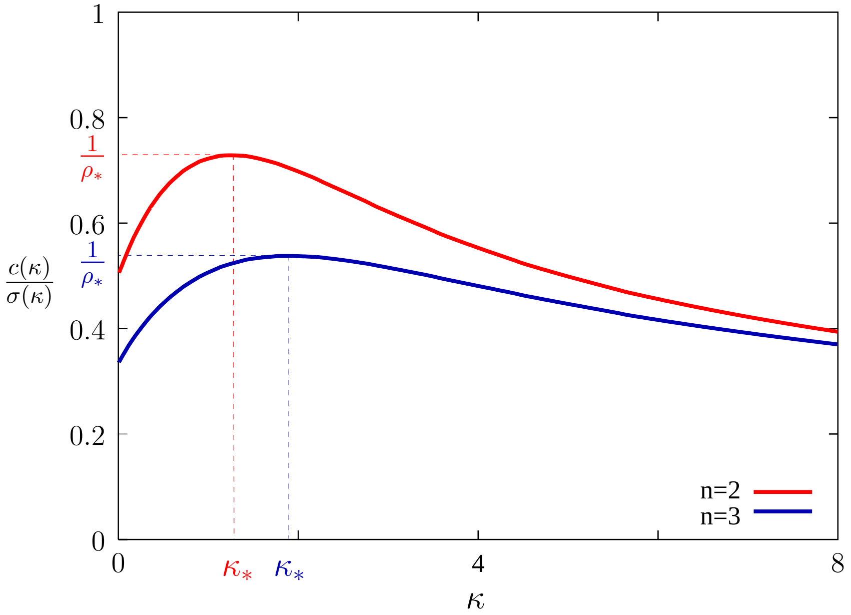


FIGURE 2. The function $\kappa \mapsto \frac{c(\kappa)}{\sigma(\kappa)}$, in dimensions 2 and 3.

We observe the following features:

- There exists a unique critical point κ_* for the function $\frac{c}{\sigma}$, corresponding to its global maximum $\frac{1}{\rho_*}$ (in dimension 2, we obtain numerically $\rho_* \approx 1.3726$ and $\kappa_* \approx 1.2619$, in dimension 3 we get $\rho_* \approx 1.8602$ and $\kappa_* \approx 1.9014$).
- The function $\frac{c}{\sigma}$ is strictly increasing in $[0, \kappa_*)$ and strictly decreasing on (κ_*, ∞).

From these properties, it follows that the solution associated to an initial condition f_0 with mass ρ can exhibit different types of behavior, depending on the three following regimes for ρ.

- If $\rho < \rho_*$, the solution converges exponentially fast to the uniform equilibrium $f_\infty = \rho$.
- If $\rho_* < \rho < n$, there are two families of stable solutions: either the uniform equilibrium $f = \rho$ or the von Mises distributions of the form $\rho M_{\kappa\Omega}$, for $\Omega \in \mathbb{S}$ where κ is the unique solution with $\kappa > \kappa_*$ of the compatibility equation (8). If f_0 is sufficiently close to one of these equilibria, there is exponential convergence to an equilibrium of the same family.

  The von Mises distributions of the other family (corresponding to solution of (8) such that $0 < \kappa < \kappa_*$) are unstable in the sense given in Theorem 3.
- If $\rho > n$ and $J_{f_0} \neq 0$, then there exists $\Omega_\infty \in \mathbb{S}$ such that f converges exponentially fast to the von Mises distribution $\rho M_{\kappa\Omega_\infty}$, where κ is the unique positive solution to the compatibility equation $\rho c(\kappa) = \sigma(\kappa)$.

At the critical point $\rho = \rho_*$, the uniform equilibrium is stable (and for any initial condition sufficiently close to it, the solution converges exponentially fast to it), but the stability of the family of von Mises distribution $\{\rho_* M_{\kappa_*\Omega}, \Omega \in \mathbb{S}\}$ is unknown..

At the critical point $\rho = n$, the family of von Mises distribution $\{nM_{\kappa_c\Omega}, \Omega \in \mathbb{S}\}$ is stable, where κ_c is the unique positive solution of (8). For any initial condition sufficiently close to $nM_{\kappa_c\Omega}$ for some $\Omega \in \mathbb{S}$, there exists Ω_∞ such that the solution converges exponentially fast to $nM_{\kappa_c\Omega_\infty}$. However, in this case, the stability of the uniform distribution $f = n$ is unknown.

As previously, in the special case $J_{f_0} = 0$, the equation reduces to the heat equation and the solution converges to the uniform equilibrium.

Since $c(\kappa)$ is an increasing function of κ, we can invert this relation $\kappa \mapsto c(\kappa)$ into $c \mapsto \kappa(c)$ and express the density $\rho = \frac{\sigma(\kappa(c))}{c}$ as a function of c. The result is depicted in Figure 3 for dimension 2 or 3. With this picture, we recover the phase diagram in a conventional way: the possible order parameters c for the different equilibria are given as functions of ρ. The dashed lines corresponds to branches of equilibria which are unstable.

We can also obtain the corresponding diagrams for the free energy and the rates of convergences. For this particular example, the free energies $\mathcal{F}(\rho)$ and $\mathcal{F}_\kappa$ (we recall that they correspond respectively to the free energy of the uniform distribution and of a von Mises distribution $\rho M_{\kappa\Omega}$ for a positive solution κ of the compatibility equation (8), including both stable and unstable branches) are given by

$$\begin{aligned}
\mathcal{F}(\rho) &= \rho \ln \rho, \\
\mathcal{F}_\kappa &= \rho \ln \rho + \langle \rho \ln M_{\kappa\Omega} \rangle_M - \frac{1}{2}\sigma^2 - \frac{1}{3}\sigma^3 \\
&= \rho \ln \rho - \rho \ln \int e^{\kappa \cos\theta} \mathrm{d}\omega - \frac{1}{6}(\kappa - \sigma) + \frac{2}{3}\sigma\kappa.
\end{aligned}$$

The plots of these functions are depicted in dimensions 2 and 3 are depicted on the left part of Figure 4. Since the functions are very close in the figure for some range of interest, we depict the difference $\mathcal{F}_\kappa - \mathcal{F}(\rho)$ in a more appropriate scale, in the right part of Figure 4. The dashed lines correspond to unstable branches of equilibria.

We observe that the free energy of the unstable non-isotropic equilibria (in dashed line) is always above that of the uniform distribution. There exist $\rho_1 \in (\rho_*, \rho_c)$ and a corresponding solution κ_1 of the compatibility solution (8) (with $\kappa_1 > \kappa_*$, corresponding to a stable family of non-isotropic equilibria) such that $\mathcal{F}_{\kappa_1} = \mathcal{F}(\rho_{\mathcal{F}})$.

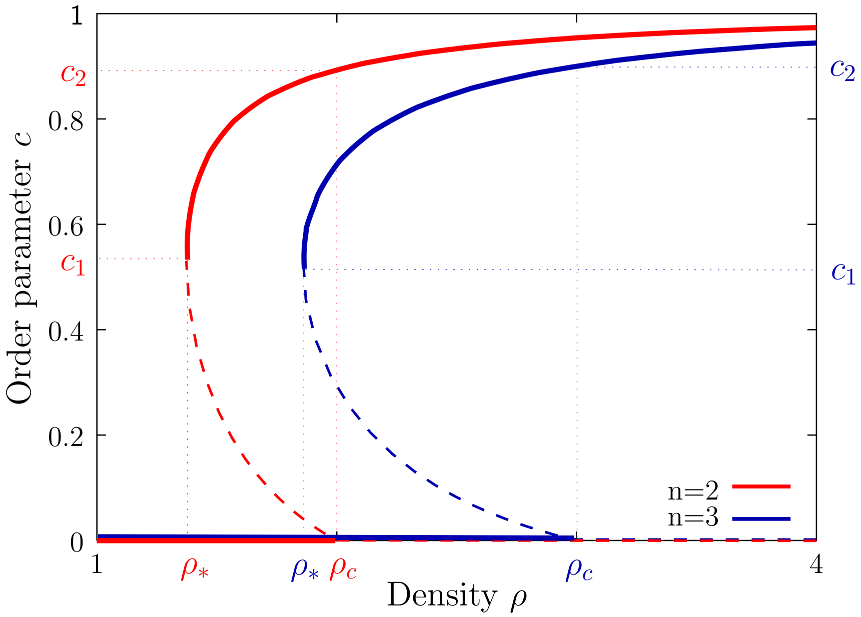


FIGURE 3. Phase diagram of the model with hysteresis, in dimensions 2 and 3.

If $\rho < \rho_1$, the global minimizer of the free energy is the uniform distribution, while if $\rho > \rho_1$, then the global minimum is reached for the family of stable von Mises equilibria. The physical relevance of this value is not clear though, as we will see in the numerical illustration of the next subsection.

The rates of convergence to the stable equilibria, following Theorems 2 and 3, are given by

$$\lambda_0 = (n-1)(1-\frac{\rho}{n}), \text{ for } \rho < \rho_c = n,$$

$$\lambda_\kappa = \frac{1}{1+\sigma}\Lambda_\kappa(1-(\frac{1}{c}-c-\frac{n-1}{\kappa})\sigma(1+2\sigma)), \text{ for } \rho > \rho_*,$$

where λ_0 is the rate of convergence to the uniform distribution ρ, and λ_κ is the rate of convergence to the stable family of von Mises distributions $\rho M_{\kappa\Omega}$, where κ is the unique solution of the compatibility condition (8) such that $\kappa > \kappa_*$. Details for the numerical computation of the Poincaré constant Λ_κ are given in the appendix of [5]. The computations in dimensions 2 and 3 are depicted in Figure 5.

4.2. **Numerical illustrations of the hysteresis phenomenon.** In order to highlight the role of the density ρ as the key parameter for this phase transition, we introduce the probability density function $\widetilde{f} = \frac{f}{\rho}$ and we get

$$\partial_t \widetilde{f} = \tau(\rho|J_{\widetilde{f}}|)\Delta_\omega \widetilde{f} - \nu(\rho|J_{\widetilde{f}}|)\nabla_\omega \cdot (\widetilde{f}\nabla_\omega(\Omega_{\widetilde{f}}\cdot\omega)). \tag{16}$$

When ρ is constant, this equation is equivalent to (1). We now consider ρ as a parameter varying slowly with time (compared to the time scale of the convergence to equilibrium, see Figure 5). If this parameter starts from a value $\rho < \rho_*$, and increases slowly, the only stable distribution is initially the uniform distribution $\widetilde{f} = 1$, and it remains stable. So we expect that the solution stays close to it,

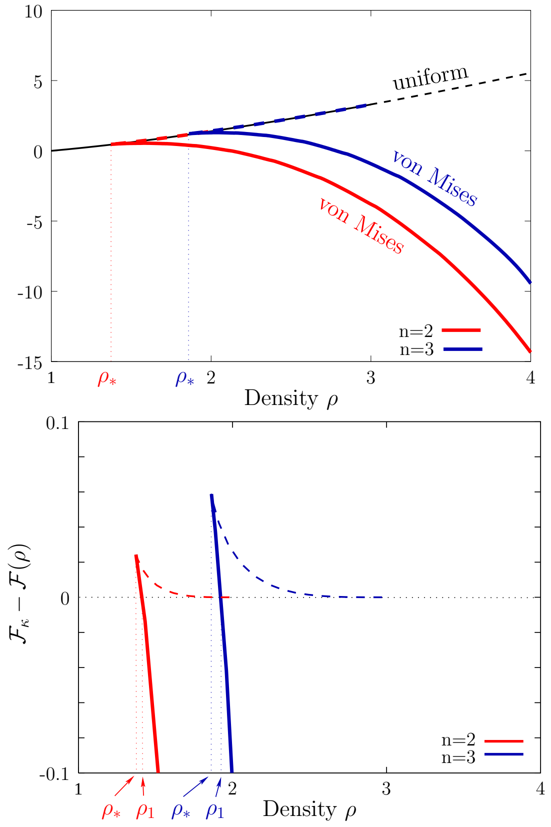


FIGURE 4. Free energy levels of the different equilibria (left), and difference of free energies $\mathcal{F}_\kappa - \mathcal{F}(\rho)$ (right), as functions of the density, in dimensions 2 and 3.

until ρ reaches the critical value ρ_c. For $\rho > \rho_c$, the only stable equilibria are the von Mises distributions, and the solution converges to one of these equilibria. The order parameter defined as $c(\widetilde{f}) = |J_{\widetilde{f}}|$, then jumps from 0 to $c_c = c(\kappa_c)$. If then the density ρ is further decreased slowly, the solution stays close to a von Mises distribution, and the order parameter slowly decreases, until ρ reaches ρ_* back. For $\rho < \rho_*$, the only stable equilibrium is the uniform distribution, and the concentration parameter jumps from $c_* = c(\kappa_*)$ to 0. This is a hysteresis phenomenon: the concentration parameter describes an oriented loop called hysteresis loop.

Let us now present some numerical simulations of the system (16) in dimension $n = 2$. We start with a initial condition which is a small perturbation of the uniform distribution, and we take $\rho = 1.75 - 0.75\cos(\frac{\pi}{T}t)$, with $T = 500$. We use a standard central finite different scheme (with 100 discretization points), implicit in time (with a time step of 0.01). The only problem with this approach is that the solution converges so strongly to the uniform distribution for $\rho < \rho_c$, so after passing ρ_c, the linear rate of explosion for $J_{\widetilde{f}}$ is given by $e^{(\frac{\rho}{\rho_c}-1)t}$, and is very slow when ρ is close to ρ_c. So since $J_{\widetilde{f}}$ is initially very small when passing the threshold $\rho = \rho_c$ we would have to wait extremely long in order to see the convergence to the stable von Mises distribution. To overcome this problem, we adding a threshold ε and

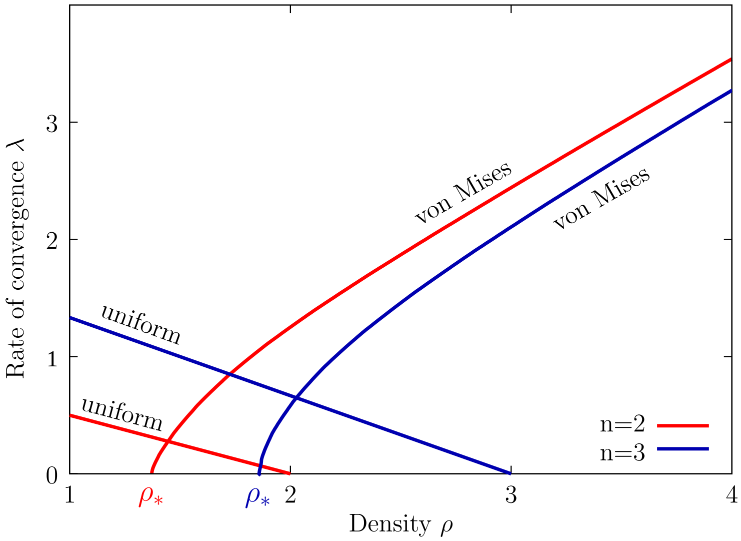


FIGURE 5. Rates of convergence to both types of stable equilibria, as functions of the density ρ, in dimensions 2 and 3.

strengthen $|J_{\widetilde{f}}|$ when $\|\widetilde{f}-1\|_\infty \leqslant \varepsilon$, by

$$\widetilde{f} \rightsquigarrow \widetilde{f} + \max(0, \varepsilon - \|\widetilde{f}-1\|_\infty)\,\Omega_{\widetilde{f}} \cdot \omega.$$

We note that this transformation that we still have $\|\widetilde{f}-1\|_\infty \leqslant \varepsilon$ if it was the case before applying the transformation.

Figure 6 depicts the result of a numerical simulation with a threshold $\varepsilon = 0.02$. We clearly see this hysteresis cycle, which agrees very well with the theoretical diagram. The jumps at $\rho = \rho_*$ and $\rho = \rho_c$ are closer to the theoretical jumps when T is very large. We were not able to see any numerical significance of the value ρ_1 (for which uniform and non-isotropic distributions have the same free energy) in all these numerical simulations. In particular, ρ_1 is close to ρ_* (see Figure 4), so in most of the cases where both uniform and non-isotropic distribution are stable, the uniform distribution is not the global minimizer of the free energy, but in practical, metastability is very strong, and the solution still converges to the uniform distribution.

5. **Conclusion.** In this note, we have given the summary of strong results on the stability and instability of the equilibrium states of the modified Smoluchowski equation (1). This allows to have a precise description of the dynamics of the solution when time goes to infinity: it converges exponentially fast to a fixed equilibrium, with explicit formulas for the rates of convergence. We have also exhibited a specific example in which we observe a first order phase transition with a hysteresis loop (in contrast with the second order phase transition of the original Smoluchowski equation with dipolar potential [11]). The details of the proofs will be found in a longer paper [6] as well as numerical comparisons between the particle and kinetic models to confirm that the hysteresis is really intrinsic to the system and not simply an artifact of the kinetic modeling.

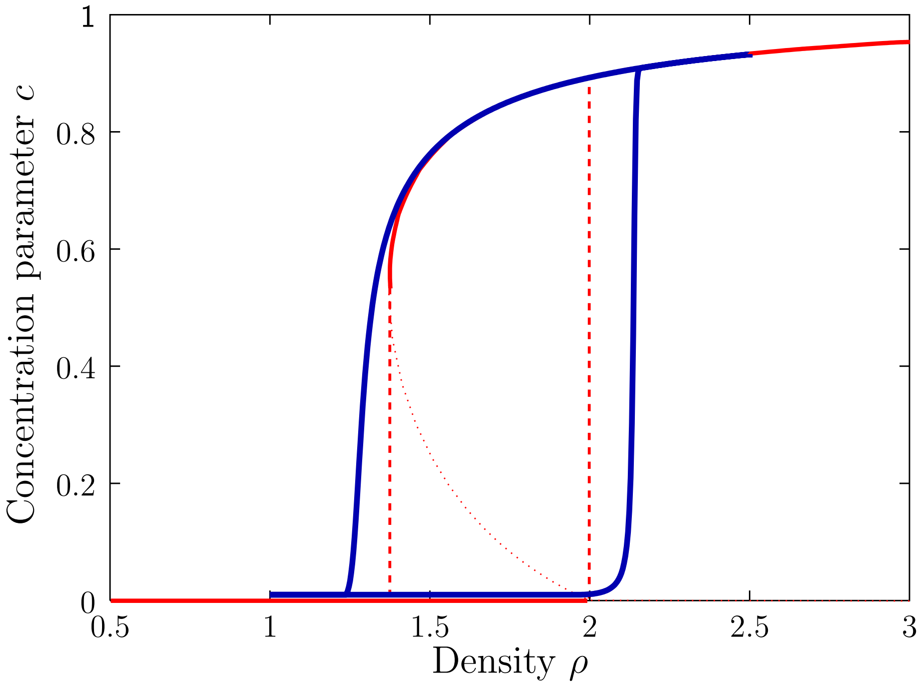


FIGURE 6. Hysteresis loop for the concentration parameter c in a numerical simulation of the kinetic equation (16), with time varying ρ, in dimension 2. The red curve is the theoretical curve, the blue one corresponds to the simulation.

Acknowledgments. JGL is grateful for the opportunity to stay and work at University Paris-Dauphine, under the sponsoring of the french 'Agence Nationale pour la Recherche (ANR)' in the frame of the contract 'CBDif-Fr' (ANR-08-BLAN-0333-01).

REFERENCES

[1] M. Aldana, H. Larralde, and B. Vásquez. On the emergence of collective order in swarming systems: A recent debate. *Int. J. Mod. Phys. B*, 23(18):3459–3483, 2009.

[2] H. Chaté, F. Ginelli, G. Grégoire, and F. Raynaud. Collective motion of self-propelled particles interacting without cohesion. *Phys. Rev. E*, 77(4):046113, 2008.

[3] W. Chen, C. Li, and G. Wang. On the stationary solutions of the 2D Doi-Onsager model. *Nonlin. Anal.*, 73(8):2410–2425, 2010.

[4] P. Constantin, I. G. Kevrekidis, and E. S. Titi. Asymptotic states of a Smoluchowski equation. *Arch. Rat. Mech. Anal.*, 174(3):365–384, 2004.

[5] P. Degond, A. Frouvelle, and J.-G. Liu. Macroscopic limits and phase transition in a system of self-propelled particles. *J. Nonlin. Sci.* 23:427-456, 2013.

[6] P. Degond, A. Frouvelle, and J.-G. Liu. Phase transitions, hysteresis, and hyperbolicity for self-organized alignment dynamics. arXiv preprint 1304.2929

[7] P. Degond and S. Motsch. Continuum limit of self-driven particles with orientation interaction. *Math. Mod. Meth. Appl. Sci.*, 18:1193–1215, 2008.

[8] M. Doi and S. F. Edwards. *The Theory of Polymer Dynamics*, volume 73 of *International Series of Monographs on Physics*. Oxford University Press, Oxford, 1999.

[9] I. Fatkullin and V. Slastikov. Critical points of the Onsager functional on a sphere. *Nonlinearity*, 18:2565–2580, 2005.

[10] I. Fatkullin and V. Slastikov. A note on the Onsager model of nematic phase transitions. *Comm. Math. Sci.*, 3(1):21–26, 2005.

[11] A. Frouvelle and J.-G. Liu. Dynamics in a kinetic model of oriented particles with phase transition. *SIAM J. Math. Anal.*, 44(2):791–826, 2012.

[12] H. Liu, H. Zhang, and P. Zhang. Axial symmetry and classification of stationary solutions of Doi-Onsager equation on the sphere with Maier-Saupe potential. *Comm. Math. Sci.*, 3(2):201–218, 2005.

[13] M. Lucia and J. Vukadinovic. Exact multiplicity of nematic states for an Onsager model. *Nonlinearity*, 23(12):3157–3185, 2010.

[14] W. Maier and A. Saupe. Eine einfache molekulare Theorie des nematischen kristallinflüssigen Zustandes. *Z. Naturforsch.*, 13:564–566, 1958.

[15] L. Onsager. The effects of shape on the interaction of colloidal particles. *Ann. New York Acad. Sci.*, 51(Molecular Interaction):627–659, 1949.

[16] H. Wang and P. J. Hoffman. A unified view on the rotational symmetry of equilibria of nematic polymers, dipolar nematic polymers and polymers in higher dimensional space. *Comm. Math. Sci.*, 6(4):949–974, 2008.

[17] H. Wang and H. Zhou. Multiple branches of ordered states of polymer ensembles with the Onsager excluded volume potential. *Phys. Lett. A*, 372(19):3423–3428, 2008.

[18] H. Zhang and P.W. Zhang. Stable dynamic states at the nematic liquid crystals in weak shear flow. *Phys. D*, 232(2):156–165, 2007.

[19] H. Zhou, H. Wang, M. G. Forest, and Q. Wang. A new proof on axisymmetric equilibria of a three-dimensional Smoluchowski equation. *Nonlinearity*, 18:2815–2825, 2005.

[20] H. Zhou, H. Wang, Q. Wang, and M. G. Forest. Characterization of stable kinetic equilibria of rigid, dipolar rod ensembles for coupled dipole–dipole and Maier–Saupe potentials. *Nonlinearity*, **20** (2007) 277–297.

E-mail address: pierre.degond@math.univ-toulouse.fr
E-mail address: frouvelle@ceremade.dauphine.fr
E-mail address: jliu@phy.duke.edu

LIPSCHITZ METRIC FOR THE TWO-COMPONENT CAMASSA–HOLM SYSTEM

Katrin Grunert

Department of Mathematical Sciences
Norwegian University of Science and Technology
NO-7491 Trondheim, Norway

Helge Holden

Department of Mathematical Sciences
Norwegian University of Science and Technology
NO-7491 Trondheim, Norway
and
Centre of Mathematics for Applications
University of Oslo
NO-0316 Oslo, Norway

Xavier Raynaud

Centre of Mathematics for Applications
University of Oslo
NO-0316 Oslo, Norway
and
SINTEF ICT
Dept. Applied Math., PO Box 124. Blindern
N-0314 Oslo, Norway

Abstract. We construct a Lipschitz metric for conservative solutions of the Cauchy problem on the line for the two-component Camassa–Holm system $u_t - u_{txx} + 3uu_x - 2u_xu_{xx} - uu_{xxx} + \rho\rho_x = 0$, and $\rho_t + (u\rho)_x = 0$ with given initial data (u_0, ρ_0). The Lipschitz metric $d_{\mathcal{D}^M}$ has the property that for two solutions $z(t) = (u(t), \rho(t), \mu_t)$ and $\tilde{z}(t) = (\tilde{u}(t), \tilde{\rho}(t), \tilde{\mu}_t)$ of the system we have $d_{\mathcal{D}^M}(z(t), \tilde{z}(t)) \leq C_{M,T} d_{\mathcal{D}^M}(z_0, \tilde{z}_0)$ for $t \in [0, T]$. Here the measure μ_t is such that its absolutely continuous part equals the energy $(u^2 + u_x^2 + \rho^2)(t)dx$, and the solutions are restricted to a ball of radius M.

1. **Introduction.** The two-component Camassa–Holm (2CH) system, which was first derived in [21, Eq. (43)], is given by

$$u_t - u_{txx} + 3uu_x - 2u_xu_{xx} - uu_{xxx} + \rho\rho_x = 0, \tag{1.1a}$$

$$\rho_t + (u\rho)_x = 0, \tag{1.1b}$$

or equivalently

$$u_t + uu_x + P_x = 0, \tag{1.2a}$$

$$\rho_t + (u\rho)_x = 0, \tag{1.2b}$$

2000 *Mathematics Subject Classification.* Primary: 35Q53, 35B35; Secondary: 35Q20.

Key words and phrases. Two-component Camassa–Holm system, Lipschitz metric, conservative solutions.

where P is implicitly defined by

$$P - P_{xx} = u^2 + \frac{1}{2}u_x^2 + \frac{1}{2}\rho^2. \tag{1.3}$$

The Camassa–Holm equation [6, 7] is obtained by considering the case when ρ vanishes identically. The aim of this article is to present the construction of a Lipschitz metric for this system on the real line with vanishing asymptotics, that is, $u \in H^1$ and $\rho \in L^2$. The conservative solutions to (1.2) are constructed in [15] for nonvanishing asymptotics. A Lipschitz metric for the system with periodic boundary conditions is given in [17]. We here combine the two approaches by constructing a Lipschitz metric for conservative, decaying solutions. The preservation of the energy is needed in the proofs so that the constuction of the metric only applies to vanishing asymptotics. Here we rather describe and motivate the general ideas behind the construction, which we hope can be of interest in the study of other related equations. For more background on the two-component Camassa–Holm system, we refer to [15] and the references therein. For related papers, see [4, 5, 19, 18].

2. **Relaxation of the equations by the introduction of Lagrangian coordinates.** The change of coordinates from Eulerian to Lagrangian coordinates has relaxation properties which are well-known for the Burgers equation, viz.

$$u_t + uu_x = 0. \tag{2.1}$$

Lagrangian coordinates are defined by characteristics

$$y_t(t,\xi) = u(t, y(t,\xi)),$$

which give the position of a particle which moves in the velocity field u and its velocity, known as the Lagrangian velocity, is given by

$$U(t,\xi) = u(t,x), \quad x = y(t,\xi).$$

The method of characteristics consists of rewriting (2.1) in terms of the Lagrangian variables and yields

$$\begin{aligned} y_t &= U, \\ U_t &= 0. \end{aligned} \tag{2.2}$$

Comparing (2.1) to (2.2), we observe that we start with a *nonlinear* and *partial* (derivatives with respect to t and x) differential equation and end up with a *linear* and *ordinary* (derivative only with respect to t) differential equation. We get rid of the nonlinear convection term, and (2.2) is nothing but Newton's law, which states that the acceleration is constant in the absence of forces. A well-known drawback of the change of coordinates from Eulerian to Lagrangian coordinates is that it doubles the dimension of the problem: We start with a scalar equation and end up with a system of dimension two. This is an important issue and we will deal with it in Section 4. However, in return, we gain the possibility to represent a larger class of objects or, more precisely in our case, to increase the regularity of the unknown functions. Let us make this imprecise statement clearer by an example and, to do so, we drop the dependence in t in the notation, as we look at singularities in the space variable. The function $u(x)$ can be represented by its graph $(x, u(x))$ but this graph can itself be represented as a parametric curve, namely, $(y(\xi), U(\xi))$ and, as

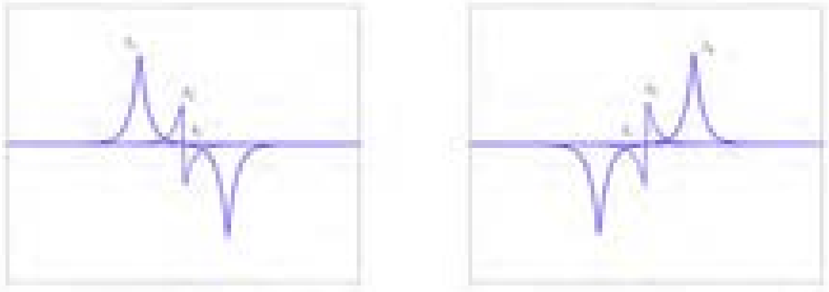


FIGURE 1. Anti symmetric peakon-antipeakon collision, before (on the left) and after (on the right) collision.

we know, the set of graphs is smaller than the set of parametric curves. As far as regularity is concerned, the Heaviside function

$$h(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 & \text{if } x \geq 0, \end{cases}$$

is only of bounded variation but it can be represented in Lagrangian coordinates by the following pair of more regular (in this case Lipschitz) functions

$$y(\xi) = \begin{cases} \xi & \text{if } \xi < 0, \\ 0 & \text{if } \xi \in [0,1), \\ \xi - 1 & \text{if } \xi \geq 1, \end{cases} \qquad H(\xi) = \begin{cases} 0 & \text{if } \xi < 0, \\ \xi & \text{if } \xi \in [0,1), \\ 1 & \text{if } \xi \geq 1 \end{cases} \tag{2.3}$$

Indeed, $(x, h(x))$ and $(y(\xi), H(\xi))$ represent one and the same curve, except for the vertical line joining the origin to the point $(0,1)$. We will return to this example later. The solution of the Camassa–Holm equation (i.e., where ρ vanishes identically) experiences in general wave breaking (i.e., loss of of regularity in the sense that the spatial derivative becomes unbounded while keeping the H^1 norm finite) in finite time ([9, 10, 11]) and the antisymmetric peakon-antipeakon solution, which is described in [18] and depicted in Figure 1, helps us to understand how the solutions can be prolonged in a way which preserves the energy.

At collision time t_c, we have

$$\lim_{t \to t_c} u(t,x) = 0 \text{ in } L^\infty, \qquad \lim_{t \to t_c} u_x(t,0) = -\infty,$$

while the H^1 norm is constant so that $\lim_{t\to t_c} \|u(t,\,\cdot\,)\|_{H^1} = \|u(0,\,\cdot\,)\|_{H^1}$. To obtain the conservative solution, we need to track the amount and the location of the concentrated energy. The function u alone cannot provide this information as $u(t_c,\,\cdot\,)$ is identically zero. Thus, we have to introduce an extra variable to describe the solutions. In Lagrangian variables, it takes the form of the *cumulative energy* $H(t,\xi)$, which is given by

$$H(t,\xi) = \int_{-\infty}^{y(t,\xi)} (u^2 + u_x^2 + \rho^2)(x)dx. \tag{2.4}$$

We will introduce later its counter-part in Eulerian variables. Equation (1.1b) transports the density ρ. Formally, after changing variables, we have $\rho(x)\,dx =$

$\rho(y)\,dy = \rho(y)y_\xi\,d\xi$, so that the Lagrangian variable corresponding to ρ is given by

$$r(t,\xi) = \rho(t, y(t,\xi))y_\xi(t,\xi). \tag{2.5}$$

Next, we rewrite (1.2) in the Lagrangian variables (y, U, H, r). We obtain the following system

$$\begin{aligned} \zeta_t &= U, \\ U_t &= -Q, \\ H_t &= U^3 - 2PU, \\ r_t &= 0, \end{aligned} \tag{2.6}$$

where $\zeta(t,\xi) = y(t,\xi) - \xi$,

$$P(t,\xi) = \frac{1}{4}\int_{\mathbb{R}} \exp(-|y(t,\xi) - y(t,\eta)|)(U^2y_\xi + H_\xi)(t,\eta)d\eta, \tag{2.7}$$

and

$$Q(t,\xi) = -\frac{1}{4}\int_{\mathbb{R}} \operatorname{sign}(y(t,\xi) - y(t,\eta))\exp(-|y(t,\xi) - y(t,\eta)|)(U^2y_\xi + H_\xi)(t,\eta)d\eta. \tag{2.8}$$

See [15] for more details on this derivation. After differentiation, we obtain

$$y_{\xi t} = U_\xi, \tag{2.9a}$$

$$U_{\xi t} = \frac{1}{2}H_\xi + (\frac{1}{2}U^2 - P)y_\xi, \tag{2.9b}$$

$$H_{\xi t} = (3U^2 - 2P)U_\xi - 2QUy_\xi, \tag{2.9c}$$

$$r_t = 0. \tag{2.9d}$$

This system is semilinear and we recognize some features observed earlier for the Burgers equation: We start from a nonlinear partial differential equation and we end up with a system of ordinary differential equations which is *semilinear*. We consider the system as an ordinary differential equation because the order of the spatial derivative is the same on both sides of the equation, so that the existence and uniqueness of solutions can be established by a contraction argument. Finally, it is important to recall in this section the geometric nature of the Camassa–Holm equation. The equation is a geodesic in the group of diffeomorphism for the H^1 norm, see, e.g., [12], as the Burgers equation for the L^2 norm. Using the connection between geometry and fluid mechanics, as presented in [1], the function $t \mapsto y(t,\xi)$ can then be understood as a path in the group of diffeomorphisms. Thus besides the relaxation properties we have just described, this interpretation adds a direct geometrical relevance to use of Lagrangian coordinates, see also [13] for the system.

3. **Semigroup in Lagrangian coordinates.** In [15, Theorem 3.2], we prove by a contraction argument that short-time solutions to (2.6) exist in a Banach space, which we will here denote E and define as follows. Let V be the Banach space defined by

$$V = \{f \in L^\infty \mid f_\xi \in L^2\}$$

and the norm of V is given by $\|f\|_V = \|f\|_{L^\infty} + \|f_\xi\|_{L^2}$. We set E

$$E = V \times H^1 \times V \times L^2$$

with the following norm $\|X\| = \|\zeta\|_V + \|U\|_{H^1} + \|H\|_V + \|r\|_{L^2}$ for any $X = (\zeta, U, H, r) \in E$. Given a constant $M > 0$, we denote by B_M the ball

$$B_M = \{X \in E \mid \|X\| \leq M\}. \tag{3.1}$$

Short-time solutions of (2.9) cannot in general be extended to global solutions. The challenge is to identify an appropriate set of initial data for which one can construct global solutions that at the same time preserve the structure of the equations, allowing us to return to the Eulerian variables. There are intrinsic relations between the variables in (2.9) that need to be conserved by the solution. This is handled by the set $\mathcal{G}$ defined below. In particular, the set $\mathcal{G}$ is preserved by the flow.

Definition 3.1. The set $\mathcal{G}$ is composed of all $(\zeta, U, H, r) \in E$ such that

$$(\zeta, U, H, r) \in \left[W^{1,\infty}\right]^3 \times L^\infty, \tag{3.2a}$$

$$y_\xi \geq 0, H_\xi \geq 0, y_\xi + H_\xi > 0 \text{ almost everywhere, and } \lim_{\xi \to -\infty} H(\xi) = 0, \tag{3.2b}$$

$$y_\xi H_\xi = y_\xi^2 U^2 + U_\xi^2 + r^2 \text{ almost everywhere,} \tag{3.2c}$$

where we denote $y(\xi) = \zeta(\xi) + \xi$.

The condition $y_\xi \geq 0$ implies that the mapping $\xi \mapsto y(\xi)$ is *almost* a diffeomorphism. The solution develop singularities exactly when this mapping ceases to be a diffeomorphism, that is, when $y_\xi = 0$ in some regions. The condition (3.2c) shows that the variables (y, U, H, r) are strongly coupled. In fact, when $y_\xi \neq 0$, we can recover H from (3.2c). It reflects the fact that H_ξ represents, in Lagrangian coordinates, the energy density of u and ρ (that is, $(u^2 + u_x^2 + \rho^2)dx$ in Eulerian coordinates) and therefore, when the solution is smooth, it can be computed from the variables y, U, and r. Note that the coupling between H and (y, U, r) disappears when $y_\xi = 0$, which is precisely the moment when collisions occur and when we need the information H provides on the energy to prolong the solution. The identity makes also clear the smoothing property of the Camassa–Holm system. If $r_0 \geq c > 0$ for some constant c, this property is preserved and then y_ξ never vanishes. The solution keeps the same degree of regularity it has initially, see [15].

As in [15, Theorem 3.6], we obtain the Lipschitz continuity of the semigroup

Theorem 3.2. *For any $\bar{X} = (\bar{y}, \bar{U}, \bar{H}, \bar{r}) \in \mathcal{G}$, the system (2.6) admits a unique global solution $X(t) = (y(t), U(t), H(t), r(t))$ in $C^1(\mathbb{R}_+, E)$ with initial data $\bar{X} = (\bar{y}, \bar{U}, \bar{H}, \bar{r})$. We have $X(t) \in \mathcal{G}$ for all times. If we equip $\mathcal{G}$ with the topology induced by the E-norm, then the mapping $S\colon \mathcal{G} \times \mathbb{R}_+ \to \mathcal{G}$ defined by*

$$S_t(\bar{X}) = X(t)$$

is a Lipschitz continuous semigroup. More precisely, given $M > 0$ and $T > 0$, there exists a constant C_M which depends only on M and T such that, for any two elements $X_\alpha, X_\beta \in \mathcal{G} \cap B_M$, we have

$$\|S_t X_\alpha - S_t X_\beta\| \leq C_M \|X_\alpha - X_\beta\| \tag{3.3}$$

for any $t \in [0, T]$.

4. **Relabeling symmetry.** The equations are well-posed in Lagrangian coordinates. We want to transport this result back to Eulerian coordinates. If the two sets of coordinates were in bijection, then it would be straightforward but, as mentioned earlier, Lagrangian coordinates increase the number of unknowns from two (u and ρ) to four (the components of X), which indicates that such a bijection does not exist. There exists a redundancy in Lagrangian coordinates and the goal of this section is precisely to identify this redundancy, in order to be able to define the correct equivalence classes. This redundancy is also present in the case of the Burgers equation when we define the Cauchy problem for both (2.1) and (2.2). To the initial condition $u(0,x) = u_0(x)$ for (2.1), there corresponds infinitely many parametrizations of the initial conditions for (2.2) given by

$$y(0,\xi) = f(\xi), \qquad\qquad U(0,\xi) = u_0(f(\xi)),$$

for an arbitrary diffeomorphism f. As also mentioned earlier, the representation of a graph is uniquely defined by a single function while there are infinitely many different parametrizations of any given curve. We will use the term *relabeling* for this lack of uniqueness in the characterization of one and the same curve.

We now define the relabeling functions as follows.

Definition 4.1. We denote by G the subgroup of the group of homeomorphisms from $\mathbb{R}$ to $\mathbb{R}$ such that

$$f - \mathrm{Id} \text{ and } f^{-1} - \mathrm{Id} \text{ both belong to } W^{1,\infty}, \tag{4.1a}$$

$$f_\xi - 1 \text{ belongs to } L^2, \tag{4.1b}$$

where Id denotes the identity function. Given $\kappa > 0$, we denote by G_κ the subset G_κ of G defined by

$$G_\kappa = \{f \in G \mid \; \|f - \mathrm{Id}\|_{W^{1,\infty}} + \left\|f^{-1} - \mathrm{Id}\right\|_{W^{1,\infty}} \leq \kappa\}.$$

We refine the definition of $\mathcal{G}$ in Definition 3.1 by introducing the subsets $\mathcal{F}_\kappa$ and $\mathcal{F}$ as

$$\mathcal{F}_\kappa = \{X = (y, U, H, r) \in \mathcal{G} \mid y + H \in G_\kappa\},$$

and

$$\mathcal{F} = \{X = (y, U, H, r) \in \mathcal{G} \mid y + H \in G\}. \tag{4.2}$$

The regularity requirement on the relabeling functions given in Definition 4.1 and the definition of $\mathcal{F}$ are introduced in order to be able to define the action of G on $\mathcal{F}$, that is, for any $X = (y, U, H, r) \in \mathcal{F}$ and any function $f \in G$, the function $(y \circ f, U \circ f, H \circ f, r \circ f f_\xi)$ belongs to $\mathcal{F}$ and we will denote it by $X \circ f$. This corresponds to the relabeling action. Note that relabeling acts differently on *primary* functions, as y, U and H (in this case, we have $(U, f) \mapsto U \circ f$) and on *derivatives* or *densities*, as y_ξ, U_ξ, H_ξ and r (in that case we have $(r, f) \mapsto r \circ f f_\xi$). The space $\mathcal{F}$ is preserved by the governing equation (2.6) and, as expected, the semigroup of solutions in Lagrangian coordinates preserves relabeling, i.e., we have the following result.

Lemma 4.2 ([15, Theorem 4.8])**.** *The mapping S_t is equivariant, that is,*

$$S_t(X \circ f) = S_t(X) \circ f$$

for any $X \in \mathcal{F}$ and $f \in G$.

Now that we have identified the redundancy of Lagrangian coordinates as the action of relabeling, we want to handle it by considering equivalence classes. However, equivalence classes are rather abstract objects which will be hard to work with from an analytical point of view. We consider instead the section defined by $\mathcal{F}_0$, which contains one and only one representative for each equivalence class, so that the quotient $\mathcal{F}/G$ is in bijection with $\mathcal{F}_0$. Let us denote by Π the projection of $\mathcal{F}$ into $\mathcal{F}_0$ defined as

$$\Pi(X) = X \circ (y + H)^{-1}$$

for any $X = (y, U, H, r) \in \mathcal{F}$. By definition, we have that X and $\Pi(X)$ belong to the same equivalence class. We can check that the mapping Π is a projection, i.e., $\Pi \circ \Pi = \Pi$, and that it is also invariant, i.e., $\Pi(X \circ f) = \Pi(X)$. It follows that the mapping $[X] \mapsto \Pi(X)$ is a bijection from $\mathcal{F}/G$ to $\mathcal{F}_0$.

5. **Eulerian coordinates.** In the method of characteristics, once the equation is solved in Lagrangian coordinates, we recover the solution in Eulerian coordinates by setting $u(t,x) = U(t, y^{-1}(t,x))$, where $y^{-1}(t,x)$ denotes—assuming it exists—the inverse of $\xi \mapsto y(t,\xi)$. The Burgers equation and the Camassa–Holm equation develop singularity because y does not remain invertible. In the case of the Burgers equation, u becomes discontinuous but the Camassa–Holm equation enjoys more regularity and u remains continuous. This is a consequence of the preservation of the H^1 norm, but it can also be seen from the Lagrangian point of view. Indeed, even if y is not invertible, we can define $u(t,x)$ as

$$u(t,x) = U(t,\xi) \text{ for any } \xi \text{ such that } x = y(t,\xi).$$

This is well-defined because if there exist ξ_1 and ξ_2 such that $x = y(t,\xi_1) = y(t,\xi_2)$, then $y_\xi(t,\xi) = 0$ for all $\xi \in [\xi_1, \xi_2]$ because y is non-decreasing, see (3.2b). Then, by (3.2c), we get $U_\xi(t,\xi) = 0$ so that $U(t,\xi_1) = U(t,\xi_2)$. Furthermore, as we explained earlier in the case of a peakon-antipeakon collision, some information is needed about the energy to prolong the solution after collision. If y is invertible, we recover the energy density in Eulerian coordinates as

$$(u^2 + u_x^2 + \rho^2)\,dx = \frac{H_\xi}{y_\xi} \circ y^{-1}\,d\xi, \tag{5.1}$$

which corresponds to the push-forward of the measure $H_\xi\,d\xi$ with respect to y, i.e.,

$$(u^2 + u_x^2 + \rho^2)\,dx = y_\#(H_\xi\,d\xi). \tag{5.2}$$

However, when y is not invertible (5.1) cannot be used and $y_\#(H_\xi\,d\xi)$ may not be absolutely continuous so that (5.2) will not hold either. It motivates the introduction of the energy μ defined here as $y_\#(H_\xi\,d\xi)$, which represents the energy of the system. The set $\mathcal{D}$ of Eulerian coordinates is defined as follows.

Definition 5.1. The set $\mathcal{D}$ consists of all triples (u, ρ, μ) such that

1. $u \in H^1$, $\rho \in L^2$, and
2. μ is a positive Radon measure whose absolutely continuous part, μ_{ac}, satisfies

$$\mu_{ac} = (u^2 + u_x^2 + \rho^2)dx. \tag{5.3}$$

It can be shown (see [15, Section 4]) that the identity (3.2c) is somehow equivalent to (5.3) but it is clear that, from an analytical point of view, it easier to deal with an algebraic identity like (3.2c) than with a property like (5.3) which immediately requires tools from measure theory. We can show that $\mathcal{D}$ and $\mathcal{F}_0$ are in bijection, and the mappings between the two are given in the following definition. The first one has been already explained.

Definition 5.2. Given any element X in $\mathcal{F}_0$, then (u, ρ, μ) defined as follows

$$u(x) = U(\xi) \text{ for any } \xi \text{ such that } x = y(\xi),$$
$$\rho(x) = y_{\#}(r d\xi), \quad \mu = y_{\#}(H_\xi d\xi),$$

belongs to $\mathcal{D}$. We denote by $M : \mathcal{F}_0 \to \mathcal{D}$ the map which to any X in $\mathcal{F}_0$ associates (u, ρ, μ).

The mapping, which we denoted by L, from $\mathcal{D}$ to $\mathcal{F}_0$ is defined as follows.

Definition 5.3. For any (u, ρ, μ) in $\mathcal{D}$ let

$$\begin{cases} y(\xi) = \sup\{y \mid \mu((-\infty, y)) + y < \xi\}, \\ H(\xi) = \xi - y(\xi), \\ U(\xi) = u \circ y(\xi), \\ r(\xi) = \rho \circ y(\xi) y_\xi(\xi). \end{cases} \tag{5.4}$$

We can see that the lack of regularity of u, which will occur when μ is singular or very large, is transformed into regions where the function y is constant or almost constant. Using the relabeling degree of freedom, we manage to rewrite functions in L^2 and measures as bounded functions (in L^∞). For example, for the peakon-antipeakon collision depicted in Figure 1, the initial data given by $u_0(x) = \rho_0(x) = 0$ and $\mu = \delta(x)\,dx$, which corresponds to the collision time, t_c, when the total energy is equal to one, yields $r(\xi) = U(\xi) = 0$ with $y(\xi)$ and $H(\xi)$ as defined in (2.3). We can check that, in this case $\delta(x)\,dx = y_{\#}(H_\xi\,d\xi)$. Finally, we define the semigroup T_t of conservative solutions in the original Eulerian variables $\mathcal{D}$ as

$$T_t := M \Pi S_t L.$$

6. **Lipschitz metric for the semigroup.** We apply the construction of the semigroup T_t in Section 5, and we can check, as done in [15, Theorem 5.2], that, for given initial data (u_0, ρ_0, μ_0), if we denote $(u(t), \rho(t), \mu_t) = T_t(u_0, \rho_0, \mu_0)$, then (u, ρ) are weak solutions to (1.2). Moreover,

$$\mu_t(\mathbb{R}) = \mu_0(\mathbb{R})$$

so that the solutions are conservative. Our goal is to define a metric on $\mathcal{D}$ which makes the semigroup Lipschitz continuous. The Lipschitz continuity is a property of a semigroup which can be used to establish its uniqueness, see [3] and [2, Theorem 2.9]. By our construction, a metric for the semigroup T_t is readily available. We can simply transport the topology of the Banach space E from $\mathcal{F}_0$ to $\mathcal{D}$ and obtain, for two elements (u, ρ, μ) and $(\tilde{u}, \tilde{\rho}, \tilde{\mu})$,

$$d_{\mathcal{D}}\big((u, \rho, \mu), (\tilde{u}, \tilde{\rho}, \tilde{\mu})\big) = \|L(u, \rho, \mu) - L(\tilde{u}, \tilde{\rho}, \tilde{\mu})\|_E\,. \tag{6.1}$$

We have

$$d_{\mathcal{D}}\big(T_t(u, \rho, \mu), T_t(\tilde{u}, \tilde{\rho}, \tilde{\mu})\big) = \|\Pi S_t L(u, \rho, \mu) - \Pi S_t L(\tilde{u}, \tilde{\rho}, \tilde{\mu})\|_E\,.$$

It can be proven that the projection Π is continuous (see [15, Lemma 4.6]), but it is not Lipschitz (at least, we have been unable to prove it). Thus, even if S_t is Lipschitz continuous, the semigroup T_t is only continuous with respect to the metric $d_{\mathcal{D}}$ defined by (6.1). In the definition (6.1) of the metric, we let the section $\mathcal{F}_0$ play a special role, but this section is arbitrarily chosen. The set $\mathcal{F}_0$ is by construction nonlinear (because of (3.2c)) and to use a linear norm to measure distances does not respect that. In fact, we want to measure the distance between equivalence classes. A natural starting point is to define, for $X_\alpha, X_\beta \in \mathcal{F}$, $\bar{J}(X_\alpha, X_\beta)$ as

$$\bar{J}(X_\alpha, X_\beta) = \inf_{f,g\in G} \|X_\alpha \circ f - X_\beta \circ g\| . \tag{6.2}$$

The function $\bar{J}$ is relabeling invariant, that is, $\bar{J}(X_\alpha \circ f, X_\beta \circ g) = \bar{J}(X_\alpha, X_\beta)$ and measures precisely the distance between two equivalence classes. However, we have to deal with the fact that the linear norm of E does not play well with relabeling: It is not invariant with respect to relabeling, i.e., we do not have

$$\|X \circ f\| = \|X\| . \tag{6.3}$$

However, such a norm exists. Let

$$B = \{X \in L^\infty \mid X_\xi \in L^1\}.$$

Then,

$$\|X \circ f\|_B = \|X \circ f\|_{L^\infty} + \|X_\xi \circ f f_\xi\|_{L^1} = \|X\|_{L^\infty} + \|X_\xi\|_{L^1} = \|X\|_B .$$

To cope with the lack of relabeling invariance of $\bar{J}$, we introduce J defined as follows.

Definition 6.1. Let $X_\alpha, X_\beta \in \mathcal{F}$, we define $J(X_\alpha, X_\beta)$ as

$$J(X_\alpha, X_\beta) = \inf_{f_1,f_2\in G} \big(\|X_\alpha \circ f_1 - X_\beta\| + \|X_\alpha - X_\beta \circ f_2\| \big). \tag{6.4}$$

The function J is not relabeling invariant, but we have $J(X_\alpha, X_\beta) = 0$ if X_α and X_β both belong to the same equivalence class. Moreover, the relabeling invariance is not strictly needed for our purpose and the following weaker property is enough. Given $X_\alpha, X_\beta \in \mathcal{F}$ and $f \in G_\kappa$, we have

$$J(X_\alpha \circ f, X_\beta \circ f) \le CJ(X_\alpha, X_\beta) \tag{6.5}$$

for some constant C which depends only on κ, see [16]. Note that, if the norm E were invariant, that is, (6.3) were fulfilled, then the function J and $\bar{J}$ would be equivalent, because we would have $\bar{J} \le J \le 2\bar{J}$.

Remark 6.2. We will make use of the following notation. The variable X is used as a standard notation for (y, U, H, r). By the L^∞ norm of X, we mean

$$\|X\|_{L^\infty} = \|y - \mathrm{Id}\|_{L^\infty} + \|U\|_{L^\infty} + \|H\|_{L^\infty} , \tag{6.6}$$

and, by the L^2 norm of the derivative X_ξ, we mean

$$\|X_\xi\|_{L^2} = \|y_\xi - 1\|_{L^2} + \|U_\xi\|_{L^2} + \|H_\xi\|_{L^2} + \|r\|_{L^2} , \tag{6.7}$$

and, similarly,

$$\|X_\xi\|_{L^\infty} = \|y_\xi - 1\|_{L^\infty} + \|U_\xi\|_{L^\infty} + \|H_\xi\|_{L^\infty} + \|r\|_{L^\infty} . \tag{6.8}$$

From J, we obtain a metric d by the following construction.

Definition 6.3. Let $X_\alpha, X_\beta \in \mathcal{F}_0$, we define $d(X_\alpha, X_\beta)$ as

$$d(X_\alpha, X_\beta) = \inf \sum_{i=1}^{N} J(X_{n-1}, X_n) \tag{6.9}$$

where the infimum is taken over all finite sequences $\{X_n\}_{n=0}^N \subset \mathcal{F}_0$ which satisfy $X_0 = X_\alpha$ and $X_N = X_\beta$.

Lemma 6.4. *The mapping $d : \mathcal{F}_0 \times \mathcal{F}_0 \to \mathbb{R}_+$ is a distance on $\mathcal{F}_0$, which is bounded as follows*

$$\frac{1}{2} \left\| X_\alpha - X_\beta \right\|_{L^\infty} \leq d(X_\alpha, X_\beta) \leq 2 \left\| X_\alpha - X_\beta \right\|. \tag{6.10}$$

Proof. The first part of the proof is identical to [16] and we reproduce it here for convenience. For any $X_\alpha, X_\beta \in \mathcal{F}_0$, we have

$$\left\| X_\alpha - X_\beta \right\|_{L^\infty} \leq 2J(X_\alpha, X_\beta). \tag{6.11}$$

We have

$$\begin{aligned} \left\| X_\alpha - X_\beta \right\|_{L^\infty} &\leq \left\| X_\alpha - X_\alpha \circ f \right\|_{L^\infty} + \left\| X_\alpha \circ f - X_\beta \right\|_{L^\infty} \\ &\leq \left\| X_{\alpha,\xi} \right\|_{L^\infty} \left\| f - \mathrm{Id} \right\|_{L^\infty} + \left\| X_\alpha \circ f - X_\beta \right\|_{L^\infty}. \end{aligned} \tag{6.12}$$

It follows from the definition of $\mathcal{F}_0$ that $0 \leq y_\xi \leq 1$, $0 \leq H_\xi \leq 1$ and $|U_\xi| \leq 1$ so that $\left\| X_{\alpha,\xi} \right\|_{L^\infty} \leq 3$. We also have

$$\left\| f - \mathrm{Id} \right\|_{L^\infty} = \left\| (y_\alpha + H_\alpha) \circ f - (y_\beta + H_\beta) \right\|_{L^\infty} \leq \left\| X_\alpha \circ f - X_\beta \right\|_{L^\infty}. \tag{6.13}$$

Hence, from (6.12), we get

$$\left\| X_\alpha - X_\beta \right\|_{L^\infty} \leq 4 \left\| X_\alpha \circ f - X_\beta \right\|_{L^\infty}. \tag{6.14}$$

In the same way, we obtain $\left\| X_\alpha - X_\beta \right\|_{L^\infty} \leq 4 \left\| X_\alpha - X_\beta \circ f \right\|_{L^\infty}$ for any $f \in G$. After adding these two last inequalities and taking the infimum, we get (6.11). For any $\varepsilon > 0$, we consider a finite sequence $\{X_n\}_{n=0}^N \subset \mathcal{F}_0$ such that $X_0 = X_\alpha$ and $X_N = X_\beta$ and $\sum_{i=1}^N J(X_{n-1}, X_n) \leq d(X_\alpha, X_\beta) + \varepsilon$. We have

$$\begin{aligned} \left\| X_\alpha - X_\beta \right\|_{L^\infty} &\leq \sum_{n=1}^N \left\| X_{n-1} - X_n \right\|_{L^\infty} \\ &\leq 2 \sum_{n=1}^N J(X_{n-1}, X_n) \\ &\leq 2(d(X_\alpha, X_\beta) + \varepsilon). \end{aligned}$$

After letting ε tend to zero, we get

$$\left\| X_\alpha - X_\beta \right\|_{L^\infty} \leq 2d(X_\alpha, X_\beta). \tag{6.15}$$

The second inequality in (6.10) follows from the definitions of J and d. Indeed, we have

$$d(X_\alpha, X_\beta) \leq J(X_\alpha, X_\beta) \leq 2 \left\| X_\alpha - X_\beta \right\|.$$

It is left to prove that d defines a metric. The symmetry is intrinsic in the definition of J while the construction of d from J takes care of the triangle inequality. From (6.10), we get that $d(X_\alpha, X_\beta) = 0$ implies $(y_\alpha, U_\alpha, H_\alpha) = (y_\beta, U_\beta, H_\beta)$. By (3.2c), we get that $r_\alpha^2 = r_\beta^2$, but we cannot yet conclude that $r_\alpha = r_\beta$. Let us define $R_\alpha(\xi) = \int_{-\infty}^{\xi} r_\alpha(\eta) e^{-|\eta|}\, d\eta$ and $R_\beta(\xi) = \int_{-\infty}^{\xi} r_\beta(\eta) e^{-|\eta|}\, d\eta$. Then, we have, for any $f \in G$,

$$\begin{aligned} R_\alpha(\xi) - R_\beta(\xi) = -\int_{\xi}^{f(\xi)} r_\alpha(\eta) e^{-|\eta|}\, d\eta + \int_{-\infty}^{\xi} r_\alpha \circ f f_\xi (e^{-|f(\eta)|} - e^{-|\eta|})\, d\eta \\ + \int_{-\infty}^{\xi} (r_\alpha \circ f f_\xi - r_\beta) e^{-|\eta|}\, d\eta, \end{aligned} \tag{6.16}$$

which implies

$$\|R_\alpha - R_\beta\|_{L^\infty} \leq \|f - \mathrm{Id}\|_{L^\infty} + \left\| \int_{-\infty}^{\xi} r_\alpha \circ f f_\xi (e^{-|f(\eta)|} - e^{-|\eta|})\, d\eta \right\|_{L^\infty} + \|r_\alpha \circ f f_\xi - r_\beta\|_{L^2}.$$

We have that

$$\int_{-\infty}^{\xi} r_\alpha \circ f f_\xi (e^{-|f(\eta)|} - e^{-|\eta|})\, d\eta = \int_{-\infty}^{\xi} r_\alpha \circ f f_\xi e^{-|f(\eta)|}(1 - e^{|f(\eta)|-|\eta|})\, d\eta$$

implies

$$\begin{aligned}\left\| \int_{-\infty}^{\xi} r_\alpha \circ f f_\xi (e^{-|f(\eta)|} - e^{-|\eta|})\, d\eta \right\|_{L^\infty} &\leq \left\| e^{|f(\xi)|-|\xi|} - 1 \right\|_{L^\infty} \|r_\alpha\|_{L^2} \left\| e^{-|\xi|} \right\|_{L^2} \\ &\leq C \|r_\alpha\|_{L^2} \|f - \mathrm{Id}\|_{L^\infty},\end{aligned}$$

for $C = e$ if we assume that $\|f - \mathrm{Id}\|_{L^\infty} \leq 1$. Since $X_\alpha \in \mathcal{F}_0$ so that $y_\xi \leq 1$, we get from (3.2c) that $\|r_\alpha\|_{L^2} \leq \|H_\alpha\|_{L^\infty}^{1/2}$. Collecting the results obtained so far, we find that

$$\|R_\alpha - R_\beta\|_{L^\infty} \leq (2 + C \|H_\alpha\|_{L^\infty}^{1/2}) \|X_\alpha \circ f - X_\beta\| \tag{6.17}$$

for any $\|f - \mathrm{Id}\|_{L^\infty} \leq 1$. Let us now assume that $d(X_\alpha, X_\beta) = 0$. For any $\varepsilon > 0$, we can find a sequence such that

$$\sum_{n=1}^{N} \|X_n \circ f_n - X_{n-1}\| \leq \varepsilon.$$

Using (6.13) and (6.14), we get $\|f_n - \mathrm{Id}\|_{L^\infty} \leq \varepsilon$ and prove by induction that

$$\|H_n\|_{L^\infty} \leq \sum_{i=1}^{n} \|X_i \circ f_i - X_{i-1}\|_{L^\infty} + \|H_\alpha\|_{L^\infty}, \tag{6.18}$$

for all $n \leq N$. Indeed, we have

$$\begin{aligned}\|H_{n+1}\|_{L^\infty} &= \|H_{n+1} \circ f_{n+1}\|_{L^\infty} \\ &\leq \|H_{n+1} \circ f_{n+1} - H_n\|_{L^\infty} + \|H_n\|_{L^\infty} \\ &\leq \sum_{i=1}^{n+1} \|X_i \circ f_i - X_{i-1}\|_{L^\infty} + \|H_\alpha\|_{L^\infty},\end{aligned}$$

after using the induction hypothesis. From (6.18), we get

$$\|H_n\|_{L^\infty} \leq \varepsilon + \|H_\alpha\|.$$

Hence, by choosing $\varepsilon \leq 1$, and using repeatedly (6.17), we obtain

$$\begin{aligned}\|R_\alpha - R_\beta\|_{L^\infty} &\leq \sum_{n=1}^{N} \|R_n - R_{n-1}\|_{L^\infty} \\ &\leq (2 + C(\varepsilon + \|H_\alpha\|_{L^\infty})^{1/2}) \sum_{n=1}^{N} \|X_\alpha \circ f - X_\beta\| \\ &\leq (2 + C(\varepsilon + \|H_\alpha\|_{L^\infty})^{1/2})\varepsilon.\end{aligned}$$

After letting ε tend to zero, this last inequality implies that $R_\alpha = R_\beta$ so that $r_\alpha = r_\beta$, which concludes the proof that d is a metric. □

The Lipschitz estimate for the semigroup S_t given in (3.3) is valid for initial data in B_M. Hence, as we want to use the same Lipschitz estimate for any of the X_n in the sequence defining the metric in (6.9), we have to redefine this metric and require that all X_n belong to $\mathcal{F}_0 \cap B_M$. The problem is that B_M is not preserved by the semigroup S_t, and we will not be able to use the same distance at later times. This is why we introduce the set

$$\mathcal{F}^M = \{X = (y, U, H, r) \in \mathcal{F} \mid \|H\|_{L^\infty} \leq M\},$$

which is preserved by *both* relabeling and the semigroup. Note that $\mathcal{F}^M$ has a simple physical interpretation as it corresponds to the set of all solutions which have total energy bounded by M. Moreover, following closely the proof of [16, Lemma 3.4], we obtain that for $X \in \mathcal{F}_0$, the sets B_M and $\mathcal{F}^M$ are in fact equivalent, i.e., there exists $\bar{M}$ depending only on M such that

$$\mathcal{F}_0 \cap \mathcal{F}^M \subset B_{\bar{M}}. \tag{6.19}$$

We set $\mathcal{F}_0^M = \mathcal{F}_0 \cap \mathcal{F}^M$ and define the metric d^M as follows.

Definition 6.5. Let d^M be the distance on $\mathcal{F}_0^M$ which is defined, for any $X_\alpha, X_\beta \in \mathcal{F}_0^M$, as

$$d^M(X_\alpha, X_\beta) = \inf \sum_{n=1}^{N} J(X_{n-1}, X_n) \tag{6.20}$$

where the infimum is taken over all finite sequences $\{X_n\}_{n=0}^N \subset \mathcal{F}_0^M$ such that $X_0 = X_\alpha$ and $X_N = X_\beta$.

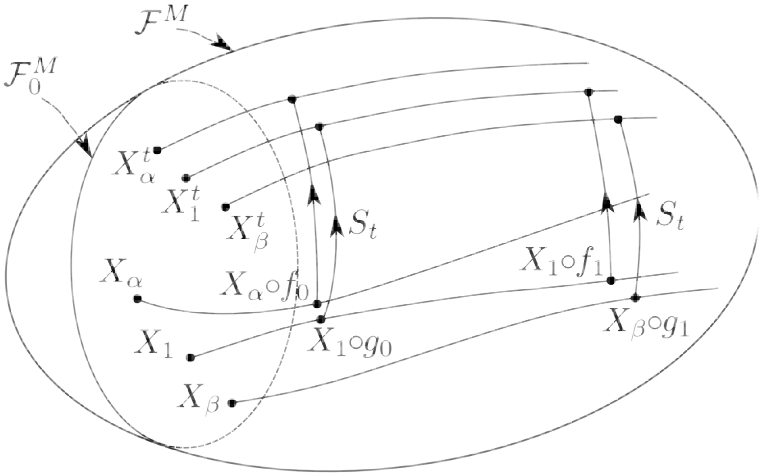


FIGURE 2. Illustration for the construction of the metric. The *horizontal* curves represent points which belong to the same equivalence class.

We can now state our main stability theorem

Theorem 6.6. *Given $T > 0$ and $M > 0$, there exists a constant $C_{M,T}$ which depends only on M and T such that, for any $X_\alpha, X_\beta \in \mathcal{F}_0^M$ and $t \in [0, T]$, we have*

$$d^M(\Pi S_t X_\alpha, \Pi S_t X_\beta) \leq C_{M,T} d^M(X_\alpha, X_\beta). \tag{6.21}$$

In fact due to the use of equivalent notations, the proof of the theorem is identical to [16, Theorem 3.6]. Here, we propose to present a simplified proof where we assume that the norm of E is invariant with respect to relabeling, that is, (6.3) holds. By doing so, we hope that some general ideas behind the construction of the metric becomes clearer. Much of the construction can be understood from the illustration in Figure 2. In this figure, we denote $X_\alpha^t = \Pi S_t(X_\alpha \circ f_0)$, $X_\beta^t = \Pi S_t(X_\beta \circ g_1)$ and $X_1^t = \Pi S_t(X_1 \circ g_0) = \Pi S_t(X_1 \circ f_1)$. Let us imagine the (very improbable) case where the infimum in (6.20) and the infimum in (6.4) both are reached, so that $d^M(X_\alpha, X_\beta) = \|X_\alpha \circ f_0 - X_1 \circ g_0\| + \|X_1 \circ f_1 - X_\beta \circ g_1\|$. Then, we have

$$\begin{aligned} d^M(X_\alpha^t, X_\beta^t) &\leq J(X_\alpha^t, X_1^t) + J(X_1^t, X_\beta^t) \\ &= J(S_t(X_\alpha \circ f_0), S_t(X_1 \circ g_0)) + J(S_t(X_1 \circ f_1), S_t(X_\beta \circ g_1)) \\ &\leq \|S_t(X_\alpha \circ f_0) - S_t(X_1 \circ g_0)\| + \|S_t(X_1 \circ f_1) - S_t(X_\beta \circ g_1)\| \\ &\leq C_{M,T}\big(\|X_\alpha \circ f_0 - X_1 \circ g_0\| + \|X_1 \circ f_1 - X_\beta \circ g_1\|\big) \\ &= C_{M,T} d^M(X_\alpha, X_\beta), \end{aligned}$$

which corresponds to the Lipschitz estimate of Theorem 6.6.

Simplified proof of Theorem 6.6. As we mentioned earlier, when the norm is invariant, then J and $\bar{J}$ are equivalent. Here, it is simpler to consider $\bar{J}$. For any $\varepsilon > 0$, there exist a finite sequence $\{X_n\}_{n=0}^N$ in $\mathcal{F}_0^M$ and functions $\{f_n\}_{n=0}^{N-1}$, $\{g_n\}_{n=0}^{N-1}$ in G such that $X_0 = X_\alpha$, $X_N = X_\beta$ and

$$\sum_{i=1}^N \|X_{n-1} \circ f_{n-1} - X_n \circ g_{n-1}\| \leq d_M(X_\alpha, X_\beta) + \varepsilon. \tag{6.22}$$

Since $B_{\bar{M}}$, where $\bar{M}$ is defined so that (6.19) holds, is preserved by relabeling, we have that $X_n \circ f_n$ and $X_n \circ g_{n-1}$ belong to $B_{\bar{M}}$. From the Lipschitz stability result given in (3.3), we obtain that

$$\|S_t(X_{n-1} \circ f_{n-1}) - S_t(X_n \circ g_{n-1})\| \leq C_{M,T} \|X_{n-1} \circ f_{n-1} - X_n \circ g_{n-1}\|, \tag{6.23}$$

where the constant $C_{M,T}$ depends only on M and T. Introduce

$$\bar{X}_n = X_n \circ f_n,\ \bar{X}_n^t = S_t(\bar{X}_n),\ \text{for } n = 0, \dots, N-1,$$

and

$$\tilde{X}_n = X_n \circ g_{n-1},\ \tilde{X}_n^t = S_t(\tilde{X}_n),\ \text{for } n = 1, \dots, N.$$

Then (6.22) rewrites as

$$\sum_{i=1}^N \left\|\bar{X}_{n-1} - \tilde{X}_n\right\| \leq d_M(X_\alpha, X_\beta) + \varepsilon \tag{6.24}$$

while (6.23) rewrites as

$$\left\|\bar{X}_{n-1}^t - \tilde{X}_n^t\right\| \leq C_{M,T} \left\|\bar{X}_{n-1} - \tilde{X}_n\right\|. \tag{6.25}$$

We have

$$\Pi(\bar{X}_0^t) = \Pi \circ S_t(X_0 \circ f_0) = \Pi \circ (S_t(X_0) \circ f_0) = \Pi \circ S_t(X_0) = \bar{S}_t(X_\alpha)$$

and similarly $\Pi(\tilde{X}_N^t) = \Pi S_t(X_\beta)$. We consider the sequence which consists of $\{\Pi\bar{X}_n^t\}_{n=0}^{N-1}$ and $\bar{S}_t(X_\beta)$. Using the property that $\mathcal{F}^M$ is preserved both by relabeling and by the semigroup, we obtain that $\{\Pi\bar{X}_n^t\}_{n=0}^{N-1}$ and $\bar{S}_t(X_\beta)$ belong to $\mathcal{F}^M$ and therefore also to $\mathcal{F}_0^M$. The endpoints are $\Pi S_t(X_\alpha)$ and $\Pi S_t(X_\beta)$. From the definition of the metric d_M, we get

$$\begin{aligned} d_M(\bar{S}_t(X_\alpha), \bar{S}_t(X_\beta)) &\leq \sum_{n=1}^{N-1} \bar{J}(\Pi\bar{X}_{n-1}^t, \Pi\bar{X}_n^t) + \bar{J}(\Pi\bar{X}_{N-1}^t, \bar{S}_t(X_\beta)) \\ &= \sum_{n=1}^{N-1} \bar{J}(\bar{X}_{n-1}^t, \bar{X}_n^t) + \bar{J}(\bar{X}_{N-1}^t, \tilde{X}_N^t), \end{aligned} \tag{6.26}$$

due to the invariance of $\bar{J}$ with respect to relabeling. By using the equivariance of S_t, we obtain that

$$\begin{aligned} \tilde{X}_n^t = S_t(\tilde{X}_n) &= S_t((\bar{X}_n \circ f_n^{-1}) \circ g_{n-1}) \\ &= S_t(\bar{X}_n) \circ (f_n^{-1} \circ g_{n-1}) = \bar{X}_n^t \circ (f_n^{-1} \circ g_{n-1}). \end{aligned} \tag{6.27}$$

Hence we get from (6.26) that

$$\begin{aligned} d_M(\bar{S}_t(X_\alpha), \bar{S}_t(X_\beta)) &\leq \sum_{n=1}^{N-1} \bar{J}(\bar{X}_{n-1}^t, \tilde{X}_n^t) + \bar{J}(\bar{X}_{N-1}^t, \tilde{X}_N^t) \\ &\leq \sum_{n=1}^{N} \left\|\bar{X}_{n-1}^t - \tilde{X}_n^t\right\| && \text{by (6.10)} \\ &\leq C_{M,T} \sum_{n=1}^{N} \left\|\bar{X}_{n-1} - \tilde{X}_n\right\| && \text{by (6.25)} \\ &\leq C_{M,T}(d_M(X_\alpha, X_\beta) + \varepsilon). \end{aligned}$$

After letting ε tend to zero, we obtain (6.21). □

The Lipschitz stability of the semigroup T_t follows then naturally from Theorem 6.6. It holds on sets of bounded energy. Let $\mathcal{D}^M$ be the subsets of $\mathcal{D}$ defined as

$$\mathcal{D}^M = \{(u, \rho, \mu) \in \mathcal{D} \mid \mu(\mathbb{R}) \leq M\}. \tag{6.28}$$

On the set $\mathcal{D}^M$ we define the metric $d_{\mathcal{D}^M}$ as

$$d_{\mathcal{D}^M}((u, \rho, \mu), (\tilde{u}, \tilde{\rho}, \tilde{\mu})) = d^M(L(u, \rho, \mu), L(\tilde{u}, \tilde{\rho}, \tilde{\mu})), \tag{6.29}$$

where the metric d^M is defined as in Definition 6.5. This definition is well-posed as, from the definition of L, we have that if $(u, \rho, \mu) \in \mathcal{D}^M$, then $L(u, \rho, \mu) \in \mathcal{F}_0^M$.

Theorem 6.7. *The semigroup $(T_t, d_{\mathcal{D}})$ is a continuous semigroup on $\mathcal{D}$ with respect to the metric $d_{\mathcal{D}}$. The semigroup is Lipschitz continuous on sets of bounded energy, that is: Given $M > 0$ and a time interval $[0, T]$, there exists a constant $C_{M,T}$, which only depends on M and T such that for any (u, ρ, μ) and $(\tilde{u}, \tilde{\rho}, \tilde{\mu})$ in $\mathcal{D}^M$, we have*

$$d_{\mathcal{D}^M}(T_t(u, \rho, \mu), T_t(\tilde{u}, \tilde{\rho}, \tilde{\mu})) \leq C_{M,T} d_{\mathcal{D}^M}((u, \rho, \mu), (\tilde{u}, \tilde{\rho}, \tilde{\mu})) \tag{6.30}$$

for all $t \in [0, T]$. Let $(u, \rho, \mu)(t) = T_t(u_0, \rho_0, \mu_0)$, then $(u(t, x), \rho(t, x))$ is weak solution of the Camassa–Holm equation (1.2).

We conclude the section about this metric by mentioning that, even if the construction of the metric is abstract, it can be compared with standard norms, cf. [16, Section 5], so that it can be used in practice, for example in the study of numerical schemes [8, 20].

REFERENCES

[1] V. Arnold and B. Khesin. *Topological Methods in Hydrodynamics.* Springer-Verlag, New York, 1998.

[2] A. Bressan. *Hyperbolic Systems of Conservation Laws. The One-Dimensional Cauchy Problem.* Oxford University Press, Oxford, 2000.

[3] A. Bressan. Contractive metrics for nonsmooth evolutions. In *Nonlinear Partial Differential Equations*, (H. Holden, K.H. Karlsen, eds.), Abel Symposia, Vol. 7, Springer, Berlin-Heidelberg, pp. 13–25, 2012.

[4] A. Bressan and A. Constantin. Global conservative solutions of the Camassa–Holm equation. *Arch. Ration. Mech. Anal.* 183:215–239, 2007.

[5] A. Bressan, H. Holden, and X. Raynaud. Lipschitz metric for the Hunter–Saxton equation. *J. Math. Pures Appl.* 94:68–92, 2010.

[6] R. Camassa and D. D. Holm. An integrable shallow water equation with peaked solutions. *Phys. Rev. Lett* 71(11):1661–1664, 1993.

[7] R. Camassa, D. D. Holm, and J. Hyman. A new integrable shallow water equation. *Adv. Appl. Mech* 31:1–33, 1994.

[8] D. Cohen and X. Raynaud. Convergent numerical schemes for the compressible hyperelastic rod wave equation. *Numerische Mathematik* 122(1):1–59, 2012.

[9] A. Constantin and J. Escher. Global existence and blow-up for a shallow water equation. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)* 26:303–328, 1998.

[10] A. Constantin and J. Escher. Wave breaking for nonlinear nonlocal shallow water equations. *Acta Math.* 181:229–243, 1998.

[11] A. Constantin and J. Escher. On the blow-up rate and the blow-up set of breaking waves for a shallow water equation. Math. Z. 233:75–91, 2000.

[12] A. Constantin and B. Kolev. Least action principle for an integrable shallow water equation. *J. Nonlinear Math. Phys.* 4:471–474, 2001.

[13] J. Escher, M. Kohlman and J. Lenells. The geometry of the two-component Camassa–Holm and Degasperis–Procesi equations. *J. Geom. Phys.* 61(2): 436–452, 2011.

[14] K. Grunert, H. Holden, and X. Raynaud. Lipschitz metric for the periodic Camassa–Holm equation. *J. Differential Equations* 250: 1460–1492, 2011.

[15] K. Grunert, H. Holden, and X. Raynaud. Global solutions for the two-component Camassa–Holm system. *Communications in Partial Differential Equations* 37(12): 2245–2271, 2012.

[16] K. Grunert, H. Holden, and X. Raynaud. Lipschitz metric for the Camassa–Holm equation on the line. *Discrete Contin. Dyn. Syst.* 33:2809–2827, 2013.

[17] K. Grunert, H. Holden, and X. Raynaud. Periodic conservative solutions for the two-component Camassa–Holm system. In *Spectral Analysis, Differential Equations and Mathematical Physics. A Festschrift in Honor of Fritz Gesztesy's 60th Birthday*, (H. Holden, B. Simon, and G. Teschl, eds.), Proc. Symp. Pure Math., Vol. 87, Amer. Math. Soc., 2013, to appear.

[18] H. Holden and X. Raynaud. Global conservative multipeakon solutions of the Camassa–Holm equation. *J. Hyperbolic Differ. Equ.* 4:39–64, 2007.

[19] H. Holden and X. Raynaud. Global conservative solutions of the generalized hyperelastic-rod wave equation. *J. Differential Equations* 233:448–484, 2007.

[20] H. Holden and X. Raynaud. A numerical scheme based on multipeakons for conservative solutions of the Camassa–Holm equation. In *Hyperbolic Problems: Theory, Numerics, Applications*, (S. Benzoni-Gavage, D. Serre, eds.) Springer, Heidelberg, 873–881, 2008.

[21] P. Olver and X. P. Rosenau Tri-Hamiltonian duality between solitons and solitary-wave solutions having compact support. *Phys. Rev. E* 53:1900–1906, 1996.

E-mail address: katrin.grunert@univie.ac.at
E-mail address: holden@math.ntnu.no
E-mail address: xavierra@cma.uio.no

POINTS OF SHOCK WAVE INTERACTION ARE 'REGULARITY SINGULARITIES' IN SPACETIME

Moritz Reintjes

Instituto Nacional de Matemática Pura e Aplicada
Estrada Dona Castorina 110
Rio de Janeiro / Brasil 22460-320, Brasil

Abstract. In this proceedings article we present the results first announced in [9]. The detailed proofs can be found in [7].

In [9, 7], we prove that the regularity of the gravitational metric tensor in spherically symmetric spacetimes cannot be lifted from $C^{0,1}$ to $C^{1,1}$ within the class of $C^{1,1}$ coordinate transformations in a neighborhood of a point of shock wave interaction in General Relativity, without forcing the determinant of the metric tensor to vanish at the point of interaction. This is in contrast to Israel's Theorem [4] which states that such coordinate transformations always exist in a neighborhood of a point on a smooth *single* shock surface. The results imply that points of shock wave interaction represent a new kind of *regularity singularity* for perfect fluid matter sources in the Einstein equations, singularities that lie in physical spacetime, that can form from the evolution of smooth initial data, but at which the spacetime is not locally Minkowskian under any coordinate transformation. In particular, at regularity singularities, delta function sources in the second derivatives of the metric exist in all coordinate systems of the $C^{1,1}$-atlas, but due to cancelation, the curvature tensor is *supnorm bounded*.

1. **Introduction.** In contrast to Newtonian gravity, Albert Einstein's theory of General Relativity (GR) generically predicts the existence of spacetime singularities. Those are points where the gravitational metric, which lies at the heart of GR, suffers a severe lack of regularity. For instance, the singularity at the center of the Schwarzschild metric or at its Schwarzschild radius, where the metric tensor fails to be bounded. The first one is an example of a non-removable singularity which persist in every coordinate system, those singularities are usually characterized by a blow-up in the scalar curvature and they lie outside of physical spacetime. The apparent singularity at the Schwarzschild radius is an example of a removable singularity, i.e., there exist coordinates in which the metric is regular enough to be non-singular. A metric is non-singular if its components, their first and second derivatives and its curvature are bounded and if it is *locally inertial*, that is, around any point p exist coordinates in which the gravitational metric at p is the Minkowski metric up to second order corrections. (The physical interpretation of the metric being locally inertial is that an observer in freefall experiences the physics of special relativity up to second order acceleration effects due to gravity.) However, the metric tensor

2000 *Mathematics Subject Classification.* Primary: 83C05, 83C75; Secondary: 76L05.

Key words and phrases. Spacetime Singularity, Shock Wave Interaction, Conservation Laws, Einstein Euler equations, PDE, Regularity Theory.

The results presented were part of my Ph.D.-thesis-work at the University of California - Davis, supervised by Blake Temple.

is governed by the Einstein equations, which are a system of PDE's, so that the Einstein equations determine the smoothness of the gravitational metric tensor by the evolution they impose. Thus the condition on spacetime that it be non-singular cannot be assumed in the beginning, but must be determined by regularity theorems for the Einstein equations.

For perfect fluids as the sources of matter and energy, this issue becomes all the more interesting and intriguing. Then the Einstein equations impose the GR compressible Euler equations as the evolution equations for the matter fields, and the compressible Euler equations create shock waves out of smooth initial data whenever the flow is sufficiently compressive [1, 11]. At a shock wave, the fluid density, pressure and velocity are discontinuous, so the Einstein equation imply that the metric is only Lipschitz continuous in some coordinates, a regularity which is too low for the metric to be non-singular in those coordinates. However, Israel's theorem asserts that a metric $C^{0,1}$ regular across a smooth *single* shock surface, is lifted to $C^{1,1}$ by the $C^{1,1}$ coordinate map to Gaussian normal coordinates, and this is again smooth enough for spacetime to be non-singular. In [2], Groah and Temple give the first general existence theory for spherically symmetric shock wave solutions of the Einstein-Euler equations allowing for *interacting* shock waves. In Standard Schwarzschild Coordinates (SSC), the gravitational metric is only $C^{0,1}$ at shock waves, and it has remained an open problem as to whether the weak solutions constructed by Groah and Temple could be smoothed to $C^{1,1}$ by coordinate transformation, like the single shock surfaces addressed by Israel.

The negative answer to the open problem of Groah and Temple was given in [9, 7] by proving there do not exist $C^{1,1}$ coordinate transformations that can lift the metric regularity from $C^{0,1}$ to $C^{1,1}$ at a point of shock wave interaction in a spherically symmetric spacetime. Consequently, in contrast to Israel's theorem for single shock surfaces, shock wave solutions cannot be continued as $C^{1,1}$ strong solutions of the Einstein equations beyond the first point of shock wave interaction. The results imply that points of shock wave interaction represent a new kind of non-removable singularity in General Relativity that can form from the evolution of smooth initial data, that lies within physical spacetime, but at which second order metric derivatives are distributional and spacetime is not locally inertial under any $C^{1,1}$ coordinate transformation. Due to cancelation, the Riemann curvature tensor is *sup-norm bounded* and *free of delta function sources* [7]. This result contrasts the common assumption about the metric being $C^{1,1}$ regular, for example, this is assumed in the singularity theorems of Hawking and Penrose, [3]. In this proceedings article we present the main result of [9, 7], sketch its proof and discuss some consequences and open problems.

To state the main result precisely, let $g_{\mu\nu}$ denote a spherically symmetric spacetime metric in SSC, that is, the metric takes the form

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = -A(t,r)dt^2 + B(t,r)dr^2 + r^2 d\Omega^2, \tag{1}$$

where either t or r can be taken to be timelike, and $d\Omega^2 = d\vartheta^2 + \sin^2(\vartheta)d\varphi^2$ is the line element on the unit 2-sphere, c.f. [13]. In Section 3 we make precise the definition of a point of regular shock wave interaction in SSC. Essentially, this is a point where two shock waves enter or leave the point p at distinct speeds, such that the metric is *Lipschitz continuous* across each shock, the Rankine Hugoniot (RH) conditions hold across the shocks, and the SSC Einstein equations hold strongly away from the shocks. The main result of [9, 7] is the following theorem, c.f. [9]:

Theorem 1.1. *Assume p is a point of regular shock wave interaction in SSC, in the sense of Definition 3.1, for the SSC metric $g_{\mu\nu}$. Then there does not exist a $C^{1,1}$ regular coordinate transformation, defined in a neighborhood of p, such that the metric components are C^1 functions of the new coordinates and such that the metric has a nonzero determinant at p.*

The proof of Theorem 1.1 is constructive in the sense that we characterize the Jacobians of (t,r) coordinate transformations that could smooth the components of the gravitational metric in a deleted neighborhood of a point p of regular shock wave interaction, and then prove that any such Jacobian must have a vanishing determinant at p itself. The proof for the full atlas of $C^{1,1}$ coordinate transformations, also allowing for changes of angular variables, can be found in [7].

Our assumptions in Theorem 1.1 apply to the upper half ($t \geq 0$) and the lower half ($t \leq 0$) of a shock wave interaction (at $t = 0$) separately, general enough to include the case of two timelike (or spacelike) interacting shock waves of opposite families that cross at the point p, but also general enough to include the cases of two outgoing shock waves created by the focusing of compressive rarefaction waves, or two incoming shock waves of the same family that interact at p to create an outgoing shock wave of the same family and an outgoing rarefaction wave of the opposite family, c.f. [11]. In particular, our framework is general enough to incorporate the shock wave interaction which was numerically simulated in [14].

Historically, the issue of the smoothness of the gravitational metric tensor across interfaces began with the matching of the interior Schwarzschild solution to the vacuum across an interface, followed by the celebrated work of Oppenheimer and Snyder who gave the first dynamical model of gravitational collapse by matching a pressureless fluid sphere to the Schwarzschild vacuum spacetime across a dynamical interface [6]. In [12], Smoller and Temple extended the Oppenheimer-Snyder model to nonzero pressure by matching the Friedmann metric to a static fluid sphere across a shock wave interface that modeled a blast wave in GR. In his celebrated 1966 paper [4], Israel gave the definitive conditions for regular matching of gravitational metrics at smooth interfaces, by showing that if the second fundamental form is continuous across a single smooth interface, then the RH conditions also hold, and Gaussian normal coordinates provide a locally inertial coordinate system at each point on the surface. In [2] Groah and Temple addressed these issues rigorously in the first general existence theory for shock wave solutions of the Einstein-Euler equations in spherically symmetric spacetimes.

Although points of shock wave interaction are straightforward to construct for the relativistic compressible Euler equations in flat spacetime, we know of no rigorous construction of a point of regular shock wave interaction in GR. However, all evidence indicates points of shock wave interaction to exist, have the structure we assume in SSC, and cannot be avoided in solutions consisting of, say, an outgoing spherical shock wave (the blast wave of an explosion) evolving inside an incoming spherical shock wave (the leading edge of an implosion). Namely, the existence theory of Temple and Groah [2] lends strong support to this claim, establishing existence of weak solutions of the Einstein-Euler equations in spherically symmetric spacetimes. The theory applies to arbitrary numbers of initial shock waves of arbitrary strength, existence is established beyond the point of shock wave interaction, and the regularity assumptions of our theorem are within the regularity class to which the Groah-Temple theory applies. Moreover, the recent work of Vogler and Temple gives a numerical simulation in which two shock waves emerge from a point of interaction where two compression waves focus into a discontinuity in density and velocity, and the numerics demonstrate that the structure of the emerging shock waves meet the assumptions of our theorem.

It is instructive at this point to clarify the difference between the essential $C^{0,1}$ singularities in the metric at points of shock wave interaction, and the essential $C^{0,1}$ singularities at surface layers like the "thin shells" introduced in Israel's illuminating paper [4]. On surface layers, the delta function sources in the energy momentum tensor, T, are the *cause* of the essential $C^{0,1}$ singularity in the metric g, because second derivatives of g must have distributional sources and consequently g cannot be $C^{1,1}$ in any regular coordinate system. For shock wave solutions of the Einstein equations, $G = \kappa T$, the issue is more delicate because T is sup-norm bounded, so that the constraint of G having delta function sources is removed and, at first sight, there is no clear obstacle to the existence of coordinate systems that smooth the metric to $C^{1,1}$. Israel's theorem confirms there is no obstacle to $C^{1,1}$ smoothness in the special case of single shock surfaces, but the methods in [2] are only sufficient to prove existence of solutions in $C^{0,1}$, and the question as to whether there is an obstacle for more complicated solutions with interactions has remained unresolved until the argument in [9, 7] resolved this issue, proving that points of shock wave interaction are non-removable $C^{0,1}$ singularities, where spacetime fails to be locally inertial.

We close the introduction by discussing possible physical implications of the presence of a regularity singularity. Since the gravitational metric is not locally inertial at points of shock wave interaction, it raises the question as to whether regularity singularities provide a physical regime where new general relativistic effects could be observed. We currently work on the question if gravitational waves crossing a regularity singularity pick up some (detectable) effect caused by the singularity [10]. So far, we believe that we isolated a mechanism which gives rise to such an effect. This mechanism is based on the results in [8], which are proofed by an extension of the methods outlined in this article.

2. Preliminaries.

Let g denote a Lorentzian metric g of signature $(-1,1,1,1)$ on a four dimensional spacetime manifold M. We call M a C^k-manifold if it is endowed with a C^k-atlas, a collection of four dimensional local diffeomorphisms from M to $\mathbb{R}^4$, such that any composition of two local diffeomorphisms x and y of the form $x \circ y^{-1}$ is C^k regular. ($x \circ y^{-1}$ is refered to as a coordinate transformation.) In this paper we consider $C^{1,1}$-manifolds, in fact, lowering the regularity to $C^{1,1}$ is the crucial step allowing for a smoothing of the metric in the presence of a single shock wave, (c.f. (19) and Theorem 5).

We use standard index notation for tensors whereby Greek vs Roman indices distinguish coordinate systems, and repeated up-down indices are assumed summed from 0 to 3. Under coordinate transformation, tensors transform by contraction with the Jacobian $J^\mu_j = \frac{\partial x^\mu}{\partial x^j}$, J^j_ν denotes the inverse Jacobian, and indices are raised and lowered with the metric and its inverse g^{ij}, which transform as bilinear forms, $g_{\mu\nu} = J^i_\mu J^j_\nu g_{ij}$, c.f. [15]. We use the fact that a matrix of functions J^μ_j is the Jacobian of a regular local coordinate transformation if and only if

$$J^\mu_{i,j} = J^\mu_{j,i} \quad \text{and} \quad Det\left(J^\mu_j\right) \neq 0, \tag{2}$$

where $f_{,j} = \frac{\partial f}{\partial x^j}$ denotes partial differentiation with respect to the coordinate x^j and $Det\left(J^\mu_j\right)$ denotes the determinant of the Jacobian.

In this article we consider the *Einstein-Euler equations*

$$G^{ij} = \kappa T^{ij}, \tag{3}$$

which couples the metric tensor g_{ij} to the perfect fluid sources

$$T^{ij} = (p+\rho)u^i u^j + pg^{ij}, \tag{4}$$

through the second order Einstein curvature tensor $G^{ij} \equiv R^{ij} - \frac{1}{2}Rg^{ij}$, and

$$Div\ T = 0 \tag{5}$$

follows from $Div\ G = 0$. Here κ is the coupling constant, ρ is the energy density, u_i the 4-velocity, and p the pressure, c.f. [15]. Equation (5) reduces to the relativistic compressible Euler equations when g_{ij} is the Minkowski metric, and the Euler equations close when an equation of state (e.g. $p = p(\rho)$) is imposed. Shock waves form from smooth solutions of the relativistic compressible Euler equations when the initial data is sufficiently compressive, [11].

Across a smooth shock surface Σ, the RH jump conditions hold,

$$[T^{\mu\nu}]n_\nu = 0, \tag{6}$$

where $[f] = f_L - f_R$ denotes the jump in f from right to left across Σ, and n_ν is the surface normal. The RH condition (6) is equivalent to the weak formulation of (5) across Σ, c.f. [11].

In this paper we restrict to time dependent spherically symmetric metrics in Standard Schwarzschild Coordinates (1). Recall, in SSC the metric takes the form,

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = -A(t,r)dt^2 + B(t,r)dr^2 + r^2 d\Omega^2.$$

A spherically symmetric metric can generically be transformed to SSC, c.f. [15]. The Einstein equations for a metric in SSC are given by

$$\begin{aligned}
B_r + B\frac{B-1}{r} &= \kappa AB^2 rT^{00} && (7)\\
B_t &= -\kappa AB^2 rT^{01} && (8)\\
A_r - A\frac{1+B}{r} &= \kappa AB^2 rT^{11} && (9)\\
B_{tt} - A_{rr} + \Phi &= -2\kappa ABr^2 T^{22}\,, && (10)
\end{aligned}$$

with

$$\Phi = -\frac{BA_tB_t}{2AB} - \frac{B_t^2}{2B} - \frac{A_r}{r} + \frac{AB_r}{rB} + \frac{A_r^2}{2A} + \frac{A_rB_r}{2B}\,.$$

Note that the first three Einstein equations in SSC imply that the metric cannot be any smoother than Lipschitz continuous if the source T is discontinuous, for example, $T^{ij} \in L^\infty$, and throughout this article we make the assumption that A and B are Lipschitz continuous, i.e., $C^{0,1}$ functions, of t and r.

3. **A point of regular shock wave interaction in SSC.** In this paper we restrict attention to radial shock waves, by which we mean hypersurfaces Σ locally parameterized by

$$\Sigma(t,\vartheta,\varphi) = (t, x(t), \vartheta, \varphi), \tag{11}$$

across which A and B are $C^{0,1}$ and T satisfies (6). Then, for each t, Σ is a 2-sphere with radius $x(t)$ and center $r = 0$.) For radial hypersurfaces in SSC, the angular variables play a passive role and it suffices to work with the so-called shock curve γ, that is, the shock surface Σ restricted to the (t,r)-plane, $\gamma(t) = (t, x(t))$, with normal 1-form $n_\sigma = (\dot{x}, -1)$.

For radial shock surfaces (11) in SSC, the RH jump conditions (6) take the simplified form

$$\begin{aligned} \left[T^{00}\right]\dot{x} &= \left[T^{01}\right], & (12)\\ \left[T^{10}\right]\dot{x} &= \left[T^{11}\right]. & (13) \end{aligned}$$

Now suppose two timelike shock surfaces Σ_i are parameterized in SSC by

$$\Sigma_i(t,\theta,\phi) = (t, x_i(t), \theta, \phi), \qquad i = 1,2. \tag{14}$$

Let $\gamma_i(t) = (t, x_i(t))$ denote their corresponding restrictions to the (t,r)-plane, with normal 1-forms $(n_i)_\sigma = (\dot{x}_i, -1),$. Denoting with $[\cdot]_i$ the jump across the $i - th$ shock curve the RH conditions read, in correspondence to (12)-(13),

$$\begin{aligned} \left[T^{00}\right]_i\dot{x}_i &= \left[T^{01}\right]_i, & (15)\\ \left[T^{10}\right]_i\dot{x}_i &= \left[T^{11}\right]_i. & (16) \end{aligned}$$

For the proof of Theorem 1.1 it suffices to restrict attention to the lower $(t < 0)$ or upper $(t > 0)$ part of a shock wave interaction that occurs at $t = 0$. That is, it suffices to impose conditions on either the lower or upper half plane

$$\mathbb{R}^2_- = \{(t,r) : t < 0\} \quad \text{or} \quad \mathbb{R}^2_+ = \{(t,r) : t > 0\},$$

respectively, whichever half plane contains two shock waves that intersect at p with distinct speeds. Thus, without loss of generality, let $t < 0$ and let $\gamma_i(t) = (t, x_i(t))$, $(i = 1,2)$, be two shock curves in the lower (t,r)-plane intersecting in a point $(0, r_0)$, for $r_0 > 0$, that is, $x_1(0) = r_0 = x_2(0)$.

We now define the notion of a point of *regular shock wave interaction in SSC*. By this we mean a point p where two shock waves collide with distinct speeds, such that the metric is smooth away from the shock curves and Lipschitz continuous across each shock curve, allowing for a discontinuous $T^{\mu\nu}$ and the RH condition to hold. Recall, we assume without loss of generality a lower shock wave interaction in $\mathbb{R}^2_-$.

Definition 3.1. Let $r_0 > 0$, and let $g_{\mu\nu}$ be an SSC metric in $C^{0,1}$, defined on $\mathcal{N} \cap \overline{\mathbb{R}^2_-}$, where $\mathcal{N} \subset \mathbb{R}^2$ is a neighborhood of a point $p = (0, r_0)$ of intersection of two timelike shock curves $\gamma_i(t) = (t, x_i(t)) \in \mathbb{R}^2_-$, $t \in (-\epsilon, 0)$. Assume the shock speeds $\dot{x}_i(0) = \lim\limits_{t\to 0}\dot{x}_i(t)$ exist and are distinct, $\dot{x}_1(0) \neq \dot{x}_2(0)$, and let $\hat{\mathcal{N}}$ denote the neighborhood consisting of all points in $\mathcal{N} \cap \mathbb{R}^2_-$ not in the closure of the two intersecting curves $\gamma_i(t)$. Then we say that p is a point of regular shock wave interaction in SSC if:

(i) The pair (g, T) is a strong solution of the SSC Einstein equations (7)-(10) in $\hat{\mathcal{N}}$, with $T^{\mu\nu} \in C^0(\hat{\mathcal{N}})$ and $g_{\mu\nu} \in C^2(\hat{\mathcal{N}})$.

(ii) The limits of $T^{\mu\nu}$ and of metric derivatives $g_{\mu\nu,\sigma}$ exist on both sides of each shock curve $\gamma_i(t)$ for all $\in (-\epsilon, 0)$.

(iii) The jumps in the metric derivatives $[g_{\mu\nu,\sigma}]_i(t)$ are C^1 function with respect to t for $i = 1,2$ and for $t \in (-\epsilon, 0)$.

(iv) The limits

$$\lim_{t\to 0}[g_{\mu\nu,\sigma}]_i(t) = [g_{\mu\nu,\sigma}]_i(0)$$

exist for $i = 1,2$.

(v) The metric g is continuous across each shock curve $\gamma_i(t)$ separately, but no better than Lipschitz continuous in the sense that, for each i there exists μ, ν such that

$$[g_{\mu\nu,\sigma}]_i (n_i)^\sigma \neq 0$$

at each point on γ_i, $t \in (-\epsilon, 0)$ and

$$\lim_{t\to 0} [g_{\mu\nu,\sigma}]_i (n_i)^\sigma \neq 0.$$

(vi) The stress tensor T is bounded on $\mathcal{N} \cap \overline{\mathbb{R}^2_-}$ and satisfies the RH conditions

$$[T^{\nu\sigma}]_i (n_i)_\sigma = 0$$

at each point on $\gamma_i(t)$, $i = 1, 2$, $t \in (-\epsilon, 0)$, and the limits of these jumps exist up to p as $t \to 0$.

4. **A Necessary and Sufficient Condition for Smoothing.** In this section we derive a necessary and sufficient pointwise condition on the Jacobians of a coordinate transformation that it lift the regularity of a $C^{0,1}$ metric tensor to C^1 across a single shock surface Σ. This is the starting point for Sections 5 and 6.

We begin with the transformation law

$$g_{\alpha\beta} = J^\mu_\alpha J^\nu_\beta g_{\mu\nu}, \tag{17}$$

for the metric components at a point on a hypersurface Σ for a general $C^{1,1}$ coordinate transformation $x^\mu \mapsto x^\alpha$, where, as customary, the indices indicate the coordinate system. Let $J^\mu_\alpha = \frac{\partial x^\mu}{\partial x^\alpha}$ denote the Jacobian of the transformation.

Now, assume the metric components $g_{\mu\nu}$ are only Lipschitz continuous across Σ with respect to coordinates x^μ, that is, smooth away from Σ but with (possibly) discontinuous derivatives across Σ. Then, differentiating (17) with respect to $\frac{\partial}{\partial x^\gamma}$ and taking the jump across Σ we obtain

$$[g_{\alpha\beta,\gamma}] = J^\mu_\alpha J^\nu_\beta [g_{\mu\nu,\gamma}] + g_{\mu\nu} J^\mu_\alpha [J^\nu_{\beta,\gamma}] + g_{\mu\nu} J^\nu_\beta [J^\mu_{\alpha,\gamma}]\,, \tag{18}$$

where $[\cdot]$ denotes the jump across the shock surface Σ, c.f. (6). Now, $g_{\alpha\beta}$ is in C^1 if and only if $[g_{\alpha\beta,\gamma}] = 0$ for every α, β, $\gamma \in \{0, ..., 3\}$. Thus, by (18), $g_{\alpha\beta}$ is C^1 regular if and only if

$$[J^\mu_{\alpha,\gamma}] J^\nu_\beta g_{\mu\nu} + [J^\nu_{\beta,\gamma}] J^\mu_\alpha g_{\mu\nu} + J^\mu_\alpha J^\nu_\beta [g_{\mu\nu,\gamma}] = 0. \tag{19}$$

We now exploit linearity in (19) to solve for the $[J^\mu_{\alpha,\gamma}]$ associated with a given $C^{1,1}$ coordinate transformation. To this end, suppose we are given a single radial shock surface Σ in SSC, introduced in (11). Now the SSC angular variables play a passive role, and thus we henceforth restrict to (t, r)-coordinate transformations.

Lemma 4.1. *Let $g_{\mu\nu} = -A(t,r)dt^2 + B(t,r)dr^2 + r^2 d\Omega^2$ be a given metric in SSC, c.f. (1), let Σ denote a single radial shock surface (11) across which g is Lipschitz continuous. Then the unique solution $[J^\mu_{\alpha,\gamma}]$ of (19) which satisfies the integrability condition, (c.f. (2)),*

$$[J^\mu_{\alpha,\beta}] = [J^\mu_{\beta,\alpha}]\ , \tag{20}$$

is given by:

$$\begin{aligned} [J^t_{0,t}] &= -\frac{1}{2}\left(\frac{[A_t]}{A} J^t_0 + \frac{[A_r]}{A} J^r_0\right); \qquad [J^t_{0,r}] = -\frac{1}{2}\left(\frac{[A_r]}{A} J^t_0 + \frac{[B_t]}{A} J^r_0\right) \\ [J^t_{1,t}] &= -\frac{1}{2}\left(\frac{[A_t]}{A} J^t_1 + \frac{[A_r]}{A} J^r_1\right); \qquad [J^t_{1,r}] = -\frac{1}{2}\left(\frac{[A_r]}{A} J^t_1 + \frac{[B_t]}{A} J^r_1\right) \end{aligned}$$

$$\begin{aligned}
[J^r_{0,t}] &= -\frac{1}{2}\left(\frac{[A_r]}{B}J^t_0 + \frac{[B_t]}{B}J^r_0\right); \qquad [J^r_{0,r}] = -\frac{1}{2}\left(\frac{[B_t]}{B}J^t_0 + \frac{[B_r]}{B}J^r_0\right) \\
[J^r_{1,t}] &= -\frac{1}{2}\left(\frac{[A_r]}{B}J^t_1 + \frac{[B_t]}{B}J^r_1\right); \qquad [J^r_{1,r}] = -\frac{1}{2}\left(\frac{[B_t]}{B}J^t_1 + \frac{[B_r]}{B}J^r_1\right). \quad (21)
\end{aligned}$$

(*We use the notation $\mu, \nu \in \{t, r\}$ and $\alpha, \beta \in \{0, 1\}$, so that t and r are used to denote indices whenever they appear on the Jacobian J.*)

Condition (19) is a necessary and sufficient condition for $[g_{\alpha\beta,\gamma}] = 0$ at a point on a smooth single shock surface. One can further prove (21) to be necessary and sufficient for lifting the metric regularity to $C^{1,1}$ in a neighborhood of a single shock curve, provided J^μ_α is the Jacobian of an actual coordinate transformation, that is, J^μ_α satisfies the integrability condition (2) everywhere on that neighborhood.

5. **Metric Smoothing on Single Shock Surfaces.** In this section we sketch an alternative constructive proof of Israel's Theorem for spherically symmetric spacetimes. For this we address the issue of how to obtain Jacobians of actual coordinate transformations defined on a whole neighborhood of a shock surface that satisfy (21). That is, we need a set of functions J^μ_α that satisfies (21), and also satisfies the integrability condition (2) in a whole neighborhood.

Theorem 5.1. (Israel's Theorem) *Suppose $g_{\mu\nu}$ is an SSC metric that is $C^{0,1}$ across a radial shock surface γ, such that it solves the Einstein equations (7) - (10) strongly away from γ, and assume $T^{\mu\nu}$ is everywhere bounded and in C^0 away from γ. Then around each point p on γ there exists a $C^{1,1}$ coordinate transformation of the (t, r)-plane, defined in a neighborhood $\mathcal{N}$ of p, such that the transformed metric components $g_{\alpha\beta}$ are $C^{1,1}$ functions of the new coordinates, if and only if the RH jump conditions (12), (13) hold on γ in a neighborhood of p.*

The main step is to construct Jacobians acting on the (t, r)-plane that satisfy the smoothing condition (21) on the shock curve, the condition that guarantees $[g_{\alpha\beta,\gamma}] = 0$. The following lemma gives an explicit formula for functions J^μ_α satisfying (21). The main point is that, in the case of single shock curves, both the RH jump conditions and the Einstein equations are necessary and sufficient for such functions J^μ_α to exist.

Lemma 5.1. Let p be a point on a single shock curve γ across which the SSC metric $g_{\mu\nu}$ is Lipschitz continuous and away from which $g_{\mu\nu}$ solves the Einstein equations (7) - (10) strongly in a neighborhood $\mathcal{N}$ of p. Then, defined on $\mathcal{N}$, there exists a set of functions J^μ_α $C^{0,1}$ across γ , that satisfies the smoothing condition (21) on $\gamma \cap \mathcal{N}$ if and only if the RH condition (12)-(13) holds on $\gamma \cap \mathcal{N}$. Furthermore, all J^μ_α that are in $C^{0,1}$ across $\mathcal{N}$ and satisfy (21) on $\gamma \cap \mathcal{N}$ are given by

$$\begin{aligned}
J^t_0(t,r) &= \frac{[A_r]\phi(t) + [B_t]\omega(t)}{4A \circ \gamma(t)}\,|x(t) - r| + \Phi(t,r) \\
J^t_1(t,r) &= \frac{[A_r]\nu(t) + [B_t]\zeta(t)}{4A \circ \gamma(t)}\,|x(t) - r| + N(t,r) \\
J^r_0(t,r) &= \frac{[B_t]\phi(t) + [B_r]\omega(t)}{4B \circ \gamma(t)}\,|x(t) - r| + \Omega(t,r) \\
J^r_1(t,r) &= \frac{[B_t]\nu(t) + [B_r]\zeta(t)}{4B \circ \gamma(t)}\,|x(t) - r| + Z(t,r)\,, \qquad (22)
\end{aligned}$$

for arbitrary functions Φ, Ω, Z, $N \in C^1(\mathcal{N})$, where

$$\phi = \Phi \circ \gamma, \quad \omega = \Omega \circ \gamma, \quad \nu = N \circ \gamma, \quad \zeta = Z \circ \gamma\,. \tag{23}$$

To complete the proof of Israel's Theorem, it remains to show the functions J^μ_α defined in (22) to be integrable to coordinates, i.e., that they satisfy the integrability condition (2) in a whole neighborhood. For this, substitute (22) into (2) and choose for N and Z arbitrary smooth functions, then (2) reduces to a system of two linear first order PDE's for the unknown functions Φ and Ω. This system is well-posed and the only obstacle to solutions Φ and Ω with the necessary C^1 regularity, is the presence of discontinous source terms $f(t)\ H(x(t)-r)$ and $h(t)\ H(x(t)-r)$, where $H(\cdot)$ denotes the Heaviside step function and f and h are C^1 functions depending on the jumps in the metric derivatives and the functions in (23). Israel's theorem is now a consequence of the following lemma which states that the discontinuous terms vanish precisely when the RH jump conditions hold on γ. (See [7] for details.)

Lemma 5.2. Assume the SSC metric $g_{\mu\nu}$ is $C^{0,1}$ across γ and solves the first three Einstein equations strongly away from γ. Then the coefficients f and g of $H(X)$ (introduced above) vanish on γ if and only if the RH conditions (12)-(13) hold on γ.

6. Shock Wave Interactions as Regularity Singularities.

The main step in the proof of Theorem 1.1 is to first prove the result for the smaller atlas of coordinate transformations of the (t,r)-plane. (See [7] for the complete proof of Theorem 1.1, taking into account the full atlas.) We formulate the main step precisely for lower shock wave interactions in $\mathbb{R}^2_-$ in the following theorem, which is the topic of this section. A corresponding result applies to upper shock wave interactions in R^2_+, as well as two wave interactions in a whole neighborhood of p.

Theorem 6.1. *Suppose that p is a point of regular shock wave interaction in SSC, in the sense of Definition 3.1, for the SSC metric $g_{\mu\nu}$. Then there does not exist a $C^{1,1}$ coordinate transformation $x^\alpha \circ (x^\mu)^{-1}$ of the (t,r)-plane, defined on $\mathcal{N} \cap \mathbb{R}^2_-$ for a neighborhood $\mathcal{N}$ of p in $\mathbb{R}^2$, such that the metric components $g_{\alpha\beta}$ are C^1 functions of the coordinates x^α in $\mathcal{N} \cap \mathbb{R}^2_-$ and such that the metric has a non-vanishing determinant at p, (that is, such that $\lim_{q\to p} Det\,(g_{\alpha\beta}(q)) \neq 0$).*

The proof of Theorem 6.1 mirrors the constructive proof of Israel's Theorem 5.1 in that it uses the extension (24) of ansatz (22) to construct all $C^{1,1}$ coordinate transformations that can smooth the gravitational metric to $C^{1,1}$ in a neighborhood of a point p of regular shock wave interaction. The negative conclusion is then reached by proving that any such coordinate transformation must have a vanishing Jacobian determinant at p. To begin with, we generalize (22) to the case of two shock curves:

Lemma 6.1. Let p be a point of regular shock wave interaction in SSC in the sense of Definition 3.1, corresponding to the SSC metric $g_{\mu\nu}$ defined on $\mathcal{N} \cap \overline{\mathbb{R}^2_-}$. Then the following is equivalent:

(i) There exists a set of functions J^μ_α, defined on $\mathcal{N} \cap \overline{\mathbb{R}^2_-}$, which is $C^{0,1}$ across $\gamma_i \cap \mathcal{N}$ and satisfies the smoothing condition (21) on $\gamma_i \cap \mathcal{N}$, for $i = 1, 2$.

(ii) The RH condition (15)-(16) holds on each shock curve $\gamma_i \cap \mathcal{N}$, for $i = 1, 2$.

Moreover, if (i) or (ii) is valid, then the functions J^μ_α assume the canonical form

$$\begin{aligned} J^t_0(t,r) &= \sum_i \alpha_i(t)\,|x_i(t)-r| + \Phi(t,r), \\ J^t_1(t,r) &= \sum_i \beta_i(t)\,|x_i(t)-r| + N(t,r), \\ J^r_0(t,r) &= \sum_i \delta_i(t)\,|x_i(t)-r| + \Omega(t,r), \\ J^r_1(t,r) &= \sum_i \epsilon_i(t)\,|x_i(t)-r| + Z(t,r)\,, \end{aligned} \tag{24}$$

where $\Phi, \Omega, Z, N \ \in C^1(\mathcal{N}\cap\overline{\mathbb{R}^2_-})$ and

$$\begin{aligned} \alpha_i(t) &= \frac{[A_r]_i\ \phi_i(t) + [B_t]_i\ \omega_i(t)}{4A\circ\gamma_i(t)}, \qquad \beta_i(t) = \frac{[A_r]_i\ \nu_i(t) + [B_t]_i\ \zeta_i(t)}{4A\circ\gamma_i(t)}, \\ \delta_i(t) &= \frac{[B_t]_i\ \phi_i(t) + [B_r]_i\ \omega_i(t)}{4B\circ\gamma_i(t)}, \qquad \epsilon_i(t) = \frac{[B_t]_i\ \nu_i(t) + [B_r]_i\ \zeta_i(t)}{4B\circ\gamma_i(t)}\ , \end{aligned} \tag{25}$$

with

$$\phi_i = \Phi\circ\gamma_i, \quad \omega_i = \Omega\circ\gamma_i, \quad \zeta_i = Z\circ\gamma_i, \quad \nu_i = N\circ\gamma_i\,. \tag{26}$$

Equation (24) gives a canonical form for all functions J^μ_α that meet the necessary and sufficient condition (21) for $[g_{\alpha\beta,\gamma}] = 0$. Assuming for contradiction that J^μ_α is integrable to a coordinate system, the free functions Φ, Ω, Z, N must meet the integrability condition (2) and be C^1 regular. In contrast to the single shock case, after substituting (24) into (2) additional mixed terms in the coefficients f and h of the discontinuous terms appear, and unlike f and g in the single shock case, these mixed terms do not vanish by the RH conditions alone. But, as a consequence of the C^1 regularity of Φ, Ω, Z and N, f and h must vanish on the shock curves, which imposes an additional constraint. Taking the limit of this constraint to the point p of shock wave interaction is the essential step for the proof of Theorem 6.1, recorded in the next lemma:

Lemma 6.2. Let $p\in\mathcal{N}$ be a point of regular shock wave interaction in SSC in the sense of Definition 3.1. Then if the integrability condition

$$J^\mu_{\alpha,\beta} = J^\mu_{\beta,\alpha} \tag{27}$$

holds in $\mathcal{N}\cap\mathbb{R}^2_-$ for the functions J^μ_α defined in (24), then

$$\frac{1}{4B}\left(\frac{\dot{x}_1\dot{x}_2}{A} + \frac{1}{B}\right)[B_r]_1[B_r]_2\,(\dot{x}_1-\dot{x}_2)\,(\phi_0\zeta_0 - \nu_0\omega_0) = 0. \tag{28}$$

To finish the proof of Theorem 6.1, observe that the first three terms in (28) are nonzero by our assumption that shock curves are non-null, and have distinct speeds at $t=0$. Thus (28) implies

$$(\phi_0\zeta_0 - \nu_0\omega_0) = 0. \tag{29}$$

But, using the canonical form (24) restricted to the shock curve and taking the determinant of the resulting J^μ_α leads directly to

$$Det\ (J^\mu_\alpha\circ\gamma_i(t)) = \left(J^t_0J^r_1 - J^t_1J^r_0\right)|_{\gamma_i(t)} = \phi_i(t)\zeta_i(t) - \nu_i(t)\omega_i(t). \tag{30}$$

Since J^μ_α is continuous, we obtain the same limit $t\to0$ for $i=1,2$,

$$\lim_{t\to0^+} Det\ (J^\mu_\alpha\circ\gamma_i(t)) = \phi_i(0)\zeta_i(0) - \nu_i(0)\omega_i(0) = \phi_0\zeta_0 - \nu_0\omega_0. \tag{31}$$

This together with (29) immediately implies $\det\left(g_{\alpha\beta}\right)\big|_p = 0$. □

We remark, at first sight the construction of the Jacobian capable of smoothing the metric seems to go through as in the single shock case, c.f. Lemma 6.1. But taking the limit to the point p of shock wave interaction, the C^1 regularity of the free functions, expressing that $[g_{\alpha\beta,\gamma}]$ vanishes at shocks, has the effect of freezing out all the freedom in Φ, Ω, Z, N, thereby forcing condition (29), implying that the determinant of the Jacobian must vanish at p. The answer was not apparent until the very last step, and thus we find the result quite remarkable and surprising.

7. The Loss of Locally Inertial Frames.

Definition 7.1. We call x^j locally inertial at p if the metric g_{ij} in coordinates x^j satisfies:

(i) $g_{ij}(p) = \eta_{ij}$, $\eta_{ij} = diag(-1, 1, 1, 1)$,
(ii) $g_{ij,l}(p) = 0$ for all $i, j, l \in \{0, ..., 3\}$,
(iii) $g_{ij,kl}$ are in L^∞ in every compact neighborhood of p where the coordinates are defined.

This condition ensures that the physical equations in curved spacetime differ from their flat counterparts by only *gravitational effects*, which are second order in the metric derivatives. By Theorem 1.1, there exist distributional second order derivatives of the metric in every neighborhood of p. Therefore, locally inertial frames cannot exist at a point of shock wave interaction [9, 7].

8. **Conclusion.** Our results show that points of shock wave interaction give rise to a new kind of (mild) singularity which is different from the well known singularities of General Relativity. The famous examples of singularities are either non-removable singularities beyond physical spacetime, (for example the center of the Schwarzschild and Kerr metrics, and the Big Bang singularity in cosmology where the curvature cannot be bounded), or else they are removable in the sense that they can be transformed to locally inertial points of a regular spacetime under coordinate transformation, (for example, the apparent singularity at the Schwarzschild radius and any apparent singularity at smooth shock surfaces that become regularized by Israel's Theorem, [4, 12, 13]). In contrast, points of shock wave interaction are non-removable $C^{0,1}$ singularities that propagate in physically meaningful spacetimes, such that the curvature is uniformly bounded, but the spacetime is essentially *not* locally inertial at the singularity and second order metric derivatives are distributional in every coordinate system. We name these *regularity singularities*. We believe that such singularities in perfect fluids are fundamental to the mathematical theory of GR shock waves.

Since the gravitational metric tensor is not locally inertial at points of shock wave interaction, it begs the question as to whether there are general relativistic gravitational effects at points of shock wave interaction that cannot be predicted from the compressible Euler equations in special relativity alone. At a regularity singularity, the unbounded second derivatives in g cancel out in the Riemann curvature tensor [7], but the curvature is not the only measurable effect of the gravitational field, so one would expect there to exist measurable general relativistic effects at points of shock wave interaction that are physical. Indeed, even if there are dissipativity terms, like those of the Navier Stokes equations, which regularize the gravitational metric at points of shock wave interaction, our results assert that the steep gradients in the second derivatives of the metric tensor at small viscosity cannot be removed uniformly while keeping the metric determinant uniformly bounded away from zero.

We thus wonder whether shock wave interactions might provide a physical regime where new general relativistic effects might be observed. In fact, we currently work on the question if gravitational waves crossing a regularity singularity pick up some (detectable) effect due to the singularity and, so far, our results look quite promising that this is the case [10].

Acknowledgments. I am grateful to Blake Temple for supervising my dissertation and proposing to work on the subject of this article for my thesis. I thank Lars Andersson and the Max Planck Institute for Gravitational Physics in Golm, Germany for their funding and hospitality during the period from August to December 2012.

REFERENCES

[1] D. Christodoulou, "The Formation of Shocks in 3-Dimensional Fluids", E.M.S. Monographs in Mathematics, 2007, ISBN: 978-3-03719-031-9.

[2] J. Groah and B. Temple, "Shock-Wave Solutions of the Einstein Equations with Perfect Fluid Sources: Existence and Consistency by a Locally Inertial Glimm Scheme", Memoirs AMS, Vol. 172, Number 813, November 2004, ISSN 0065-9266.

[3] S.W. Hawking and G.F.R. Ellis, "The Large Scale Structure of Spacetime", Cambridge University Press, 1973.

[4] W. Israel, *Singular hypersurfaces and thin shells in general relativity*, Il Nuovo Cimento, Vol. XLIV B, No. 1, 1966, pp. 1-14.

[5] P. Lax, *Hyperbolic systems of conservation laws, II*, Comm. Pure Appl. Math., 10(1957), pp. 537-566.

[6] J. R. Oppenheimer and J. R. Snyder, *On continued gravitational contraction*, Phys. Rev., 56 (1939), pp. 455-459.

[7] M. Reintjes, *The 'regularity singularity' at points of general relativistic shock wave interactions*, arXiv:1112.1803, v.2, Dec. 2012, 47 pages.

[8] M. Reintjes, *At points of GR-shock-wave interaction gravitational effects can enter the laws of physics to first order in every coordinate system*, in preparation, (Sep. 2013).

[9] M. Reintjes and B. Temple, *Points of general relativistic shock wave interaction are 'regularity singularities' where space-time is not locally flat*, Proc. R. Soc. A, 2012, vol. 468 no. 2146, 2962-2980.

[10] M. Reintjes and B. Temple, *The scattering of gravity waves by regularity singularities* in preparation, (Sep. 2013).

[11] [10.1007/978-1-4612-0873-0] J. Smoller, "Shock Waves and Reaction-Diffusion Equations," Springer-Verlag, 1983.

[12] J. Smoller and B. Temple, *Shock wave solutions of the Einstein equations: the Oppenheimer-Snyder model of gravitational collapse extended to the case of non-zero pressure*, Archive Rational Mechanics and Analysis, 128 (1994), pp. 249-297, Springer-Verlag 1994.

[13] J. Smoller and B. Temple, *Shock wave cosmology inside a black hole*, PNAS, Vol. 100, no. 20, 2003, pp. 11216-11218.

[14] Z. Vogler and B. Temple, *Simulation of general relativistic shock wave interactions by a locally inertial Godunov method featuring dynamical time dilation*, Proc. R. Soc. A, 2012, 468:1865-1883; doi:10.1098/rspa.2011.0355.

[15] S. Weinberg, "Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity", John Wiley & Sons, New York, 1972.

E-mail address: moritzreintjes@googlemail.com

TWO FLUID FLOW IN POROUS MEDIA

Michael Shearer

Dept. of Mathematics, N.C. State University
Raleigh, NC 27695, USA

Abstract. The Gray-Hassanizadeh model, for flow in a porous medium of two immiscible fluids such as oil and water, is a scalar equation for the evolution of the saturation of one of the fluids. The model is based on Darcy's law, coupled with a constitutive equation for the capillary pressure that incorporates a rate-dependence to capture the relaxation of interfacial energy towards equilibrium. The model has interesting properties, including the structure of traveling waves explored in this paper. In particular, we find that for certain forms of relative permeability, there are undercompressive shocks that are degenerate in that the corresponding smooth traveling wave drops to zero saturation in finite time, due to the singularity in the PDE at zero saturation. In the second half of the paper, we report on a two-dimensional stability result for the shock wave, regarded as a plane wave in two dimensions. We find a criterion for linearized stability that predicts that some Lax shocks are stable, while others are unstable. This analysis relates to the Saffman-Taylor instability [9] and is a version of a result of Yortsos and Hickernell [15] for stability of smooth traveling waves of the full system including capillary pressure. Whereas matched asymptotics are used in [15], at the hyperbolic level of this paper the analysis of stability and instability is much more transparent.

1. **Introduction.** The flow of two immiscible fluids in a homogeneous porous medium is relevant for a variety of applications, including oil recovery, groundwater flow with contaminants, and carbon sequestration. It is widely acknowledged that simple models of such flow omit important features of the flow, such as the strong heterogeneity of typical subsurface geology, including the presence of fissures, and the stochastic nature of the pore distribution. Nonetheless, simple models can capture gross macroscopic features of the propagation of a front separating the two fluids. Two features are of primary interest, namely (i) the structure and speed of the interface between the fluids, and (ii) the stability of the front.

In this paper, I summarize and refine recent results concerning plane wave solutions of the model of Gray and Hassanizadeh [4, 5], and present a result that characterizes when the plane waves are stable or unstable according to linear analysis. The stability result, slightly generalizing an earlier version [13], mimics the classical calculation of Saffman and Taylor [9]. Of particular interest is the fact that although plane shocks are determined from a scalar conservation law (the Buckley-Leverett equation [2]), the multidimensional stability analysis involves a system of equations.

2000 *Mathematics Subject Classification.* Primary: 35K25, 35L67; Secondary: 76S99.

Key words and phrases. Porous media, shock waves, traveling waves, conservation laws.

The author is supported by NSF grant DMS 0968258.

The main result is that there is a stability boundary that separates pairs of left and right states that correspond to stable shocks from those that are subject to a fingering instability analogous to the Saffman-Taylor instability.

In Section 2, I formulate the model system, based on Darcy's law for conservation of momentum. Darcy's law expresses the proportionality of fluid velocity and pressure gradient. The extension to two-phase flow is generally taken to be of the same form in each of the two fluids, but with constants of proportionality that depend on the local volume fraction (saturation) of the fluids as well as their viscosities. The pressures in the two fluids are consequently different in general, and should be related through the effect of surface tension at fluid interfaces. It is not settled in the literature how to do this in a mixture theory of flow in porous media, so the pressure difference, known as the capillary pressure is typically related to the saturations in an ad hoc way, guided by experimental evidence.

In Section 3 results on planar traveling waves are presented. In [14] an interesting traveling wave was identified numerically, dubbed a *sharp shock* by the authors. Here, we explain how sharp shocks appear as orbits for a vector field, and demonstrate their presence in a numerical simulation of an initial value problem that generates a pair of plane wave solutions, one of which is a sharp shock. The main stability results are derived in Section 4. The foundation of the linearized stability argument has been improved from the treatment in [12], and we consider more general relative permeability functions.

2. **Darcy's Law and Capillary Pressure.** We consider horizontal flow of two immiscible fluids in a homogenous isotropic porous medium. Each fluid has it's own saturation, the volume fraction of pore space occupied by that fluid. For definiteness, consider the fluids to be water and oil. If we let $u(x, y, t)$ denote the saturation of water, then the saturation of oil is $1 - u$. Similarly, each fluid has it's own pressure p and velocity $\mathbf{v}$. To begin with, we distinguish between the two fluids with superscripts ^{w,o} for water, oil respectively. Darcy's law was originally formulated as an empirical law relating the flow rate of a single fluid through a porous medium to spatial changes in pressure. The generalization to two-fluid flow is used widely to replace the law of conservation of momentum:

$$\mathbf{v}^w = -\lambda^w(u)\nabla p^w; \quad \mathbf{v}^o = -\lambda^o(u)\nabla p^o. \tag{1}$$

This law states that the velocity of each fluid is proportional to it's pressure gradient, with a constant of proportionality, the mobility λ depending on the local saturation. The mobilities are related to saturations through the relative permeabilities $k^{w,o}$, with positive constants K, the permeability of the medium, and $\mu^{w,o}$, the viscosities:

$$\lambda^w(u) = K\frac{k^w(u)}{\mu^w}; \quad \lambda^o(u) = K\frac{k^o(1-u)}{\mu^o}. \tag{2}$$

The following notation is convenient to simplify the statement of the governing equations:

$$\mathbf{v}^T = \mathbf{v}^w + \mathbf{v}^o; \quad \Lambda = \lambda^w + \lambda^o.$$

We will also focus on the saturation and pressure of water as the unknowns. For this reason, we write

$$\lambda = \lambda^w, \quad p = p^w, \quad p^c = p^o - p.$$

The pressure difference p^c is the *capillary pressure* that results from capillary forces at the interface between the two fluids at the microscopic pore level. The resulting pressure jump is proportional to mean curvature and surface tension. Then it is natural for the capillary pressure, representing this pressure jump averaged locally within the porous medium, to depend on the local saturation u.

Mass conservation of the water phase balances the rate of change of the saturation against the volume flux:

$$\varphi \frac{\partial u}{\partial t} + \nabla \cdot \mathbf{v}^{\mathbf{w}} = 0. \tag{3}$$

The porosity φ of the medium, the ratio of pore volume to total volume, is taken to be constant. Mass conservation of oil leads to a similar equation for the evolution of the oil saturation $1-u$. Consequently, the aggregate fluid is incompressible:

$$\nabla \cdot \mathbf{v}^{\mathbf{T}} = 0, \tag{4}$$

We use Darcy's law (1) to reduce the system of equations to two equations for the unknowns u, p, p^c :

$$\begin{array}{rcll} \varphi \dfrac{\partial u}{\partial t} - \nabla \cdot (\lambda(u)\nabla p) & = & 0, & (a) \\ \nabla \cdot (\Lambda(u)\nabla p + (\Lambda(u) - \lambda(u))\nabla p^c) & = & 0. & (b) \end{array} \tag{5}$$

In equation (5)(b), the variable p^c needs to be specified with a constitutive law. The classic formulation lets p^c be a decreasing function of saturation:

$$p^c = p^e(u),$$

where $p^e(u)$ is the equilibrium pressure at a particular saturation u. The Gray-Hassanizadeh model introduces a rate-dependence in the capillary pressure, which in it's simplest form is

$$p^c = p^e(u) - \tau u_t. \tag{6}$$

In this model, $\tau > 0$ is a relaxation time for the capillary pressure to relax to equilibrium (for which $u_t = 0$).

The relative permeabilities $k^w(u)$ and $k^o(1-u)$ are typically specified as powers of their respective saturations:

$$k^w(u) = \kappa^w u^m; \quad k^o(1-u) = \kappa^o(1-u)^n. \tag{7}$$

Here, the exponents can be fractional, with $m \geq 1, n \geq 1$. For $m, n > 1$, $f(u)$ has a classical 'S' shape; a common choice is $m = n = 2$, but this value turns out to separate different behavior in traveling wave solutions for the Gray-Hassanizadeh model [14], as we discuss in the next section.

3. **Traveling Waves.**

We consider one-dimensional traveling waves, in which the flow is entirely in one direction, which we can take as parallel to the x-axis. Then the fluid velocities are parallel to the axis, and from (4) we see that $\mathbf{v}^T = (V, 0)$ is a function of t, which we assume is constant in order to realize traveling waves. Correspondingly, from (5)(b) we have

$$\Lambda(u)p_x + (\Lambda(u) - \lambda(u))p^c_x = V.$$

We eliminate p_x in favor of p^c_x :

$$p_x = \frac{\Lambda(u) - \lambda(u)}{\Lambda(u)} p^c_x + \frac{1}{\Lambda(u)} V$$

Then equation (5)(a) becomes

$$u_t + f(u)_x = -\left(H(u)p^c_x\right)_x, \tag{8}$$

in which

$$f(u) = \frac{\lambda(u)}{\Lambda(u)}\frac{V}{\varphi}, \quad H(u) = (\Lambda(u) - \lambda(u))\frac{\lambda(u)}{\varphi\Lambda(u)}. \tag{9}$$

Traveling wave solutions of (8) with speed $s > 0$ correspond to saturations $u(x,t) = \overline{u}(x - st)$. Let $\xi = x - st$, with $d_\xi = \frac{d}{d\xi}$. Then $\overline{u}$ satisfies the ODE

$$-sd_\xi\overline{u} + d_\xi f(\overline{u}) = -d_\xi(H(\overline{u})d_\xi p^c).$$

We can integrate this equation once, and use far-field conditions at infinity: $\overline{u}(\pm\infty) = u_\pm$:

$$-s(\overline{u}-u_-)+f(\overline{u})-f(u_-) = -H(\overline{u})d_\xi p^c, \text{ and } s = (f(u_+)-f(u_-))/(u_+-u_-), \tag{10}$$

assuming $d_\xi p^c \to 0$ at infinity.

If we neglect capillary pressure by setting $p^c \equiv 0$, then equation (8) is a scalar conservation law

$$u_t + f(u)_x = 0, \tag{11}$$

with a flux $f(u)$ that is typically S-shaped (if $m > 1, n > 1$ in (7)), hence concave-convex. Shock wave solutions

$$u(x,t) = \overline{u}(\xi) = \begin{cases} u_-, & x < st \\ u_+, & x > st \end{cases} \tag{12}$$

with $s = (f(u_+) - f(u_-))/(u_+ - u_-)$ are discontinuous traveling wave solutions of (11).

We are interested in understanding when shock waves correspond to traveling wave solutions of the enhanced equation (8). As indicated in the introduction, this depends on the dependence of p^c on u and it's derivatives. In the simplest non-trivial case, $p^c = p^e(u)$ is a strictly decreasing function of $u : g(u) = \frac{d}{du}p^e(u) < 0$. Then (10) is a first order ODE, and solutions must connect adjacent equilibria. In this case, we recover the (generalized) Lax entropy condition as a necessary and sufficient condition for a shock wave to have a corresponding traveling wave [7]:

$$f'(u_+) \le s \le f'(u_-). \tag{13}$$

Note that the boundary cases in which $s = f'(u_+)$ or $s = f'(u_-)$ are included, so that the shock is characteristic on one side. In fact a boundary case is included in the original construction of Buckley and Leverett [2], in which a solution of (11) includes a rarefaction wave connected to a shock wave to represent a water injection displacing oil, i.e., from $u = 1$ upstream to $u = 0$ ahead of the shock.

For the Gray-Hassanizadeh model, the situation is more complicated. We have $p^c = p^e(u) - \tau u_t$, so that the PDE reads

$$u_t + f(u)_x = -\left(H(u)g(u)u_x\right)_x + \tau\left(H(u)u_{xt}\right)_x. \tag{14}$$

Correspondingly, the ODE (10) becomes the second order equation

$$-s(\overline{u} - u_-) + f(\overline{u}) - f(u_-) = -H(\overline{u})g(\overline{u})\overline{u}' - s\tau H(\overline{u})\overline{u}'', \quad \text{where } ' = d_\xi. \tag{15}$$

We rewrite this equation as a first order system, change variables from ξ to $\eta = \xi/\sqrt{s\tau}$, and drop the bar from u. Thus, $u(\eta) = \overline{u}(\xi)$:

$$\begin{aligned} u' &= v \\ v' &= -\frac{1}{\sqrt{s\tau}}g(u)v + \frac{1}{H(u)}\left[s(u-u_-) - f(u) + f(u_-)\right]. \end{aligned} \tag{16}$$

We say a shock wave (12) is an *admissible* solution of (11) if system (16) has a solution (u, v) satisfying

$$(u, v)(\pm\infty) = u_\pm. \tag{17}$$

To describe which shocks are admissible, we need to specify the relative permeabilities (7) and the equilibrium pressure $p^e(u)$. For simplicity, w take $p^e(u) = -u$, so that $g(u) = -1$. There are also various positive parameters in the model. However, nondimensionalization greatly reduces the number of parameters. We introduce length and time scales L, T related to the total velocity V and a scale P for the pressure:

$$\frac{L}{T} = \frac{V}{\varphi}, \quad P = LV\frac{\varphi\mu^w}{K\kappa^w}.$$

Since there are free scales here, we take $\varphi = 1$ in what follows. With mobilities given by (2) and relative permeabilities given by (7),

$$f(u) = \frac{u^m}{u^m + M(1-u)^n}, \quad H(u) = \frac{u^m(1-u)^n}{u^m + M(1-u)^n}, \tag{18}$$

where $M = \frac{\kappa^o\mu^w}{\kappa^w\mu^o} < 1$ is the mobility ratio. Typically, the $\kappa^{w,o} \sim 1$, so this ratio is essentially the ratio of viscosities. When a specific value of M is needed for numerical simulations, we take $M = 0.2$.

From Rolle's Theorem we deduce that for $m, n > 1$, $f(u)$ has an inflection point at some $u_I \in (0,1)$. We shall assume that $f''(u)(u_I - u) > 0$ for $u \neq u_I$.

We need two other key values α, β of u, defined by

$$f'(\alpha) = (f(\alpha) - 1)/(\alpha - 1), \quad f'(\beta) = f(\beta)/\beta, \quad 0 < \alpha < \beta < 1.$$

3.1. **Equilibria of (16).**

For specified values of the parameters u_-, s, system (16) has between one and three equilibria, in which $v = 0$ and u satisfies $s(u - u_-) - f(u) + f(u_-) = 0$. Two equilibria coincide when $s = f'(u)$.

For $0 < u < 1$ there is $u^*(u) \in (\alpha, \beta)$ such that $f'(u^*) = (f(u) - f(u^*))/(u - u^*)$, and $u^*(u_I) = u_I$. From the convex-concave property of f, we calculate easily that $(u^*)'(u) < 0$, and in $(u^*)'(u_I) = -1/\sqrt{2}$. The inverse function $u_*(u)$ is also significant. In fact, for each pair (u_-, u_+) between the curves $u_+ = u^*(u_-), u_+ = u_*(u_-)$, there are three equilibria u_-, u_+, u_m, with u_m between u_- and u_+, and $s = (f(u_+) - f(u_-))/(u_+ - u_-)$. As indicated in Fig. 1, values of u_-, u_+ corresponding to shocks satisfying the Lax entropy condition are bounded by the diagonal and the curve $u_+ = u_*(u_-)$. These properties are spelt out in more detail in [11].

For (u_-, u_+) corresponding to Lax shocks, $(u, 0)$ is an unstable node or spiral and $(u_+, 0)$ is a saddle. However, there is no guarantee that there will be a trajectory from u_- to u_+ as the stable manifold entering $(u_+, 0)$ may be kept away from $(u_-, 0)$ by a third saddle point equilibrium whose unstable manifold surrounds $(u_-, 0)$.

If (u_-, u_+) is in either region labeled 3 in Fig. 1, then $(u_\pm, 0)$ are saddle points, and there is the possibility of a trajectory satisfying the boundary conditions (17). Such a saddle-saddle trajectory is not stable to perturbations (by changing u_+ for example), but has codimension one, so these trajectories can be found using a shooting method with a single parameter. We choose to fix $\tau > 0$ and u_-, and vary u_+.

When there is such a trajectory, we say the corresponding shock wave (12) is *undercompressive*, because it fails to satisfy the Lax entropy condition for "compressive" shocks. Nonetheless, undercompressive shocks are admissible, and in fact,

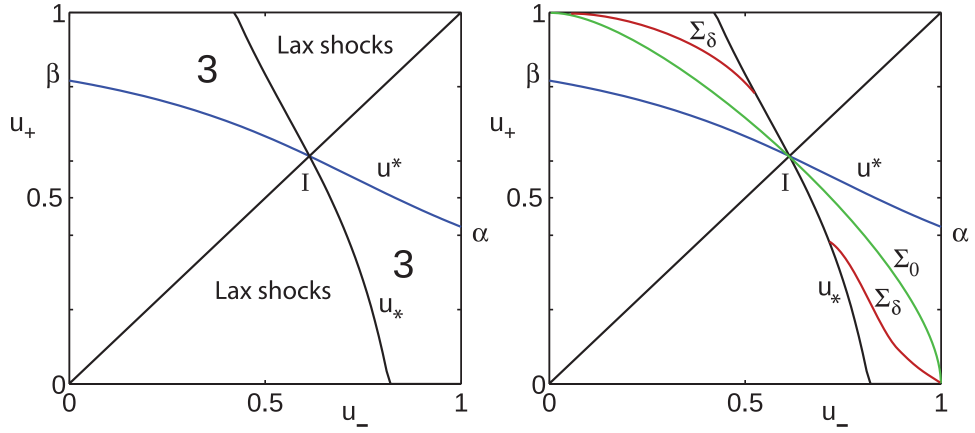


FIGURE 1. (Left) Values of $u_\pm$ (in region '3') for which there is a middle equilibrium $u = u_m, v = 0$ for (16), and values corresponding to Lax shocks. (Right) Σ curves recording saddle-saddle connections for $p = q = 2$.

they render some compressive shocks inadmissible, as indicated above. Saddle-saddle connections necessarily have $v = u'$ of a single sign, so that they can be parameterized by $u:\ v = v(u)$. Let $\gamma = \frac{1}{\sqrt{s\tau}}$. Then the ODE system can be written

$$v\frac{dv}{du} = \gamma v + \frac{s(u - u_-) - f(u) + f(u_-)}{H(u)}. \tag{19}$$

Consequently, for a saddle-to-saddle trajectory, since $v(u_\pm) = 0$,

$$\int_{u_-}^{u_+} G(y; u_-, u_+)\, dy = -\gamma \int_{u_-}^{u_+} v(y)\, dy, \tag{20}$$

where $G(y; u_-, u_+) = \{s(u-u_-)-f(u)+f(u_-)\}/H(u)$, and $s = (f(u_+)-f(u_-))/(u_+ - u_-)$.

Let $(u, v_+(u))$ be the locus of the stable manifold entering $(u_+, 0)$, on the side of $(u_-, 0)$, and let $(u, v_-(u))$ denote points on the unstable manifold leaving $(u_-, 0)$ towards $(u_+, 0)$. We can form a useful separation function by evaluating $v_\pm$ at the middle equilibrium u_m. Since γ depends on the shock speed, and hence on u_-, u_+, we remove this dependence by setting $\delta = \frac{1}{\sqrt{\tau}}$. Now define the separation function R by

$$R(u_-, u_+, \delta) = v_+(u_m) - v_-(u_m). \tag{21}$$

It takes a bit of checking that $v_\pm$ are both defined at u_m, but this is a simple analysis of the phase portraits. Then for (u_-, u_+) in region 3, there is a traveling wave from u_- to u_+ precisely when

$$R(u_-, u_+, \delta) = 0. \tag{22}$$

Such a simple separation function has been used successfully for numerical calculations in many contexts [6], although a more useful version allows for greater analytic simplicity [12, 10]. For each fixed δ, equation (22) defines a curve Σ_δ in each of the two region 3's of the (u_-, u_+) diagram.

Let's first consider the case $\tau = \infty$, i.e., $\delta = 0$. Then the existence of a trajectory joining saddle points at $u_\pm$ reduces to the equation

$$\int_{u_-}^{u_+} G(y; u_-, u_+)\, dy = 0. \tag{23}$$

For $m = n = 2$ we showed in [12] that there is a curve in the u_-, u_+ diagram through the inflection point I and ending in the corners $(0,1), (1,0)$. In fact, the behavior near the corners can be quantified in terms of the mobility ratio M. The asymptotics are somewhat delicate as $G(y; u_-, u_+)$ is singular at $y = 0$ and at $y = 1$. However, the logarithmic singularities in the integrals cancel and the curve Σ_0 of values satisfying (23) is well defined as described. For $\delta > 0$, the curves Σ_δ fill up portions of each region 3 between Σ_0 on the side of the boundaries $u_+ = 0, 1$, as shown in Fig 1 (right).

If either exponent m or n in the relative permeability functions is less than 2, then the picture changes. In this case, the function G is integrable at $u = 0, 1$, with the consequence that the Σ_0 curve reaches the boundaries $u_+ = 1, u_- = 1$ away from the corners of the u_-, u_+ square. This is shown in Fig. 2. As pointed out by van Duijn et al [14], if τ is sufficiently large, there may be no traveling waves corresponding to undercompressive shocks if (u_-, u_+) is close to one of the corners $(0,1), (1,0)$. Instead, there are *sharp shocks,* identified in [14, 3].

To explain this, we focus on a particular case, in which $m = n = \frac{3}{2}$. Then, setting $u_- = 0, u_+ = 1$, we have $s = 1$, and

$$G(y, 0, 1) = \frac{y - f(y)}{H(y)} = (1-y)^{\frac{3}{2}} y^{\frac{3}{2}} \{y^{\frac{5}{2}} - y^{\frac{3}{2}} + My(1-y)^{\frac{3}{2}}\}.$$

Thus, for $M = 0$, $\int_0^1 G(y, 0, 1) dy = 2$, and $\frac{d}{dM} \int_0^1 G(y, 0, 1) dy = 2$. Consequently, $\int_0^1 G(y, 0, 1) dy > 0$ for all $M > 0$. By continuity, for any fixed $M > 0$, for u_- near $u_- = 0$, and u_+ near $u_+ = 1$, $\int_{u_-}^{u_+} G(y, u_-, u_+) dy > 0$. However, if we keep u_- fixed near zero and decrease u_+, it is easy to see that the positive contribution to $\int_{u_-}^{u_+} G(y, u_-, u_+) dy$ increases while the negative part decreases. Consequently $\int_{u_-}^{u_+} G(y, u_-, u_+) dy$ can never reach zero and there are no traveling waves from u_- near zero to $u_+ > u_*(u_-)$, for $\delta = 0$. Moreover, since the Σ_δ curves lie above the Σ_0 curve in the u_-, u_+ diagram, for $u_- < u_I$, there are no traveling wave solutions for any value of $\delta > 0$ for this same range of u_-. A corresponding argument for (u_-, u_+) near $(1, 0)$ helps to justify the intersection of Σ_δ curves with the line $u_- = 1$ shown in Fig. 2.

In [14, 3] traveling waves are found numerically in this forbidden range that have a corner at $u_- = 1$, at a finite value of ξ, and then run smoothly to a small value of u_+. The speed of these waves is $s = (f(u_+) - 1)/(u_+ - 1)$, which corresponds to a shock from $u_- = 1$ to u_+. However, although $(u, v) = (1, 0)$ is an equilibrium for this value of s, the stable manifold from $(u_+, 0)$ intersects the line $u = 1$ at a point $v = u' < 0$, as shown in Fig. 2 (Right). This is related to the separation function not changing sign for $u_+ \sim 0$ and $u_- \sim 1$ as explained above. Consequently, there is a traveling wave as predicted with a finite negative slope at $u = 1$. An unexplained aspect of these observations is how the speed s is selected, since the same behavior of the stable manifold is present in the phase portrait for any nearby value of s.

To verify the occurrence of sharp shocks in the PDE, we ran numerical PDE simulations for the full Gray-Hassanizadeh PDE, in which initial data are chosen

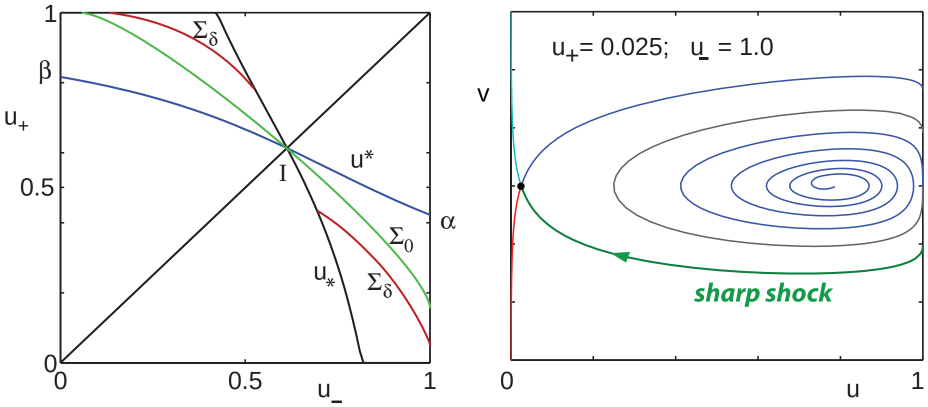


FIGURE 2. Left: Sigma curves showing undercompressive shocks for $p = q = 3/2$. Right: Phase portrait for a sharp shock TW

corresponding to a jump from $u_- < 1$ but near 1, to u_+ near zero, specifically a smoothed jump from $u = 0.9$ to $u = 0.025$:

$$u(x,0) = \frac{1}{2}(0.925 - 0.875\tanh(10x)) \tag{24}$$

The resulting solutions are shown at a series of times in Fig. 3. After an initial transient, the solution quickly finds the level $u = 1$, and evolves into a combination of traveling waves. The slower wave corresponds to a Lax shock from $u = 0.9$ to $u = 1$, with oscillations corresponding to the spiral equilibrium at $u = 0.9$. This is preceded by a faster sharp shock, with monotonic traveling wave that has a corner at $u = 1$, and approaches $u = 0.025$ exponentially, as suggested by the phase portrait in Fig. 2(left).

4. **Stability of Plane Waves.** The shock waves (12) of the previous section are plane wave solutions of system (5) in which $p^c = 0$. This reduced system is hyperbolic-elliptic:

$$\begin{aligned} \frac{\partial u}{\partial t} - \nabla\cdot(\lambda(u)\nabla p) &= 0, \\ \nabla\cdot(\Lambda(u)\nabla p) &= 0. \end{aligned} \tag{25}$$

In this section, I discuss linearized stability of planar shocks to two-dimensional perturbations.

A shock wave solution of (25) is a weak solution $(u(x,y,t), p(x,y,t))$ of the system in which u has a discontinuity along a curve $x = \hat{x}(y,t)$, p is continuous there, but ∇p is discontinuous. The normal to the discontinuity surface $x = \hat{x}(y,t)$ in t, x, y is $(-\hat{x}_t, 1, -\hat{x}_y)$. With the notation $u_\pm(y,t) = u(\hat{x}(y,t)^\pm, y, t)$, and similarly for p_x, p_y, we have jump conditions:

$$\begin{aligned} -\hat{x}_t[u] - [\lambda(u)p_x] + \hat{x}_y[\lambda(u)p_y] &= 0, \\ -[\Lambda(u)p_x] + \hat{x}_y[\Lambda(u)p_y] &= 0, \end{aligned} \tag{26}$$

where $[\cdots]$ denotes the jump.

Consider a planar Lax shock related to the piecewise constant saturation function (12), in which the shock is located for fixed t on the line $x = st$, with velocities

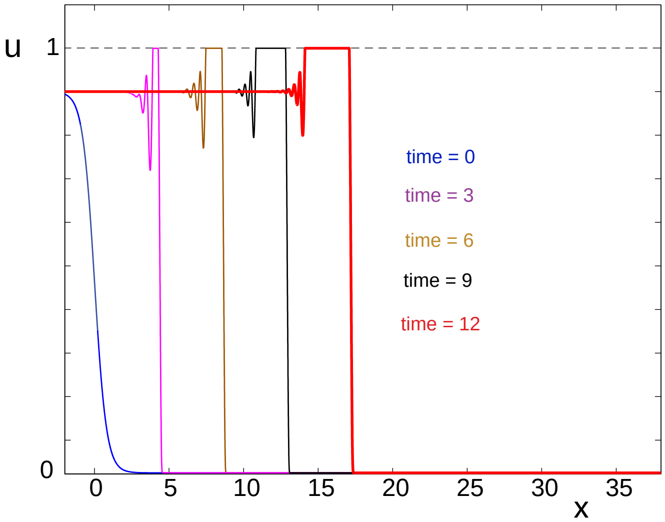


FIGURE 3. PDE simulations of equation (14) showing a sharp shock emerging from monotonic initial data.

$v_\pm = -\lambda_\pm \partial_x p_\pm$ constant on either side of the shock. The mobilities $\lambda_\pm = Kk(\bar{u}_\pm)/\mu$ depend on the constant saturations $u = \bar{u}_\pm$. The pressure p is continuous; thus (up to a constant):

$$p_\pm = \bar{p}_\pm \cdot (x - st) = -\frac{v_\pm}{\lambda_\pm} z = -\frac{V}{\Lambda_\pm} z, \quad z = x - st.$$

It is convenient to work in the frame moving with speed s, for which the shock location is $z = \bar{z} = \bar{x} - st = 0$.

We seek solutions of the linearized system. To this end, we perturb the variables u, p and the shock location.

$$u = \bar{u}_\pm + U_\pm(x,y,t), \quad p = \bar{p}_\pm \cdot (x - st) + P_\pm(x,y,t), \quad x = st + \hat{z}(y,t).$$

Equations (25) are linearized about the shock on each side, giving linearized equations (dropping the subscripts ±):

$$\begin{aligned} U_t - \lambda'(\bar{u})\bar{p}\,U_x - \lambda(\bar{u})\Delta P &= 0, \\ \Lambda'(\bar{u})\bar{p}\,U_x + \Lambda(\bar{u})\Delta P &= 0. \end{aligned} \tag{27}$$

Thus, $\Delta P = -\dfrac{\Lambda'(\bar{u})}{\Lambda(\bar{u})}\bar{p}\,U_x$. Substituting back into the first equation, we eliminate $P:$

$$U_t - \bar{p}\left(\lambda'(\bar{u}) - \lambda(\bar{u})\frac{\Lambda'(\bar{u})}{\Lambda(\bar{u})}\right) U_x = 0.$$

This is a linear equation for $U = U_\pm$, with constant coefficients. Moreover, although dependence on y has been eliminated, we need to consider perturbations that are dependent on y as they can affect P. Solutions have the form $U(x,y,t) = w(x - st)e^{i\alpha y}e^{\sigma t}$. Here, $\alpha \in \mathbb{R}$ is the wave number, $\sigma \in \mathbb{C}$ gives the time response: $Re\,\sigma > 0$

corresponds to an unstable wave, and $Re\,\sigma < 0$ is needed for stability. The traveling wave w is required to be bounded. With $z = x - st, w' = \frac{dw}{dz}$, we obtain

$$\sigma w - sw' - \bar{p}\,\lambda(\bar{u})\left(\frac{\lambda'(\bar{u})}{\lambda(\bar{u})} - \frac{\Lambda'(\bar{u})}{\Lambda(\bar{u})}\right) w' = 0.$$

But $f(u) = V\lambda(u)/\Lambda(u), \quad \bar{p}\,\lambda(\bar{u}) = -f(\bar{u})$ so after some calculation we find

$$\sigma w = (s - f'(\bar{u}))\, w'.$$

This ODE has solutions $w(z) = a_\pm e^{\beta_\pm z}$, with

$$\sigma = \beta_\pm(s - f'(\bar{u}_\pm)).$$

The Lax entropy condition reads $f'(\bar{u}_+) < s < f'(\bar{u}_-)$. Consequently, for a Lax shock, $\pm Re\ \beta_\pm > 0$, if $Re\ \sigma > 0$. Therefore, $w(z)$ does not decay at $z = \pm\infty$ unless $w \equiv 0$. If $Re\ \sigma < 0$, then the perturbation decays in time, at least to leading order.

We proceed to consider perturbations with $U \equiv 0$. We still have the equation for the perturbation $P = P_\pm$ of the base pressure. We seek solutions of the form $u = \bar{u}_\pm,\ p = \bar{p}_\pm z + P_\pm,\ P_\pm(z, y, t) = q_\pm(z)e^{i\alpha y + \sigma t}$, with sharp interface at $\hat{z} = \hat{x} - st = ae^{i\alpha y + \sigma t},\ z = x - st$. By scale invariance, we have $\sigma = \sigma_1\alpha,\ \alpha > 0$, and we wish to determine the sign of σ_1 when there are non-trivial solutions of the linearized equations.

Since $U \equiv 0$, equations (27) reduce to $\Delta P = 0$, leading to

$$q''_\pm - \alpha^2 q_\pm = 0. \tag{28}$$

Thus, $q_\pm = b_\pm e^{\mp\alpha z},\ \pm z > 0$, are the solutions that decay as $|z| \to \infty$. It remains to write three linear equations for the three constants $a, b_\pm$. Then there are non-trivial solutions if and only if the determinant of coefficients is singular, thus leading to an expression for σ_1, which appears in the coefficiant matrix. The analysis is an explicit version of the derivation of the Lopatinski determinant condition for stability of a shock wave solution of a system of hyperbolic conservation laws. (See [1], for example.) Two of the three conditions come from linearizing the jump conditions (26), and the third condition ensures that the pressure $p(x, y, t)$ is continuous.

The linearized jump conditions, after setting $U \equiv 0$, become

$$\hat{z}_t[\bar{u}] + [\lambda(\bar{u})P_x] \quad = \quad 0 \tag{29}$$

$$[\Lambda(\bar{u})P_x] \quad = \quad 0. \tag{30}$$

Substituting in the forms for $\hat{z}$ and $P_\pm$, we get

$$\sigma_1\alpha a[\bar{u}] + [\lambda(\bar{u})q'] \quad = \quad 0 \tag{31}$$

$$[\Lambda(\bar{u})q'] \quad = \quad 0 \tag{32}$$

Equation (32) relates $b_\pm$:

$$\Lambda(\bar{u}_+)b_+ = -\Lambda(\bar{u}_-)b_-, \tag{33}$$

from which (31) implies

$$b_-\Lambda(\bar{u}_-)(f(\bar{u}_+) - f(\bar{u}_-)) = -\sigma_1\, a\,(\bar{u}_+ - \bar{u}_-)V.$$

But $(f(\bar{u}_+) - f(\bar{u}_-))/(\bar{u}_+ - \bar{u}_-) = s$, the shock speed, so we get a second equation,

$$b_-\Lambda(\bar{u}_-)s = -a\,\sigma_1\, V. \tag{34}$$

The third equation comes from continuity of pressure at $z = \hat{z}(y,t) = ae^{i\alpha y+\sigma t}$, recalling that $p = \bar{p}_\pm z + q_\pm(z)e^{i\alpha y+\sigma t}, q_\pm = b_\pm e^{\mp\alpha z}\ (\pm z > 0):$

$$\bar{p}_+ a + b_+ = \bar{p}_- a + b_- \tag{35}$$

Thus, $(\bar{p}_+ - \bar{p}_-)a = -\sigma_1 \frac{V}{s} a \left(\frac{1}{\Lambda(\bar{u}_+)} + \frac{1}{\Lambda(\bar{u}_-)}\right)$, from (33), (34). Since $\bar{p}_\pm = -\dfrac{V}{\Lambda(\bar{u}_\pm)}$, we obtain, for $a \neq 0:$

$$\sigma_1 = s\frac{\Lambda(\bar{u}_-) - \Lambda(\bar{u}_+)}{\Lambda(\bar{u}_-) + \Lambda(\bar{u}_+)}.$$

Since $s > 0$, and $\Lambda(\overline{u}_\pm) > 0$, we conclude that the planar shock (12) is linearly stable to transverse perturbations if $\Lambda(\bar{u}_-) < \Lambda(\bar{u}_+)$, and is linearly unstable if $\Lambda(\bar{u}_-) > \Lambda(\bar{u}_+)$. Thus, the curve

$$\Lambda(\bar{u}_-) = \Lambda(\bar{u}_+) \tag{36}$$

separates regions in the u_-, u_+ plane where the shock is stable from regions where the shock is unstable. Moreover, although the analysis used the Lax entropy condition, which also defines regions within the u_-, u_+ plane, the condition for instability is valid for shocks that fail to satisfy the Lax condition, since the eigenfunctions with $U \equiv 0$ are still present.

With quadratic relative permeabilities, the stability boundary (36) simplifies to a straight line $u_+ = -u_- + \frac{2M}{M+1}$, which necessarily cuts through the region of Lax shocks. Thus, some Lax shocks are stable and others are unstable. Somewhat surprisingly, there are arbitrarily weak Lax shocks that are stable and others that are unstable, since the stability boundary crosses the diagonal $u_- = u_+$ transversally. The stability boundary was tested numerically for a pair of cases in [13]. More detailed simulations have been performed elsewhere [8], but based on the detailed stability analysis of traveling waves of Yortsos and Hickernell [15] rather than the comparatively simple analysis here.

A further consequence in the case of quadratic relative permeabilities is that undercompressive shocks are necessarily unstable. It would be interesting to know whether there are relative permeability functions for which at least some undercompressive shocks are stable. This is more complicated than it might appear however, since for undercompressive shocks there can be eigenfunctions with U non-zero on the left of the shock (the side on which characteristics leave the shock). Then the generalization of the Lopatinski determinant condition involves quantifying the kinetic relation, an additional condition that determines which shocks are admissible. In the previous section, this was $u_+ = \hat{u}_+(u_-)$, and was derived numerically from computations of traveling waves.

Acknowledgments. The author thanks Kim Spayd and Zhengzheng Hu for their contributions to this paper. Discussion with John Hunter clarified the stability argument of the last section.

REFERENCES

[1] S. Benzoni-Gavage and H. Freistühler, Effects of Surface Tension on the Stability of Dynamical Liquid-Vapor Interfaces, *Arch. Rational Mech. Anal.* **174,** 111–150, 2004.

[2] S. E. Buckley and M. C. Leverett. Mechanism of fluid displacement in sands. *Petroleum Trans. AIME*, **146,** 107–116, 1942.

[3] Y. Fan, Dynamic Capillarity in Porous Media – Mathematical Analysis. Ph.D. Thesis, Eindhoven University, 2012.

[4] S. M. Hassanizadeh and W. G. Gray. Mechanics and thermodynamics of multiphase flow in porous media including interphase boundaries. *Adv. Water Resources*, **13,** 169–186, 1990.

[5] S. M. Hassanizadeh and W. G. Gray. Thermodynamic basis of capillary pressure in porous media. *Water Resources Research*, **29,** 3389–3405, 1993.

[6] B. Hayes and M. Shearer, Undercompressive shocks and Riemann problems for scalar conservation laws with nonconvex fluxes. *Proc. Royal Society Edinburgh,* **129A**, 733–754, 1999.

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

[8] A. Riaz and H. A. Tchelepi. Numerical simulation of immiscible two-phase flow in porous media. *Physics of Fluids*, **19**: 072103, 2007.

[9] P. G. Saffman and G. I. Taylor. The penetration of a fluid into a porous medium or hele-shaw cell containing a more viscous liquid. *Proc. Royal Society London Series A*, 245:312–329, 1958.

[10] M. Shearer and Y. Yang, The Riemann problem for a system of conservation laws of mixed type with a cubic nonlinearity. *Proc. Royal Society Edinburgh,* **125A,** 675-699, 1995.

[11] K. R. Spayd, Two Phase Flow in Porous Media: Traveling Waves and Stability Analysis. Ph.D. Thesis, North Carolina State University, 2012.

[12] K.R. Spayd and M. Shearer, The Buckley–Leverett equation with dynamic capillary pressure, *SIAM J. Appl. Math.* **71,** 1088–1108, 2012.

[13] K.R. Spayd, M. Shearer and Z. Hu, Stability of plane waves in two-phase porous media flow, *Applicable Analysis* **91**, 295-308, 2012.

[14] C.J. van Duijn, Y. Fan, L.A. Peletier and I.S. Pop, Travelling wave solutions for degenerate pseudo-parabolic equation modelling two-phase flow in porous media, *Nonlinear Analysis: Real World Appl.* **14(3)**,1361–1383, 2013

[15] Y. C. Yortsos and F. J. Hickernell. Linear stability of immiscible displacement in porous media. *SIAM J. Appl. Math.*, **49**, 730–748, 1989.

E-mail address: `shearer@ncsu.edu`

Part 3

Contributed Talks

NUMERICAL SIMULATION OF THREE-PHASE FLOW IN HETEROGENEOUS MEDIA WITH SPATIALLY VARYING NONLINEAR HYPERBOLIC-PARABOLIC FLUX FUNCTIONS

Eduardo Abreu

Department of Applied Mathematics
IMECC, University of Campinas (UNICAMP)
Campinas, SP 13083-859, Brazil

Abstract. We address the issue of numerical simulation of a three-phase flow model in a two-dimensional heterogeneous porous media with an isolated umbilic point inside the three-phase domain and taking into account gravity effects and explicit spatially varying flux functions in the both hyperbolic and parabolic operators in the differential three-phase governing equations. Our computational approach is based on an operator-splitting procedure for decoupling the nonlinear three-phase flow system with mixed discretization methods, leading to purely hyperbolic, parabolic and elliptic subproblems. We were able to numerically reproduce semi-analytical nonclassical results for three-phase flow calculations for one-dimensional homogeneous media. Several numerical experiments were performed in order to show evidence of stable nonclassical waves for the three-phase flow system with gravity effects at hand based on flux functions with explicit spatial variation.

1. **Introduction.** We discuss a computational approach for solving and simulation of a immiscible and incompressible three-phase flow model which in turn includes the gravity effects in inhomogeneous porous medium in two space dimensions taking into account hyperbolic-parabolic flux functions with explicit spatial variation. We consider a fluid composed of phases gas, oil and water, mixed at macroscopic level, and indicated by the subscripts g, o, and w. Furthermore, there are no sources or sinks and compressibility and thermal effects are considered to be negligible. We also assume that the whole pore space is occupied by the fluid phases. As a consequence [11, 4], any pair of saturations inside the saturation triangle (three-phase domain) defined by $\triangle := \{(S_w, S_g) \,|\, 0 \leq S_i \leq 1; S_w + S_g + S_o = 1\}$, can be chosen to describe the state of the fluid phases (S_o, S_w, S_g) in the porous medium. Therefore, following [2, 1], we choose S_w, S_g and p_o as the three independent variables of the differential problem. Thus, the system of conservation laws governing immiscible

2000 *Mathematics Subject Classification.* Primary: 76T30; Secondary: 68U20.

Key words and phrases. Three-phase flow, Nonclassical wave, Gravity, nonlinear convection-diffusion system, mixed hybrid finite element, operator splitting.

E. Abreu acknowledge financial support from State of São Paulo Research Foundation grants 2011/11897-6 and 2012/19874-8 - FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) and from grant 519.292-785/11 by Fundo de Apoio ao Ensino, Pesquisa e Extensão da Universidade Estadual de Campinas (UNICAMP/FAEPEX).

and incompressible three-phase flow in porous media reads,

$$\begin{aligned}
&\frac{\partial}{\partial t}(\phi(\mathbf{x})S_w) + \nabla\cdot(\mathbf{F}_w(\mathbf{S},\mathbf{x})) = \nabla\cdot(\mathbf{D}_w(\mathbf{S},\mathbf{x})),\\
&\frac{\partial}{\partial t}(\phi(\mathbf{x})S_g) + \nabla\cdot(\mathbf{F}_g(\mathbf{S},\mathbf{x})) = \nabla\cdot(\mathbf{D}_g(\mathbf{S},\mathbf{x})),\\
&\nabla\cdot\mathbf{v} = 0, \quad \mathbf{v} = -K(\mathbf{x})\lambda(\mathbf{S})\nabla p_o + \epsilon_{P}(\mathbf{v}_{wo}+\mathbf{v}_{go}) + \epsilon_{G}\mathbf{v}_G,
\end{aligned} \tag{1}$$

where $\mathbf{S} = (S_w, S_g)$ and the spatially varying hyperbolic terms are denoted by,

$$\begin{aligned}
\mathbf{F}_w(\mathbf{S},\mathbf{x}) &\equiv \mathbf{v} f_w(\mathbf{S},\mathbf{x}) + \epsilon_{G} K(\mathbf{x})[\lambda_w(1-f_w)\rho_{wo} - \lambda_w f_g\rho_{go}]\,\mathrm{g}\nabla Z,\\
\mathbf{F}_g(\mathbf{S},\mathbf{x}) &\equiv \mathbf{v} f_g(\mathbf{S},\mathbf{x}) + \epsilon_{G} K(\mathbf{x})[\lambda_g(1-f_g)\rho_{go} - \lambda_g f_w\rho_{wo}]\,\mathrm{g}\nabla Z.
\end{aligned} \tag{2}$$

The diffusive terms of the three-phase model are represented by the right-hand side of the system (1) incorporating capillary pressure effects (that in principle are experimentally measured functions of all saturations), which in turn are given by,

$$[\mathbf{D}_w, \mathbf{D}_g]^\top = \epsilon_{P} G(\mathbf{S},\mathbf{x})\,[\nabla S_w, \nabla S_g]^\top. \tag{3}$$

Matrix $G(\mathbf{S},\mathbf{x})$ can be rewritten conveniently into the product of the uniformly positive scalar permeability field $K(\mathbf{x})$ along with the following two 2×2 matrices,

$$\Theta = \begin{bmatrix} \lambda_w(1-f_w) & -\lambda_w f_g \\ -\lambda_g f_w & \lambda_g(1-f_g) \end{bmatrix} \text{ and } \Psi = \begin{bmatrix} \partial p_{ow}/\partial S_w & \partial p_{ow}/\partial S_g \\ \partial p_{og}/\partial S_w & \partial p_{og}/\partial S_g \end{bmatrix}. \tag{4}$$

Since experimental data are hardly available we use physical arguments of Aziz and Settari (see, e.g., [11, 2, 1]) that the capillary pressure functions for three-phase flow can be taken to have the form $p_{wo} = -P_{ow}(S_w)$ and $p_{go} = -P_{og}(S_g)$, where P_{ow} and P_{og} are certain monotone decreasing functions (as described below). Thus in this approach, three-phase capillary pressures can be expressed in terms of well-known two-phase capillary pressures [11, 2, 1]. So, assumptions on capillary pressures leads to conditions on $G(\mathbf{S},\mathbf{x})$. In short, for our choice of P_{ow} and P_{og} we have an operator $\nabla\cdot[\mathbf{D}_i(\mathbf{S},\mathbf{x})]$, $i = w, g$, associated with (1) that is strictly parabolic in the interior of the saturation triangle $\triangle$ and system (1)-(4) is well-posed in the three-phase domain [11]. To take the qualitative behavior of the effect of heterogeneity (imposed by the permeability $K(\mathbf{x})$ and the porosity $\phi(\mathbf{x})$) we use a Leverett J-function to scale the capillary pressure curve for each grid block in the porous media (see [1] for details): $p_{io}(S_w, S_g, S_o) = \frac{-P_{oi}}{\sigma_{io}\ cos\,\psi_{io}}\sqrt{\frac{K(\mathbf{x})}{\phi(\mathbf{x})}}$, $i = w,\ g$, where σ_{io} is the interfacial tension and ψ_{io} is the contact angle, the wetting preference of a solid surface in contact with two fluids, along with the following models $P_{ow} = 5\left(S_w^{-2} - (1-S_w)^{-2}\right)$ and $P_{og} = \left(S_g^{-2} - (1-S_g)^{-2}\right)$. For the three-phase flow system (1)-(4), $K(\mathbf{x}) > 0$ and $\phi(\mathbf{x}) > 0$ are the absolute permeability and porosity of the porous medium; K_i, ρ_i and μ_i are, respectively, the relative permeability, density and viscosity of phase i; p_o is the oil pressure; the correction velocities $\mathbf{v}_{wo}$ and $\mathbf{v}_{go}$ are defined by $\mathbf{v}_{ij} = -K(\mathbf{x})\lambda_i(S_w,S_g)\nabla p_{ij}$, $i \neq j$ and $\mathbf{v}_G(S_w,S_g) = K(\mathbf{x})\lambda_\rho\,\mathrm{g}\,\nabla Z$ is the velocity correction due to gravity, where $\lambda_\rho(S_w,S_g) \equiv \lambda_w\rho_w + \lambda_o\rho_o + \lambda_g\rho_g$; ρ_i is the density of phase i. The magnitude of the gravity is g and Z is the depth. In the above definitions $\lambda_i = K_i/\mu_i$ denotes the mobility of phase i, the ratio between relative permeability and viscosity; $\lambda(S_w,S_g) = \sum_i \lambda_i$ is the total mobility and $f_i(S_w,S_g) = \lambda_i/\lambda$ is the fractional flow function of phase i. We use the model by Corey [11] for relative permeabilities: $K_w = S_w^2$, $K_o = S_o^2$ and $K_g = S_g^2$. For other models of three-phase flow used in petroleum engineering, such as certain models of Stone [11], the umbilic point in the Corey model is generally replaced by an elliptic region, which in turn the characteristic speeds are not real. After nondimensionalizing the three-phase equations (1)-(4) two dimensionless groups appears: $\epsilon_{_G}$ quantifies the ratio between the gravity effects to convective effects and $\epsilon_{_P}$ quantifies the ratio between convective to diffusive effects. In the numerical experiments, to be reported later on, we will specify the corresponding values of the dimensionless groups.

2. **The Numerical Method.** We limit ourselves to a very short description of the numerical method. The interested reader is referred to [1, 2] for further information. Following [2], our computational method is an operator-splitting procedure [10, 8, 6] for decoupling the nonlinear three-phase flow system (1)-(4). The splitting allows time steps for the pressure-velocity (elliptic subproblem) calculation,

$$\nabla \cdot \mathbf{v} = 0, \quad \mathbf{v} = -K(\mathbf{x})\lambda(\mathbf{S})\nabla p_o + \epsilon_{_P}(\mathbf{v}_{wo} + \mathbf{v}_{go}) + \epsilon_{_G}\mathbf{v}_G, \tag{5}$$

that are longer than those for the purely diffusive calculation,

$$\frac{\partial}{\partial t}(\phi(\mathbf{x})S_w) = \nabla \cdot (\mathbf{D}_w(\mathbf{S},\mathbf{x})), \qquad \frac{\partial}{\partial t}(\phi(\mathbf{x})S_g) = \nabla \cdot (\mathbf{D}_g(\mathbf{S},\mathbf{x})), \tag{6}$$

which in turn are longer than those for purely convection calculation,

$$\frac{\partial}{\partial t}(\phi(\mathbf{x})S_w) + \nabla \cdot (\mathbf{F}_w(\mathbf{S},\mathbf{x})) = 0, \qquad \frac{\partial}{\partial t}(\phi(\mathbf{x})S_g) + \nabla \cdot (\mathbf{F}_g(\mathbf{S},\mathbf{x})) = 0. \tag{7}$$

Of course, appropriate boundary and initial conditions [1, 2] must be specified to solve system (1)-(4), or (5)-(7). A mixture of water and gas is injected at a uniform and constant rate along the left boundary $(x, y) \in \{0\} \times [0, Y]$ for temporally constant boundary conditions with production of a three-phase fluid mixture at $(x, y) \in \{X\} \times [0, Y]$. No flow is allowed across the boundaries with $(x, y) \in [0, X] \times \{0, Y\}$. In order to capture viscous fingering effects keep a constant reference pressure p_o at $(x, y) \in \{X\} \times [0, Y]$. Therefore, we have a solvable elliptic problem [1, 2]. The oil pressure p_o and the Darcy velocity $\mathbf{v}$ in (5) are approximated at times $t^m = m\Delta t_p$, $m = 0, 1, 2, \ldots$. Locally conservative mixed finite elements [1, 2] are used to discretize the pertinent elliptic equation (5) and the spatial operators in the diffusion system (6). In the diffusion step, the saturations S_w and S_g are approximated at times $t_n = n\Delta t_d$, $n = 1, 2, \ldots$; the time discretization of the latter is performed by means of the implicit backward Euler method [1, 2]. In practice, the convective transport (7) might be computed at intermediate times $t_{n,\kappa} = t_n + k\Delta t_c$ for $t_n < t_{n,\kappa} \leq t_{n+1}$, which in turn the time steps Δt_c are subject dynamically to a CFL condition [1]. Indeed, our main goal in the current study is the accurate numerical simulation of the three-phase system (1)-(4) in discontinuous porous. Therefore, in order to avoid an overestimation of the shock layer, we take Δt_c= Δt_d=Δt_p since the time step Δt_c should not be considerably larger than $O(\epsilon_P)$ (see [8, 1] for details). This choice is an attempt to perform an accurate computation of the nonlinear self-sharpening mechanism between convection and diffusion.

3. **A central scheme formulation for variable porosity and flux functions with spatial variation.** We discuss very briefly a possible implementation of a central differencing scheme for the numerical approximation of the hyperbolic system (7) taking into account heterogeneous permeability $K(\mathbf{x})$ and porosity $\phi(\mathbf{x})$ fields. First, we notice that Karlsen and Towers [9] gave a convergence proof for the Lax-Friedrichs finite difference scheme for non-convex genuinely nonlinear scalar conservation laws of the form $u_t + f(u, k(x, t))_x = 0$, with $u = u(x, t)$, where

the coefficient $k(x,t)$ is allowed to be discontinuous along curves in the (x,t) plane. Karlsen and Towers also proved stability, and uniqueness, for an extended Kruzhkov entropy solution, provided that the flux function satisfies a so-called crossing condition, and that strong traces of the solution exist along the curves where $k(x,t)$ is discontinuous. In this way they were able to show that a convergent subsequence of approximations produced by the Lax-Friedrichs scheme to the above equation converges to such an entropy solution [9]. On the other hand, we recall that the celebrated central scheme, introduced by Nessyahu and Tadmor [13], is also based on the Lax-Friedrichs scheme in a staggered grid. Nessyahu and Tadmor proved that the resulting scalar scheme satisfies both the Total Variation Diminishing property and a local cell entropy inequality in order to get convergence to the unique entropy solution, at least in the genuinely non-linear scalar case.

Thus, the idea here is to build a locally conservative central approximation scheme for system (7) respecting the local equilibria linked to the capillary pressure discontinuities associated with the parabolic system (6). Following Nessyahu-Tadmor [13] and Karlsen-Towers [9] we performed a consistent discretization of the flux functions $\mathbf{F}_i(\mathbf{S},\mathbf{x})$, $i=w,g$ for the hyperbolic system (7) in a staggered grid. This is the same locally conservative approach as performed by Karlsen-Towers [9] and Nessyahu-Tadmor [13]. For more details we refer the reader to [1, 2].

We remark that the spatial integration of the discontinuous flux functions is performed over the entire Riemann fan [13]. This is the distinctive feature of the central scheme approach. Indeed, this integration eliminates the need of any detailed knowledge about the exact (or approximate) Riemann problem. Furthermore, it facilitates more accurate computation of the numerical flux $\int_t^{t+\Delta t_c}\mathbf{F}_i(\mathbf{S},K(\mathbf{x}))d\tau$, whose values are extracted from the smooth interface of two noninteracting (local) Riemann problems taking into account discontinuous quantities $\mathbf{S}$, $\phi(\mathbf{x})$ and $K(\mathbf{x})$. In other words, such values share the benefit of the location of the "well-defined" integration points due to the staggered solution strategy to take into account hyperbolic-parabolic flux functions with explicit spatial variation on $K(\mathbf{x})$ and $\phi(\mathbf{x})$ associated with the system (1)-(4). To avoid the error incurred in the calculation of cell averages associated with $\partial(\phi(\mathbf{x})\mathbf{S})/\partial t$ in system (7), we use a projection by means of a locally conservative definition of the porosity $\phi(\mathbf{x})$ on the same staggered grid. This numerical approach seems to be appropriate to resolve the discontinuous coefficients $K(\mathbf{x})$ and $\phi(\mathbf{x})$ without spurious oscillation in the numerical numerical approximations to equation (5)-(7).

The nonlinear parabolic subsystem (6) associated to the system (1)-(4) is handled by a mixed finite element approach [2, 1]. In particular, it has been shown in [5] that discontinuous capillary pressure field yield undercompressive discontinuities, in opposition to what was commonly accepted in the literature [3, 12]. This effect seems to be of great importance on the behavior of the solution, even when the capillary pressure seemed to be neglected. In our formulation the discontinuous capillary pressure appears in the hyperbolic-parabolic fluxes (5)-(7). Furthermore, our numerical procedure combines a domain decomposition technique with an implicit time backward Euler method (see [2, 1]) in the construction of a local iterative method for system (6), which in turn allows for variable porosity $\phi(\mathbf{x})$.

In the above algorithm, we first assume that the saturations $\{S_w,S_g\}$ and their fluxes $\{\mathbf{D}_w,\mathbf{D}_g\}$ are known. Then, the phase velocity corrections $\mathbf{v}_{wo}$ and $\mathbf{v}_{go}$ can be calculated using $\{\mathbf{D}_w,\mathbf{D}_g\}$ [2, 1]. Finally, the total velocity $\mathbf{v}$ can be recovered from the elliptic subsystem (5).

4. **Numerical Experiments.** Several 1D and 2D numerical experiments are performed to exhibit the relevance of the simulation of three-phase flow in discontinuous porous media. For the one-directional injection problem depicted here, the spatial domain for (1)-(4) is an idealized semi-infinite core for $x > 0$. In this setting, we are interested in solutions for initial conditions at $t = 0$, corresponding to the Riemann data: left state equal to $(S_w, S_g) = (0.15, 0.05)$ and for temporally constant boundary conditions at $x = 0$, representing a steady injection rate with left state $(S_w, S_g) = (0.721, 0.279)$.

It is also worth mentioning that the presence of an umbilic point or an elliptic region almost always prohibits the existence of Riemann invariants. This fact is at the root of the complexity of solutions when strict hyperbolicity fails, as it does for the model of immiscible three-phase flow (1)-(4). Only viscosity ratios are relevant, but for the purpose to generate representative solutions related to nonclassical structure for three-phase flow model (5)-(7), we will choose the following dimensionless viscosities $\mu_w = 0.5$, $\mu_o = 1.0$, and $\mu_g = 0.3$. In conjunction with the capillary pressures functions, we take interfacial tension values $\sigma_{go} = 23$ and $\sigma_{wo} = 51$ and contact angle with limit values of $\psi_{wo} \approx 90^o$ and $\psi_{go} \approx 90^o$. For the experiments with gravity we consider: $\rho_o = 0.7$, $\rho_w = 1.0$, and $\rho_g = 5.76 \times 10^{-2}$.

The one-dimensional numerical results reported in Figure 1 were computed with a grid having 512 uniform cells. In these frames are shown saturation profiles for gas (top) and oil (bottom) as a function of distance. We compare simulations with (solid profiles) and without gravity (dashed profiles). For the numerical experiments we take $\epsilon_{_P} = 10^{-3}$ and two values for the dimensionless group $\epsilon_{_G}$ are used: 5.5×10^{-3} (top solid profiles) and 8.5×10^{-2} (bottom solid profiles).

Notice in the frames of Figure 1 **left column** computed with gravity that the oil saturation front (solid line from S^* to $2CS$, from $2CS$ to $1CS$ and from $1CS$ to SR) are slightly slower than the solution computed without gravity (dashed line). We point out that in the top picture of Figure 1 the wave front in the gas saturation profile computed with gravity (solid line) also move very slightly ahead when compared with the case without gravity. These results are in agreement with the action of gravity on the flow, as it is in the opposite direction with respect to the injection. In addition, we remark that a stable nonclassical structure is simulated in the numerical results with gravity for a small value of $\epsilon_{_G}$ (see Figure 1).

We now turn to the discussion of the numerical experiments displayed on the **right column** of Figure 1 with $\epsilon_{_G} = 8.5 \times 10^{-2}$. The relative importance of gravity in this simulation is stronger. We point out that the oil saturation fronts move considerably slower (bottom-solid) and while the gas saturation profile moves faster (top-solid). In addition, the separate identity of the stable nonclassical wave from $1CS$ to $2CS$ and the Buckley-Leverett front from SR to $1CS$ seems to be lost (dashed curves in Figure 1). The findings in this numerical experiment address the issue of the effect of gravity on three-phase flow with structurally stable wave group solutions. Similar results were obtained for other capillary pressure functions [1].

4.1. **Two dimensional numerical computations.** The current study looks to the question of simultaneous propagation of multiple stable waves in discontinuous porous media using numerical simulation for system (5)-(7). We point out that the propagation of nonclassical undercompressive waves have a strong dependency upon the parabolic-diffusion mechanism being modeled [7, 11, 2]; such waves have been observed only in one-dimensional for three-phase flow problems (see, e.g., [7, 14, 11, 4]). The understanding of the interplay between rock heterogeneity and

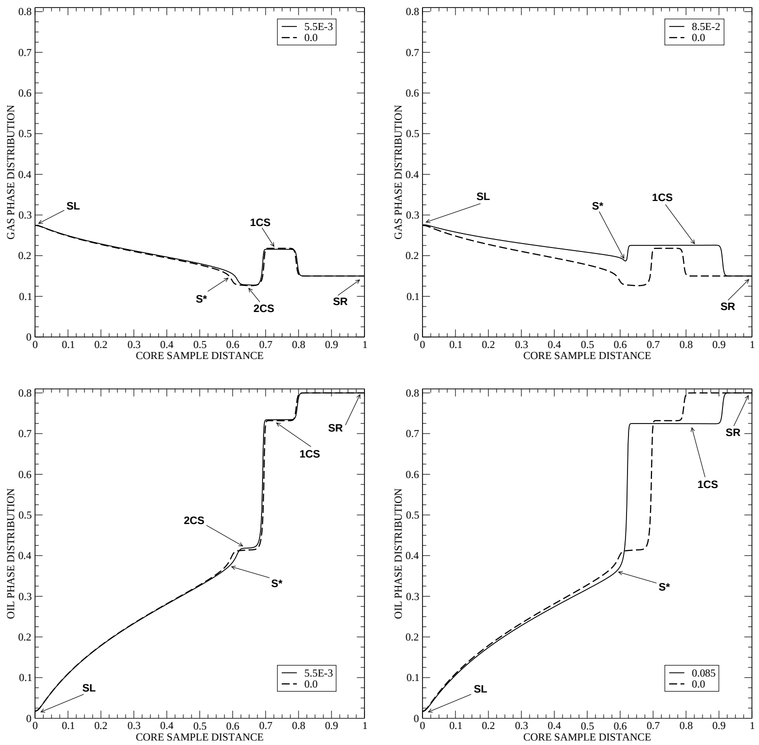


Figure 1. From top to bottom are shown Gas and Oil saturation profiles. On the left column, under influence of gravity ($\epsilon_G = 5.5 \times 10^{-3}$), we have a nonclassical structure: the slow wave group comprises a strong slow rarefaction fan from SL to S^* and an adjoining slow front wave from S^* to the constant state $2CS$. The fast wave group is a Buckley-Leverett front from the second constant state $1CS$ to SR. Between the slow and fast wave groups is a nonclassical front with left state $2CS$ and right state $1CS$ [11].

nonlinearities (viscous forces with or without gravity) is a open issue for three-phase flow; see references [2, 1] for an attempt on this question in the same lines of the current study.

As a model for multi-length scale rock heterogeneity we consider scalar, log-normal permeability fields, so that $\xi(\mathbf{x}) = \log K(\mathbf{x})$ is Gaussian and its distribution is determined by its mean and covariance function. We use the same multiscale field $\xi(\mathbf{x})$ to construct a correlated variable porosity field $\phi(\mathbf{x})$. This model is based on a random field model for anomalous diffusion in heterogeneous porous media (see, e.g., [2, 1] and references therein). The spatially variable permeability $K(\mathbf{x})$ and porosity $\phi(\mathbf{x})$ fields are defined on 512×128 grids with three values for the coefficient of variation CV (standard deviation/mean): CV is used as a dimensionless measure on the relative strength of heterogeneity of the porous medium. In this study, a mixture of water and gas is injected at a uniform and constant rate of 0.2 pore volume per year along the left boundary $(x, y) \in \{0\} \times [0, Y]$, $Y = 128$ for temporally constant boundary conditions for system (5)-(7), with production of the three-phase fluid mixture on the right boundary, in the same setting as in the one-dimensional numerical experiments.

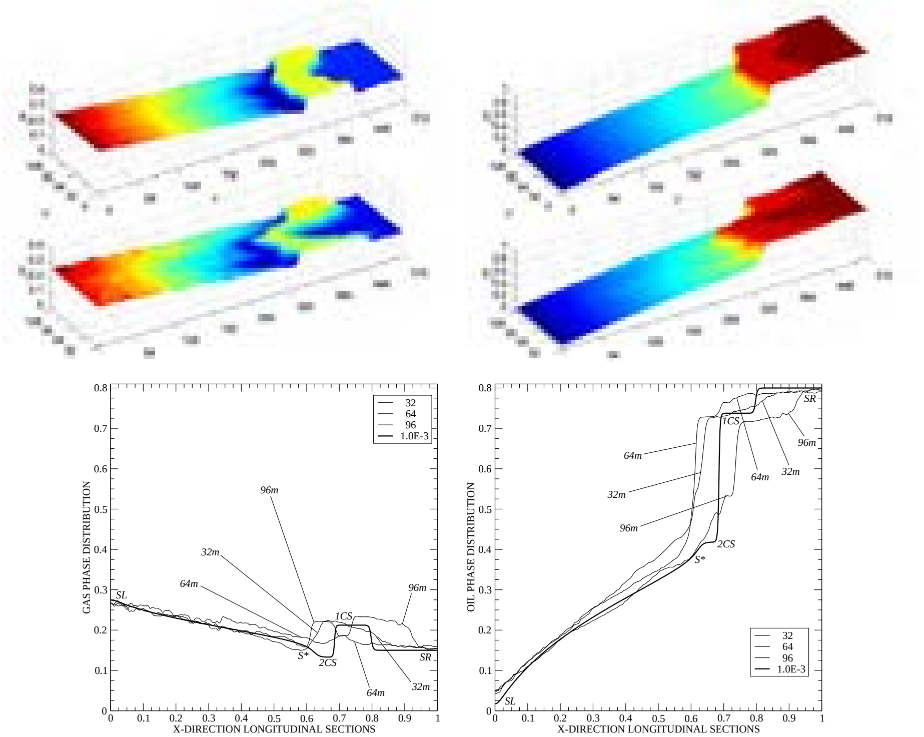


FIGURE 2. Gas (resp. Oil) phase solutions are shown at the left (resp. right) column as a function of distance for values $CV_k = 1.0$ and $CV_\phi = 0.25$. The one dimensional solutions (bottom frames) are $x-$direction longitudinal cross-sections for gas (left) and oil (right). We notice the ocurrence of a nonclassical wave [1, 2].

4.1.1. *Evidence of stable nonclassical wave in three-phase flow with gravity.* In order to better explain the numerical experiments, we consider "x-directional longitudinal cross-sections" at positions y = 32m, y = 64m and y = 96m (see Figure 2) of two-dimensional simulations for the three-phase system (5)-(7) in conjunction with superimposed one-dimensional numerical solutions obtained on grids having 512 cells. In Figure 2 oil and gas saturation surface plots are shown as a function of distance and gravity is present in the all numerical experiments with $\epsilon_G = 4.75 \times 10^{-2}$ and $\epsilon_P = 10^{-3}$. For completeness we describe the nonclassical solution [11]: the slow wave group comprises a strong slow rarefaction fan from SL to S^* and an adjoining slow front wave from S^* to the constant state $2CS$. The fast wave group is a Buckley-Leverett front wave from the second constant state $1CS$ to SR. Between the slow and fast wave groups is a nonclassical front wave (located at 320m-384m along the $x-$direction in the two-dimensional solutions) with left state $2CS$ and right state $1CS$. The numerical simulations with gravity effects reported in Figure 2 indicate that the nonclassical wave group is stable under the presence of heterogeneities imposed for long-range correlations and stronger heterogeneity induced by the variability of the permeability $K(\mathbf{x})$ and the porosity $\phi(\mathbf{x})$ fields (for both $CV_k = 0.5$ and $CV_k = 1.0$) with $CV_\phi = 0.25$. In addition, the nonclassical solution with left state $2CS$ and right state $1CS$ persist and it is accurately captured by our computational method, as shown by comparisons with reliable one-dimensional simulations and calculations for discontinuous media (Figure 2). This behavior with respect to the structurally stable wave group solutions is also observed in the one-dimensional experiments (see Figure 1) for homogeneous media with and without gravity, where there is not coupling between the transport equations (6)-(7) and the velocity field $\mathbf{v}$ dictated by the pressure system (5). This nonlinear behavior have not been reported on rigorous mathematical grounds in the literature (see, e.g., [11, 14, 7, 2, 1]). Although not exhaustive, the numerical simulations of the three-phase flow equations (5)-(7) show some evidence of stable nonclassical waves for two-dimensional problems taking into account gravity and hyperbolic-parabolic flux functions with explicit spatial variation. This important nonlinear mechanism between heterogeneity and viscous forces with gravity for three-phase flows is not yet understood. Therefore, a distinct study might be required to determine the dynamic behavior of three-phase flow solutions in discontinuous porous media.

REFERENCES

[1] E. Abreu, *Numerical modelling of three-phase immiscible flow in heterogeneous porous media with gravitational effects*, to appear in Mathematics and Computers in Simulation.

[2] E. Abreu and J. Douglas Jr. and F. Furtado and D. Marchesin and F. Pereira, *Three-phase immiscible displacement in heterogeneous petroleum reservoirs*, Mathematics and Computers in Simulation, **73** (2006), 2-20.

[3] Adimurthi, J. Jaffré and G. D. Veerappa Gowda, *Godunov-type methods for conservation laws with a flux function discontinuous in space*, SIAM J. Numer. Anal., **42** (2004), 179-208.

[4] A. Azevedo. and A. Souza and F. Furtado and D. Marchesin and B. Plohr, *The solution by the wave curve method of three-phase flow in virgin reservoirs*, Transport in Porous Media, **83** (2010), 99-125.

[5] C. Cancès, *Asymptotic behavior of two-phase flows in heterogeneous porous media for capillarity depending only on space. II. Nonclassical shocks to model oil-trapping.* SIAM J. Math. Anal., **42** (2010), 972-995.

[6] S. E. Gasda, M. W. Farthing, C. E. Kees, C. T. Miller, *Adaptive split-operator methods for modeling transport phenomena in porous medium systems*, Advances in Water Resources, **34** (2011), 1268-1282.

[7] E. Isaacson and D. Marchesin and B. Plohr, *Transitional waves for conservation laws*, SIAM J. Math. Anal., **21** (1990), 837-866.

[8] K. H. Karlsen, N. H. Risebro, *Corrected operator splitting for nonlinear parabolic equations*, SIAM Journal on Numerical Analysis, **37** (2000), 980-1003.

[9] K. H. Karlsen and J. D. Towers, *Convergence of the Lax-Friedrichs Scheme and Stability for Conservation Laws with a Discontinuous Space-Time Dependent Flux.* Chinese Annals of Mathematics (CAM), **25**(3) (2004), 287-318.

[10] A. Kurganov, G. Petrova, B. Popov, *Adaptive Semi-discrete Central-upwind Schemes*, SIAM J. Sci. Comput., **29** (2007), 2381-2401.

[11] D. Marchesin and B. Plohr, *Wave structure in wag recovery*, Society of Petroleum Engineering Journal, **71314** (2001), 209-219.

[12] S. Mishra and J. Jaffré, *On the upstream mobility scheme for two-phase flow in porous media.* Comput. Geosci., **14** (2010), 105-124.

[13] N. Nessyahu and E. Tadmor. *Non-oscillatory central differencing for hyperbolic conservation laws.* J. of Computational Physics, **87** (1990), 408-463.

[14] M. Shearer, *Nonuniqueness of admissible solutions of Riemann initial value problems for a system of conservation laws of mixed type*, Arch. Rational Mech. Anal., **93** (1986), 45-59.

E-mail address: eabreu@ime.unicamp.br

ON NONLINEAR CONSERVATION LAWS REGULARIZED BY A RIESZ-FELLER OPERATOR

Franz Achleitner and Sabine Hittmeir

Institute for Analysis and Scientific Computing
Vienna University of Technology
Wiedner Hauptstr. 8-10, 1040 Vienna, Austria

Christian Schmeiser

Faculty of Mathematics
University of Vienna
Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria

Abstract. Scalar one-dimensional conservation laws with nonlocal diffusion term are considered. The wellposedness result of the initial-value problem with essentially bounded initial data for scalar one-dimensional conservation laws with fractional Laplacian is extended to a family of Riesz-Feller operators.

The main interest of this work is the investigation of smooth traveling wave solutions. In case of a genuinely nonlinear smooth flux function we prove the existence of such traveling waves, which are monotone and satisfy the standard entropy condition. Moreover, the dynamic nonlinear stability of the traveling waves under small perturbations is proven, similarly to the case of the standard diffusive regularization, by constructing a Lyapunov functional.

Apart from summarizing our results in the article Achleitner et al. (2011), we provide the wellposedness of the initial-value problem for a larger class of Riesz-Feller operators.

1. **Introduction.** We consider one-dimensional conservation laws with nonlocal diffusion term

$$\partial_t u + \partial_x f(u) = \partial_x \mathcal{D}^\alpha u \tag{1}$$

for a scalar quantity $u : \mathbb{R}_+ \times \mathbb{R}$, $(t,x) \mapsto u(t,x)$, a smooth flux function $f : \mathbb{R} \to \mathbb{R}$ and a non-local operator

$$(\mathcal{D}^\alpha u)(x) = \frac{1}{\Gamma(1-\alpha)} \int_{-\infty}^{x} \frac{u'(y)}{(x-y)^\alpha} dy\,, \tag{2}$$

with $0 < \alpha < 1$.

1.1. **Motivation.** Conservation laws with nonlocal diffusion term of the form (1) appear in viscoelasticity - modeling the far-field behavior of uni-directional viscoelastic waves [11] - as well as in fluid mechanics - modeling the internal structure of hydraulic jumps in near-critical single-layer flows [9]. Moreover the nonlocal operator $\mathcal{D}^{1/3}$ appears in Fowler's equation

$$\partial_t u + \partial_x u^2 = \partial_x^2 u - \partial_x \mathcal{D}^{1/3} u\,, \tag{3}$$

2000 *Mathematics Subject Classification.* Primary: 47J35, 26A33, 35C07.

Key words and phrases. Nonlocal evolution equation, fractional derivative, traveling wave.

The authors acknowledge support by the Austrian Science Fund under grant numbers W8 and P18367.

which models the uni-directional evolution of sand dune profiles [7].

Equation (1) is closely related to

$$\partial_t u + \partial_x f(u) = D^{\alpha+1} u \tag{4}$$

with a fractional Laplacian $D^{\alpha+1} = (-\frac{\partial^2 u}{\partial x^2})^{(\alpha+1)/2}$, $0 < \alpha < 1$. This kind of nonlinear conservation law with nonlocal regularization has been studied e.g. in [3, 5].

Remark 1. The nonlocal operators $\partial_x \mathcal{D}^\alpha$, $0 < \alpha < 1$, and the fractional Laplacian $D^{\alpha+1}$, $0 < \alpha < 1$, are Fourier multiplier operators, i.e.

$$\mathcal{F}(\partial_x \mathcal{D}^\alpha u)(\xi) = -(\sin(\alpha\pi/2) - i\cos(\alpha\pi/2)\,\mathrm{sgn}(\xi))|\xi|^{\alpha+1}\mathcal{F}u(\xi)$$

and

$$\mathcal{F}(D^{\alpha+1}u)(\xi) = -|\xi|^{\alpha+1}\mathcal{F}u(\xi)\,,$$

whereat the Fourier transform $\mathcal{F}$ is defined as $\mathcal{F}\varphi(\xi) = \widehat{\varphi}(\xi) = \frac{1}{\sqrt{2\pi}} \int e^{-ix\xi}\varphi(x)dx$.

1.2. **Riesz-Feller operators.** Riesz-Feller operators [6, 13, 8] are Fourier multiplier operators

$$(\mathcal{F}D_{a,\theta}f)(\xi) = -\psi_{a,\theta}(-\xi)(\mathcal{F}f)(\xi)$$

whose multiplier $\psi_{a,\theta}(\xi) = |\xi|^a e^{(i\,\mathrm{sgn}(\xi)\theta\pi/2)}$ is the logarithm of the characteristic function of a general Lévy strictly stable probability density with *index of stability* $0 < a \leq 2$ and asymmetry parameter $|\theta| \leq \min(a, 2-a)$. The nonlocal operators $\partial_x \mathcal{D}^\alpha$, $0 < \alpha < 1$, and the fractional Laplacian $D^{\alpha+1}$, $0 < \alpha < 1$, are Riesz-Feller operators, see also Remark 1 and Figure 1.

Theorem 1.1. *For $1 < a \leq 2$ and $|\theta| \leq \min\{a, 2-a\}$, the Riesz-Feller operator $D_{a,\theta}$ generates a strongly continuous, convolution semigroup*

$$T(t) : L^p(\mathbb{R}) \to L^p(\mathbb{R})\,, \quad u_0 \mapsto T(t)u_0 = K(t,\cdot) * u_0\,,$$

with $1 \leq p < \infty$ and a convolution kernel $K(t,x) = \mathcal{F}^{-1}\exp(-t\psi(-.))(x)$ satisfying - for all $x \in \mathbb{R}$, $t > 0$ and $m \in \mathbb{N}$ - the properties

- *(non-negative) $K(t,x) \geq 0$,*
- *(integrable) $\|K(t,.)\|_{L^1(\mathbb{R})} = 1$,*
- *(scaling) $K(t,x) = t^{-\frac{1}{a}}K(1, xt^{-\frac{1}{a}})$,*
- *(smooth) $K(t,x)$ is $\mathcal{C}^\infty$ smooth,*
- *(bounded) there exists $B_m \in \mathbb{R}_+$ such that*

$$\left|\frac{\partial^m K}{\partial x^m}\right|(t,x) \leq t^{-\frac{1+m}{a}} \frac{B_m}{1 + t^{-\frac{2}{a}}|x|^2}\,.$$

The initial-value problem

$$\partial_t u + \partial_x f(u) = D_{a,\theta} u\,, \quad u(0,x) = u_0(x)\,, \tag{5}$$

for Riesz-Feller operators $D_{a,\theta}$ with *index of stability* $1 < a \leq 2$ and asymmetry parameter $a - 2 \leq \theta \leq 2 - a$ covers the special cases (1) and (4).

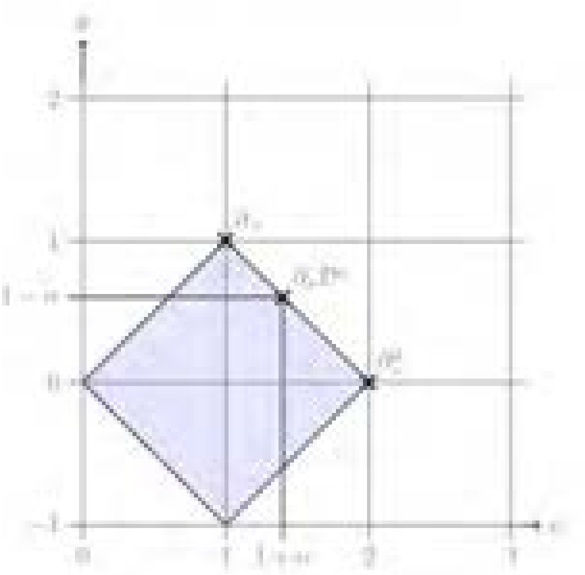


FIGURE 1. The family of Fourier multipliers $\psi_{a,\theta}(\xi) = |\xi|^a e^{(i\,\mathrm{sgn}(\xi)\theta\pi/2)}$ has two parameters a and θ. Some associated Fourier multiplier operators $(\mathcal{F}Tf)(\xi) = -\psi_{a,\theta}(-\xi)(\mathcal{F}f)(\xi)$ are displayed in the parameter space (a,θ). The Riesz-Feller operators $D_{a,\theta}$ are those operators, that take their parameters in the blue set, also known as Feller-Takayasu diamond. The family of operators $\partial_x\mathcal{D}^\alpha$, $0<\alpha<1$, interpolates formally between the first derivative ∂_x and second derivative ∂_x^2. Thus the limiting cases of equation (1) are a hyperbolic conservation law (for $\alpha = 0$) and a viscous conservation law (for $\alpha = 1$) [11].

Theorem 1.2. *Suppose $1 < a \le 2$ and $a-2 \le \theta \le 2-a$. If $u_0 \in L^\infty$, then there exists a unique solution $u \in L^\infty((0,\infty)\times\mathbb{R})$ of (5) satisfying the mild formulation*

$$u(t,x) = K(t,.) * u_0(x) - \int_0^t \left[\frac{\partial K}{\partial x}(t-\tau,.) * f(u(\tau,.))\right](x)\,\mathrm{d}\tau \tag{6}$$

almost everywhere. In particular

$$\|u(t,.)\|_\infty \le \|u_0\|_\infty, \qquad \text{for } t > 0\,,$$

and, in fact, u takes its values between the essential lower and upper bounds of u_0. Moreover, the solution has the following properties:

(i) $u \in C^\infty((0,\infty)\times\mathbb{R})$ and $u \in C_b^\infty((t_0,\infty)\times\mathbb{R})$ for all $t_0 > 0$.
(ii) u satisfies equation (5) in the classical sense.
(iii) $u(t) \to u_0$, as $t \to 0$, in $L^\infty(\mathbb{R})$ weak-$*$ and in $L^p_{loc}(\mathbb{R})$ for all $p \in [1,\infty)$.

Sketch of proof. The analysis of the initial-value problem for (4) by Droniou, Gallouët and Vovelle [5] depends on the properties in Theorem 1.1 of the semigroup (and its convolution kernel $K(t,x)$) generated by the fractional Laplacian $D^{\alpha+1}$ for $0 < \alpha < 1$. However all Riesz-Feller operators $D_{a,\theta}$ with *index of stability* $1 < a \le 2$

and asymmetry parameter $a-2 \le \theta \le 2-a$ share these properties. Thus the analysis in [5] carries over to the initial-value problem (5). □

2. Traveling wave solutions.

Definition 2.1. Suppose $(u_-, u_+, s) \in \mathbb{R}^3$. A traveling wave solution of (1) is a solution of the form $u(t,x) = \bar{u}(\xi)$ with $\xi := x - st$ and some function $\bar{u} : \mathbb{R} \to \mathbb{R}$ that connects the distinct endstates $\lim_{\xi \to \pm\infty} \bar{u}(\xi) = u_\pm$.

Inserting a traveling wave ansatz in (1) and integrating with respect to ξ yields the traveling wave equation

$$h(u) := f(u) - su - \big(f(u_-) - su_-\big) = \mathcal{D}^\alpha u = \frac{1}{\Gamma(1-\alpha)} \int_{-\infty}^{x} \frac{u'(y)}{(x-y)^\alpha}\, \mathrm{d}y\,, \qquad (7)$$

which is translation invariant.

If a smooth profile $\bar{u}$ approaches the endstates sufficiently fast, then the formal limit $\xi \to \infty$ in (7) leads to the Rankine-Hugoniot condition $f(u_+) - f(u_-) = s(u_+ - u_-)$.

If f is a convex flux function, then the vector field h is non-positive for values between u_- and u_+. Thus and due to the right-hand side of (7), a monotone traveling wave solution has to be monotone decreasing and the standard entropy condition $u_- > u_+$ has to hold.

The profile $\bar{u}$ of a traveling wave solution is governed by (7), whence its value at $\xi \in \mathbb{R}$ depends (only) on its values on the interval $(-\infty, \xi)$. Therefore, first the existence of a profile on an interval $(-\infty, \xi_\varepsilon]$ is established, subsequently its monotonicity and boundedness are verified and finally its global existence is deduced from an continuation argument.

The integral operator

$$\mathcal{D}^\alpha u(\xi) = \frac{1}{\Gamma(1-\alpha)} \int_{-\infty}^{\xi} \frac{u'(y)}{(\xi-y)^\alpha}\, \mathrm{d}y$$

is of Abel type and can be inverted by multiplying it with $(z-\xi)^{-(1-\alpha)}$ and integrating with respect to ξ from $-\infty$ to z. Thus the traveling wave problem

$$h(u) = \mathcal{D}^\alpha u\,, \qquad \lim_{\xi \to -\infty} \bar{u}(\xi) = u_-\,, \qquad \lim_{\xi \to +\infty} \bar{u}(\xi) = u_+\,, \qquad (8)$$

and

$$u(\xi) - u_- = \mathcal{D}^{-\alpha}(h(u))(\xi) := \frac{1}{\Gamma(\alpha)} \int_{-\infty}^{\xi} \frac{h(u(y))}{(\xi-y)^{1-\alpha}}\, \mathrm{d}y \qquad (9)$$

are equivalent if $u \in C_b^1(\mathbb{R})$ and $u' \in L^1(\mathbb{R}_-)$, and in particular if $u \in C_b^1(\mathbb{R})$ is monotone. Equation (9) is a nonlinear Volterra integral equation with a locally integrable kernel, where a well developed theory exists for problems on bounded intervals.

The linearizations of (8) and (9) at $\xi = -\infty$ (or, equivalently, at $u = u_-$) are

$$h'(u_-)v = \mathcal{D}^\alpha v \quad \text{and} \quad v = h'(u_-)\mathcal{D}^{-\alpha} v\,, \qquad (10)$$

respectively. Both linearizations have solutions of the form $v(\xi) = be^{\lambda\xi}$ with $\lambda = h'(u_-)^{1/\alpha}$ and arbitrary $b \in \mathbb{R}$, see also [4]. We will need that these are the only non-trivial solutions of (10) in the space $H^2(-\infty, \xi_0]$ for some $\xi_0 \le 0$. In particular, we assume that

$$\mathcal{N}\big(id - h'(u_-)\mathcal{D}^{-\alpha}\big) = \operatorname{span}\{\exp(\lambda\xi)\} \qquad \text{with} \qquad \lambda = h'(u_-)^{1/\alpha}\,, \qquad (11)$$

which is reasonable due to our analysis in [1, Appendix A].

In the existence result both formulations (8) and (9) will be used.

Theorem 2.2 ([1, Theorem 2])**.** *Suppose $f \in C^\infty(\mathbb{R})$ is a convex flux function, the shock triple (u_-, u_+, s) satisfies the Rankine-Hugoniot condition $f(u_+) - f(u_-) = s(u_+ - u_-)$ as well as the entropy condition $u_- > u_+$, and condition (11) holds. Then there exists a decreasing solution $u \in C_b^1(\mathbb{R})$ of the traveling wave problem (8). It is unique (up to a shift) among all $u \in u_- + H^2((-\infty, 0)) \cap C_b^1(\mathbb{R})$.*

Remark 2 (Extensions)**.** In [1] we prove the result assuming only

$$\begin{aligned} &h \in C^\infty([u_+, u_-])\,, \quad h(u_+) = h(u_-) = 0\,, \quad h < 0 \text{ in } (u_+, u_-)\,, \\ &\quad \exists\, u_m \in (u_+, u_-) \text{ such that } \quad h' < 0 \text{ in } (u_+, u_m) \text{ and } h' > 0 \text{ in } (u_m, u_-]\,. \end{aligned} \tag{12}$$

This is a little less than asking for convexity of f and the Lax entropy condition, since it covers the case $f'(u_+) \le s < f'(u_-)$.

The case of an concave flux function f can be analyzed in a similar way.

Idea of proof. The nonlinear problem has, up to translations, only two nontrivial solutions u_{down} and u_{up}, which can be approximated for large negative ξ by $u_- - e^{\lambda\xi}$ and $u_- + e^{\lambda\xi}$, respectively. The choice 1 of the modulus of the coefficient of the exponential is irrelevant due to the translation invariance of the traveling wave equations (7) and (9).

The traveling wave equation (7) involves a causal integral operator, i.e. to evaluate $\mathcal{D}^\alpha \bar{u}(\xi)$ at a point ξ the profile $\bar{u}$ on the interval $(-\infty, \xi]$ is needed. Thus, for $\varepsilon > 0$ and $\xi_\varepsilon := \log \varepsilon / \lambda$, we investigate the existence of solution $u_{down} : I_\varepsilon \to \mathbb{R}$ of (7) on the interval $I_\varepsilon = (-\infty, \xi_\varepsilon]$

$$\lim_{\xi \to -\infty} u_{down}(\xi) = u_- \quad \text{and} \quad u_{down}(\xi_\varepsilon) = u_- - \varepsilon\,. \tag{13}$$

Due to the analysis of the linearized equation (10) and assumption (11), the solution is written as $u_{down}(\xi) = u_- - \exp(\lambda\xi) + v$. Thus the perturbation v satisfies a boundary value problem (BVP)

$$\big(\mathcal{D}^\alpha - h'(u_-)\big)v = h(u_- - \exp(\lambda\xi) + v) + h'(u_-)\big(\exp(\lambda\xi) - v\big)\,, \quad v(\xi_\varepsilon) = 0\,.$$

This can be formulated as a fixed point problem for a given right-hand side in $H^2(I_\varepsilon)$ and an application of Banach's fixed point theorem yields the existence of u_{down} which is unique among all functions u satisfying (13) and $\|u - u_-\|_{H^2(I_\varepsilon)} \le \delta$ for some sufficiently small δ, which is independent of ε. Moreover

$$\|u_{down} - u_- + e^{\lambda\xi}\|_{H^2(I_\varepsilon)} \le C\varepsilon^2 \tag{14}$$

for some ε-independent constant C. The boundedness and monotonicity of u_{down},

$$u_{down}(\xi) < u_- \quad \text{and} \quad u'_{down}(\xi) < 0 \qquad \forall \xi \in I_\varepsilon\,,$$

follows from (14), a Sobolev embedding $H^2(\mathbb{R}) \hookrightarrow C^1(\mathbb{R})$ and the properties of $u_- - \exp(\lambda\xi)$.

Next, the continuation of the solution $u_{down} : (-\infty, \xi_\varepsilon] \to \mathbb{R}$ is proven. The boundedness and monotonicity of u_{down} imply that u_{down} is also a solution of (9). Due to the causality of the integral operator, (9) can be written as a Volterra integral equation on a bounded interval $[\xi_\varepsilon, \xi_\varepsilon + \delta)$ for some $\delta > 0$

$$u(\xi) = f(\xi) + \frac{1}{\Gamma(\alpha)} \int_{\xi_\varepsilon}^{\xi} \frac{h(u(y))}{(\xi - y)^{1-\alpha}}\, \mathrm{d}y\,.$$

with a well-defined inhomogeneity $f(\xi) = u_- + \frac{1}{\Gamma(\alpha)} \int_{-\infty}^{\xi_\varepsilon} \frac{h(u(y))}{(\xi-y)^{1-\alpha}} \,\mathrm{d}y$. The (local) existence of a smooth solution for sufficiently small δ is a standard result in the theory of Volterra integral equations on bounded intervals, see e.g. Linz [10].

Then, the boundedness and monotonicity of these continued solutions is proven, such that the argument for local existence can be iterated to imply the existence of a solution

$$u_{down} \in C_b^1(\mathbb{R}) \quad \text{with} \quad \lim_{\xi\to\infty} u_{down}(\xi) = u_- \,.$$

Finally, the proof of Theorem 2.2 is completed by proving $\lim_{\xi\to\infty} u(\xi) = u_+$. Assuming to the contrary $\lim_{\xi\to\infty} u(\xi) > u_+$, would imply $\lim_{\xi\to\infty} h(u(\xi)) < 0$. Then, however, $-\mathcal{D}^{-\alpha} h(u) = u_- - u$ would increase above all bounds, which is impossible by the boundedness of the solution. □

Remark 3 (Discussion of previous results)**.** Sugimoto and Kakutani [11, 12] studied the existence of traveling wave solutions of (1). They prove that bounded continuous traveling wave solution may exist, but give no analytical proof of existence, instead they construct numerical solutions and study the asymptotic behavior analytically.

In case of Burgers' equation with fractional Laplacian (4), Biler et al. [3] showed that no continuous traveling wave solutions can exist for $\alpha \in (-1, 0]$, however they provide no existence result for the case $\alpha \in (0, 1)$.

Alvarez-Samaniego and Azerad [2] proved the existence of traveling wave solutions of (3) with perturbation methods.

Remark 4 (Comparison with previous results)**.** The dynamical systems approach to prove the existence of traveling wave solutions in [1, Theorem 2], parallels the one in case of viscous conservation laws. This approach is possible due to the causality of the operator $\mathcal{D}^\alpha$ in (7) and the monotonicity of the profiles.

In contrast in case of a conservation law with fractional Laplacian (4) the traveling wave equation for traveling wave solutions $u(t,x) = \bar{u}(\xi)$ with $\bar{u} \in C_b^2(\mathbb{R})$ can be written as

$$h(u) := f(u) - su - \big(f(u_-) - su_-\big) = \frac{1}{\Gamma(1-\alpha)} \int_{-\infty}^{\infty} \frac{u'(y)}{|x-y|^\alpha} \,\mathrm{d}y \,.$$

Thus the value of a profile $\bar{u}$ at $\xi \in \mathbb{R}$ depends on the entire profile $\bar{u}$, such that a different approach is needed.

Whereas in case of Fowler's equation (3) the profile of a traveling wave solution is not necessarily monotone, such that the boundedness of a profile is difficult to establish.

2.1. Asymptotic stability of traveling wave solutions.

To study the asymptotic stability of traveling wave solutions ϕ of (1), equation (1) is cast in a moving coordinate frame $(t, x) \to (t, \xi = x - st)$,

$$\partial_t u + \partial_\xi (f(u) - su) = \partial_\xi \mathcal{D}^\alpha u \,, \tag{15}$$

such that a traveling wave solution becomes a stationary solution of (15). Analogous to viscous conservation laws asymptotic stability of ϕ is only to be expected for integrable zero-mass perturbations $U_0 := u_0 - \phi$, i.e.

$$\int_{\mathbb{R}} U_0(\xi) \,\mathrm{d}\xi = 0 \,. \tag{16}$$

The evolution of a perturbation $U := u - \phi$ is governed by

$$\partial_t U + \partial_\xi (f(\phi + U) - f(\phi) - sU) = \partial_\xi \mathcal{D}^\alpha U \,. \tag{17}$$

However the L^2-norms of the perturbation U and its derivative are not enough to construct a Lyapunov functional. Therefore the primitive

$$W(t,\xi) = \int_{-\infty}^{\xi} U(t,\eta)\ \mathrm{d}\eta$$

of the perturbation U has to be considered.

The flux function will be assumed to be convex between the far-field values $u_\pm$ of the traveling wave solution ϕ, i.e.

$$f''(\phi(\xi)) \geq 0 \quad \text{for all} \quad \xi \in \mathbb{R}\,. \tag{18}$$

Theorem 2.3 ([1, Theorem 4])**.** *Suppose $f \in C^\infty(\mathbb{R})$, the conditions (12) and (18) hold and ϕ is a traveling wave solution of (1) as in Theorem 2.2. Let u_0 be such that $W_0(\xi) = \int_{-\infty}^{\xi}(u_0(\eta) - \phi(\eta))\ \mathrm{d}\eta$ satisfies $W_0 \in H^2(\mathbb{R})$. If $\|W_0\|_{H^2}$ is small enough, then the initial-value problem for equation (15) with initial datum u_0 has a unique global solution converging to the traveling wave solution ϕ in the sense that*

$$\lim_{t\to\infty} \int_t^\infty \|u(\tau,.) - \phi\|_{H^1}^2\ \mathrm{d}\tau = 0\,. \tag{19}$$

Proof. First, the local-in-time wellposedness of the initial-value problem

$$\partial_t W + (f(U+\phi) - f(\phi) - sU) = \partial_\xi \mathcal{D}^\alpha W\,, \quad W(0,x) = W_0(x)\,, \tag{20}$$

is established by an fixed point argument [1, Proposition 2].

Then a (Lyapunov) functional

$$J(t) = \frac{1}{2}(\|W\|_{L^2}^2 + \gamma_1\|U\|_{L^2}^2 + \gamma_2\|\partial_\xi U\|_{L^2}^2)$$

is defined with positive constants $\gamma_1, \gamma_2 > 0$. The functional $J : H^2(\mathbb{R}) \to \mathbb{R}$, $W(t) \mapsto J(t)$, is equivalent to $\|W(t)\|_{H^2}^2$, since $\gamma_*\|W(t)\|_{H^2}^2 \leq 2J(t) \leq \gamma^*\|W(t)\|_{H^2}^2$ with $\gamma_* = \min\{1,\gamma_1,\gamma_2\}$ and $\gamma^* = \max\{1,\gamma_1,\gamma_2\}$. Combining the energy estimates of the perturbation U, its primitive W and its derivative $\partial_\xi U$, and using a Gagliardo-Nirenberg inequality yields

$$\begin{aligned}\frac{d}{dt}J + a_\alpha\left(\|W\|_{\dot{H}^{(1+\alpha)/2}}^2 + \gamma_1\|W\|_{\dot{H}^{(3+\alpha)/2}}^2 + \gamma_2\|W\|_{\dot{H}^{(5+\alpha)/2}}^2\right)&\\ - \gamma_1 C_0\|U\|_{L^2}^2 - \gamma_2 C_1\|U\|_{H^1}^2 - L(\|W\|_{H^2})\|W\|_{H^2}\|U\|_{H^{(5+\alpha)/4}}^2 &\leq 0\,,\end{aligned}$$

where $a_\alpha = \sin(\alpha\pi/2) > 0$ and $\dot{H}^s$ denotes the homogeneous Sobolev space of order s. Finally, the constants $\gamma_1, \gamma_2 > 0$ are chosen such that

$$\begin{aligned}&\gamma_1 C_0\|U\|_{L^2}^2 + \gamma_2 C_1\|U\|_{H^1}^2\\ &\qquad\leq \frac{a_\alpha}{2}\left(\|W\|_{\dot{H}^{(1+\alpha)/2}}^2 + \gamma_1\|W\|_{\dot{H}^{(3+\alpha)/2}}^2 + \gamma_2\|W\|_{\dot{H}^{(5+\alpha)/2}}^2\right)\,,\end{aligned}$$

which implies the final estimate

$$\begin{aligned}\frac{d}{dt}J + \left(\frac{a_\alpha}{2} - \frac{1}{\gamma_*}L(\|W\|_{H^2})\|W\|_{H^2}\right)\left(\|W\|_{\dot{H}^{(1+\alpha)/2}}^2 + \gamma_1\|W\|_{\dot{H}^{(3+\alpha)/2}}^2\right)&\\ + \gamma_2\left(\frac{a_\alpha}{2} - \frac{1}{\gamma_*}L(\|W\|_{H^2})\|W\|_{H^2}\right)\|W\|_{\dot{H}^{(5+\alpha)/2}}^2 &\leq 0\,.\end{aligned}$$

For initial data such that $J(0)$ is sufficiently small, the functional $J(t)$ - being equivalent to $\|W(t)\|_{H^2}^2$ - is non-increasing for all times. This implies the global-in-time existence of $W(t)$ as a solution of (20) and moreover (19). □

Remark 5. In case of Burgers' flux $f(u) = u^2$ and $\alpha > 1/2$, asymptotic stability of a traveling wave solution ϕ is established in case of $W_0 \in H^1(\mathbb{R})$, see also [1, Theorem 3].

Due to a Sobolev embedding $H^1(\mathbb{R}) \hookrightarrow C_b(\mathbb{R})$, the asymptotic stability result $\lim_{t\to\infty} \|U(t)\|_{H^1} = 0$ implies also $\lim_{t\to\infty} \|U(t)\|_{L^\infty} = 0$.

REFERENCES

[1] F. Achleitner, S. Hittmeir and C. Schmeiser, *On nonlinear conservation laws with a nonlocal diffusion term* J. Diff. Equ., **250** (2011), 2177–2196.

[2] B. Alvarez-Samaniego and P. Azerad, *Existence of travelling-wave solutions and local well-posedness of the Fowler equation*, Discrete Contin. Dyn. Syst. Ser. B, **12(4)** (2009), 671–692.

[3] P. Biler, T. Funaki and W. Woyczynski, *Fractal Burgers equations*, J. Diff. Equ., **148** (1998), 9–46.

[4] L.M.B.C. Campos, *On the solution of some simple fractional differential equations*, Internat. J. Math. & Math. Sci., **13(3)** (1990), 481–196.

[5] J. Droniou, T. Gallouët and J. Vovelle, *Global solution and smoothing effect for a non-local regularization of a hyperbolic equation*, J. Evol. Equ., **3** (2003), 499–521.

[6] W. Feller, "An Introduction to Probability Theory and Its Applications", Volume 2, 2^{nd} edition, John Wiley and Sons, 1971.

[7] A.C. Fowler, *Evolution equations for dunes and drumlins*, RACSAM, Rev. R. Acad. Cienc. Exactas Fís. Nat., Ser. A Mat., **96(3)** (2002), 377–387.

[8] R. Gorenflo and F. Mainardi, *Random walk models for space-fractional diffusion processes*, Fract. Calc. Appl. Anal., **1(2)** (1998), 167–191.

[9] A. Kluwick, E.A. Cox, A. Exner and C. Grinschgl, *On the internal structure of weakly nonlinear bores in laminar high Reynolds number flow*, Acta Mech., **210** (2010), 135–157.

[10] P. Linz, "Analytical and Numerical Methods for Volterra Equations", SIAM, Philadelphia, 1985.

[11] N. Sugimoto and T. Kakutani, *'Generalized Burgers equation' for nonlinear viscoelastic waves*, Wave Motion, **7** (1985), 447–458.

[12] N. Sugimoto, *'Generalized' Burgers equations and fractional calculus*, in "Nonlinear Wave Motion" (Edited by Alan Jeffrey), Longman Sci. Tech., (1989), 162–179.

[13] V.M. Zolotarev, "One-dimensional Stable Distributions", American Mathematical Society, Providence (RI), 1986.

E-mail address: franz.achleitner@tuwien.ac.at
E-mail address: sabine.hittmeir@tuwien.ac.at
E-mail address: christian.schmeiser@univie.ac.at

ON QUANTITATIVE COMPACTNESS ESTIMATES FOR HYPERBOLIC CONSERVATION LAWS

Fabio Ancona

Dipartimento di Matematica
Università degli Studi di Padova
Via Trieste 63, 35121 Padova, Italy

Olivier Glass

Ceremade, Université Paris-Dauphine, CNRS UMR 7534
Place du Maréchal de Lattre de Tassigny
75775 Paris Cedex 16, France

Khai T. Nguyen

Department of Mathematics
Penn State University
235, McAllister buiding, PA 16802, USA

Abstract. We are concerned with the compactness in L^1_{loc} of the semigroup $(S_t)_{t\geq 0}$ of entropy weak solutions generated by hyperbolic conservation laws in one space dimension. This note provides a survey of recent results establishing upper and lower estimates for the Kolmogorov ε-entropy of the image through the mapping S_t of bounded sets in $L^1 \cap L^\infty$, both in the case of scalar and of systems of conservation laws. As suggested by Lax [16], these quantitative compactness estimates could provide a measure of the order of "resolution" of the numerical methods implemented for these equations.

1. **Introduction.** Consider a system of conservation laws in one space dimension

$$u_t + f(u)_x = 0, \tag{1}$$

where $u : [0, +\infty) \times \mathbb{R} \to \mathbb{R}^N$ is the state variable, $f : \Omega \to \mathbb{R}^N$ is a twice continuously differentiable map, and Ω is an open set of $\mathbb{R}^N$. Assume that the above system is strictly hyperbolic, i.e, that the Jacobian matrix $Df(u)$ has N real, distinct eigenvalues $\lambda_1(u) < ... < \lambda_N(u)$ for all $u \in \Omega$. The fundamental paper of Bianchini and Bressan [5] shows that (1) generates a unique (up to the domain) Lipschitz continuous semigroup $\{S_t : \mathcal{D}_0 \to \mathcal{D}_0\}_{t\geq 0}$, defined on a closed domain $\mathcal{D}_0 \subset L^1(\mathbb{R}, \mathbb{R}^N)$, with the properties:

2000 *Mathematics Subject Classification.* Primary: 35L65, 47B06; Secondary: 94A17.

Key words and phrases. Hyperbolic equations, conservation laws, characteristics, compactness estimates, Kolmogorov entropy.

The first and third authors are partially supported by he European Union Programme [FP7-PEOPLE-2010-ITN] under grant agreement n.264735-SADCO, and by the ERC Starting Grant 2009 n.240385 ConLaws. The second author is partially supported by the Agence Nationale de la Recherche, Project CISIFS, grant ANR-09-BLAN-0213-02.

(i)

$$\Big\{u \in L^1(\mathbb{R},\Omega) \,\Big|\, \text{Tot.Var.}(u) \le \delta_0\Big\} \subset \mathcal{D}_0 \subset \Big\{u \in L^1(\mathbb{R},\Omega) \,\Big|\, \text{Tot.Var.}(u) \le 2\delta_0\Big\}, \tag{2}$$

for some suitable constant $\delta_0 > 0$.

(ii) For every $\overline{u} \in \mathcal{D}$, the semigroup trajectory $t \mapsto S_t u_0 \doteq u(t,\cdot)$ provides an entropy admissible weak solution of the Cauchy problem for (1), with initial data $u(0,\cdot) = \overline{u}$, that satisfy the following admissibility criterion proposed by T.P. Liu in [17], which generalizes the classical stability conditions introduced by Lax [15].

Liu stability condition. A shock discontinuity of the k-th family $(u^L,\, u^R)$, traveling with speed $\sigma_k[u^L,\, u^R]$, is *Liu admissible* if, for any state u lying on the k-th Hugoniot curve between u^L and u^R, the shock speed $\sigma_k[u^L, u]$ of the discontinuity (u^L, u) satisfies

$$\sigma_k[u^L,\, u] \ge \sigma_k[u^L,\, u^R]\,. \tag{3}$$

The uniform BV-bound on the elements of the domain $\mathcal{D}_0$ in (2) yield the compactness of the semigroup map S_t, for every $t > 0$. Actually, in the scalar case, when $f : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}$, the semigroup map $\{S_t\}_{t\ge0}$ generated by the equation (1) is defined on the whole space $L^1(\mathbb{R})$ (cfr. [9], [13]). This is in general not true for general hyperbolic systems. However, in the case of hyperbolic systems of conservation laws of Temple class [19, 20] with all characteristic family genuiney nonlinear or linearly degenerate, one can construct a continuous semigroup of solutions $\{S_t : \mathcal{D} \to \mathcal{D}\}_{t\ge0}$, defined on domains $\mathcal{D}$ of L^∞-functions with possibly unbounded variation (see [7], [4]). Relying on Oleĭnik-type inequalities on the decay of positive waves, one can still recover the compactness of the semigroup map S_t on domains of functions with unbounded variation in the case of scalar conservation laws with convex flux or of Temple systems with genuinely nonlinear characteristic families. Aim of this note is to discuss some recent results which have provided quantitative estimates of these compactness properties that reflect the irreversibility feature of these equations. Namely, following a suggestion of Lax [14], De Lellis and Golse have analized the Kolmogorov *ε-entropy* in L^1 of the image set $S_t(\mathcal{L})$ for bounded subsets $\mathcal{L}$ of the domain of S_t, and they have provided an upper bound on such a quantity in the case of scalar, convex, conservation laws [10]. We have next supplemented this estimate with a lower bound on the ε-entropy of $S_t(\mathcal{L})$, both in the case of scalar conservation laws [1] and in the case of systems [2], and we have established an upper bound on this quantity for Temple systems with genuinely nonlinear characteristic families [2]. We recall the following

Definition 1.1. Let (X, d) be a metric space and K a totally bounded subset of X. For $\varepsilon > 0$, let $N_\varepsilon(K)$ be the minimal number of sets in a cover of K by subsets of X having diameter no larger than 2ε. Then the ε-entropy of K is defined as

$$H_\varepsilon(K \mid X) \doteq \log_2 N_\varepsilon(K).$$

We remark that entropy numbers play a central roles in various areas of information theory and statistics as well as of learning theory. In the present setting, this concept could provide a measure of the order of "resolution" of a numerical method for (1), as suggested in [16].

2. **Scalar conservation laws.** In this section we assume that $N = 1$, $\Omega = \mathbb{R}$, and that $f : \mathbb{R} \to \mathbb{R}$ is a twice continuously differentiable, (uniformly) strictly convex function:

$$f''(u) \geq c > 0 \qquad \forall\ u \in \mathbb{R}. \tag{4}$$

Without loss of generality, by possibly performing a space and flux transformation, we may suppose that

$$f'(0) = 0. \tag{5}$$

The scalar equation (1) generates an L^1-contractive semigroup $S_t : L^1(\mathbb{R}) \to L^1(\mathbb{R})$ that associates to every initial data $\overline{u} \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$, the unique entropy solution $u(t,\cdot) = S_t\overline{u}$ of (1), with initial data $u(0,\cdot) = \overline{u}$. Now, given any $L, m, M > 0$, consider the set of bounded, compactly supported, initial data

$$\mathcal{L}_{[L,m,M]} \doteq \Big\{ \overline{u} \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R}) \mid \mathrm{Supp}\,(\overline{u}) \subset [-L, L],\ \ \|\overline{u}\|_{L^1} \leq m,\ \ \|\overline{u}\|_{L^\infty} \leq M \Big\}. \tag{6}$$

Since an entropy solution $u(t,x)$ of (1) satisfies the Oleĭnik estimate

$$\frac{u(t,y) - u(t,x)}{y - x)} \leq \frac{1}{c\,t} \qquad \forall\ x < y, \quad t > 0\,, \tag{7}$$

it follows that every map $x \mapsto S_t\overline{u}(x) - \frac{x}{ct}$, $\overline{u} \in \mathcal{L}_{[L,m,M]}$, is a nonincreasing function. Relying on this observation, and providing an estimates on the ε-entropy for class of bounded, nonincreasing functions with compact support, De Lellis and Golse have established in [10] the following

Theorem 2.1 ([10])**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be a twice continuously differentiable map, satisfying (4), (5). Then, given any $L, m, M, T > 0$, for $\varepsilon > 0$ sufficiently small, one has*

$$H_\varepsilon\big(S_T(\mathcal{L}_{[L,m,M]}) \mid L^1(\mathbb{R})\big) \leq \frac{\Gamma^+}{cT} \cdot \frac{1}{\varepsilon}\,, \tag{8}$$

where $\Gamma^+ \doteq 24\big(L + 2\sup_{|z|\leq M} |f''(z)| \sqrt{2mT/c}\,\big)^2$.

In fact, we have shown in [1] that the upper bound on the ε-entropy of $S_T(\mathcal{L}_{[L,m,M]})$ provided by (8) is actually optimal (w.r.t. the ε-dependence) as stated in the following

Theorem 2.2 ([1])**.** *Under the same assumptions of Theorem 2.1, given any L, m, $M, T > 0$, for $\varepsilon > 0$ sufficiently small, one has*

$$H_\varepsilon\big(S_T(\mathcal{L}_{[L,m,M]}) \mid L^1(\mathbb{R})\big) \geq \frac{\Gamma^-}{|f''(0)|\,T} \cdot \frac{1}{\varepsilon}\,, \tag{9}$$

where $\Gamma^- \doteq L^2/(48 \cdot \ln(2))$.

The main steps of the proof of the lower bound (9) are the following:

1. We introduce a two-parameter class $\mathcal{F}_{n,h}$ of piecewise affine functions and show that any element of such a class can be obtained, at any given time t, as the value $u(t,\cdot)$ of an entropy admissible weak solution of (1), with initial data in $\mathcal{L}_{[L,m,M]}$.
2. We provide an optimal estimate (w.r.t. the parameters n, h) of the maximum number of functions of $\mathcal{F}_{n,h}$ contained in a subset of $S_T(\mathcal{L}_{[L,m,M]})$ having diameter $\leq 2\varepsilon$. This estimate is established with a similar combinatorial argument as the one used in [3].

Similar upper and lower bounds as the ones stated in Theorem 2.1 and Theorem 2.2 have been derived in [1] also for solutions of scalar balance laws

$$u_t + f(u)_x = g(t,x,u), \tag{10}$$

with flux function $f(u)$ satisfying the assumptions (4), (5).

3. General hyperbolic systems of conservation laws.

In this section we assume that $f : \Omega \to \mathbb{R}^N$ is a twice continuously differentiable vector valued map, defined on an open, connected domain $\Omega \subset \mathbb{R}^N$ containing the origin, and that the system (1) is strictly hyperbolic. Then, given any $L, m, M > 0$, consider the set of bounded, compactly supported, initial data

$$\mathcal{L}_{[L,m,M]} \doteq \Big\{ \overline{u} \in \mathcal{D}_0 \mid \mathrm{Supp}\,(\overline{u}) \subset [-L,L],\ \ \|\overline{u}\|_{L^1} \leq m,\ \ \|\overline{u}\|_{L^\infty} \leq M \Big\}, \tag{11}$$

where $\mathcal{D}_0$ denotes a domain of the semigroup $\{S_t\}_{t\geq 0}$ generated by (1), satisfying (2). For the image set $S_t(\mathcal{L}_{[L,m,M]})$ of such a class we have established in [2] the following extension to general hyperbolic systems of the upper and lower bounds provided by Theorem 2.1 and Theorem 2.2 for scalar conservation laws.

Theorem 3.1 ([2]). *Let $f : \Omega \to \mathbb{R}^N$ be a map satisfying the above assumptions, and suppose that the system (1) is strictly hyperbolic. Then, given any $L, m, M, T > 0$, for $\varepsilon > 0$ sufficiently small, the following estimates hold.*

(i)

$$H_\varepsilon\Big(S_T\big(\mathcal{L}_{[L,m,M]}\big) \mid L^1(\mathbb{R},\Omega)\Big) \geq \frac{N^2 L^2}{T} \cdot \frac{\big(\min\big\{c_1, c_2 \frac{T}{L}\big\}\big)^2}{\max\big\{c_3,\ c_4 \frac{N^2 L}{T},\ c_5 \frac{NL}{\delta_0 T}\big\}} \cdot \frac{1}{\varepsilon}\,, \tag{12}$$

where c_l, $l = 1, \dots, 5$, are nonegative constants which depend only on the eigenvalues $\lambda_i(u)$ of the Jacobian matrix $Df(u)$, on the corresponding right and left eigenvectors $r_i(u), l_i(u)$, and on their derivatives, in a neighbourhood of the origin.

(ii)

$$H_\varepsilon\Big(S_T\big(\mathcal{L}_{[L,m,M]}\big) \mid L^1(\mathbb{R},\Omega)\Big) \leq 48 N \delta_0 \cdot L_T \cdot \frac{1}{\varepsilon}\,, \tag{13}$$

where

$$L_T \doteq L + \frac{\Delta_\vee \lambda}{2} \cdot T, \qquad \Delta_\vee \lambda \doteq \sup\big\{\lambda_N(u) - \lambda_1(v)\,;\ u,v \in \Omega\big\}. \tag{14}$$

Sketch of the Proof.

The upper bound stated in (ii) can be easily obtained relying on the upper estimates for the covering number of classes of functions with uniformly bounded total variation established in [3]. Therefore, we shall provide here only an outline of the proof of (i).

Step 1: *(A class of classical solutions)* We first consider a family of *simple waves*, i.e. of piecewise C^1 solutions of (1) that take values on the integral curves of the eigenvectors of the Jacobian matrix Df. Next, we construct a family of classical solutions of (1) with initial data given by the profiles of N simple waves (one for each characteristic family) supported on N disjoint sets.

For every k-th characteristic family, let $s \mapsto R_k(s)$ denote the integral curve of the eigenvector r_k, passing through the origin at $s = 0$. Then, given any $d, b > 0$, consider the class of functions

$$\mathcal{PC}^1_{[d,b]} \doteq \Big\{ \beta : \mathbb{R} \to [-d,d] \ \Big|\ \beta \text{ is piecewise } C^1 \text{ and } |\dot{\beta}(x)| \leq b \Big\}, \tag{15}$$

and, for every $\beta \in \mathcal{PC}^1_{[d,b]}$, define the map

$$\phi_k^\beta(x) \doteq R_k(\beta(x)) \qquad x \in \mathbb{R}\,. \tag{16}$$

Observe that, setting $x_k(t,y) \doteq y + \lambda_k(\phi_k^\beta(x))$, and $\alpha_1 \doteq \sup\{|\nabla\lambda_k(u)|\,;|u| \leq d\}$, it follows that the map $y \mapsto x_k(t,y)$ is one-to-one in $\mathbb{R}$, for all $t \leq \frac{1}{2\alpha_1\, b}$. Then, if $b \leq \frac{1}{2\alpha_1\, T}$, one can show that, setting $z_k(t,\cdot) \doteq\ x_k^{-1}(t,\cdot)$, the function

$$u(t,x) \doteq \phi_k^\beta(z_k(t,x)) \tag{17}$$

provides a classical solution of (1) on $[0,T]\times\mathbb{R}$, with initial data $u(0,x) = \phi_k^\beta(x)$.

Next, given $T, L > 0$, assume that $T \geq L/\Delta_\wedge\lambda$, $\Delta_\wedge\lambda \doteq \min\limits_k\big\{\lambda_{k+1}(0) - \lambda_k(0)\big\}$, and consider an N-tuple $\beta \doteq (\beta_1, \dots, \beta_N)$ of maps $\beta_k \in \mathcal{PC}^1_{[d,b]}$ supported on the disjoint intervals $I_i \doteq [-L/2 - \lambda_i(0)\,T,\, L/2 - \lambda_i(0)\,T]$, of the same length $|I_k| = L$. Let $\phi^\beta : \mathbb{R} \to \Omega$ denote the map that coincides with $\phi_k^{\beta_k}$ on every I_k, and vanishes elsewhere. Then, relying on the above analysis and deriving standard a-priori bounds on the solution of (1) and on its spatial derivatives (e.g. see [12, Section 4.2]), one can show that, if we take d, b sufficiently small, the Cauchy problem for (1) with initial data $u(0,x) = \phi^\beta(x)$ admits a classical solution $u(t,x)$ on $[0,T]\times\mathbb{R}$. Moreover, the following uniform bounds hold

$$\begin{gathered}\mathrm{Supp}(u(T,\cdot)) \subseteq [-\alpha_2 L,\, \alpha_2 L],\\ \|u(t,\cdot)\|_{L^\infty(\mathbb{R},\Omega)} \leq \alpha_3 N\cdot d, \qquad \|u_x(t,\cdot)\|_{L^\infty(\mathbb{R},\Omega)} \leq \alpha_4 N\cdot b, \qquad \forall\; t\in[0,T]\,,\end{gathered} \tag{18}$$

for some positive constants $\alpha_2, \alpha_3, \alpha_4$ (depending on λ_k, r_k, l_k, and on their derivatives in a neighbourhood of the origin). We shall denote by $\mathcal{PC}^{1,N}_{[L,d,b,T]}$ the class of N-tuples $\beta \doteq (\beta_1, \dots, \beta_N)$ of maps with the above properties.

Step 2: *(A controllability result)* Given any $L, m, M, T > 0$, we can show now that, setting

$$\mathcal{A}_{[L,d,b,T]} \doteq \Big\{\psi \in C(\mathbb{R},\Omega)\;\Big|\; \psi = \phi^\beta(-\cdot) \quad \text{for some} \quad \beta = (\beta_1,\dots,\beta_N) \in \mathcal{PC}^{1,N}_{[L,d,b,T]}\Big\}, \tag{19}$$

there holds

$$\mathcal{A}_{[\widetilde{L},d,b,T]} \subset S_T\big(\mathcal{L}_{[L,m,M]}\big)\,, \qquad \widetilde{L} \doteq L\cdot\min\big\{1/\alpha_2,\, \Delta_\wedge\lambda\cdot(T/L)\big\}, \tag{20}$$

if we take d sufficiently small and

$$b \leq \frac{1}{T}\cdot\frac{1}{\max\{2\alpha_1,\, \alpha_5\frac{N^2L}{T},\, \alpha_6\frac{NL}{\delta_0 T}\}}, \tag{21}$$

with α_1 as above and α_5, α_6 positive constants depending on λ_k, r_k, l_k, and on their derivatives in a neighbourhood of the origin. In fact, one can prove that for every $\psi \in \mathcal{A}_{[\widetilde{L},h,b,T]}$ there exists a classical solution $u(t,x)$ of (1) with initial data $\overline{u} \in \mathcal{L}_{[L,m,M]}$ so that $u(T,\cdot) = \psi$. This is accomplished with the same strategy adopted in [1] by reversing the direction of time and constructing a (classical) backward solution to (1) that starts at time T from ψ. The existence of such a solution $u(t,x)$ on $[0,T]\times\mathbb{R}$ is guaranteed by the analysis at Step 1, while the uniform bounds (18) imply that $u(0,\cdot) \in \mathcal{L}_{[L,m,M]}$.

Step 3: *(A class of profiles of superposition of simple waves)* Given any integer $n \geq 2$ and any constant $h > 0$, for every k-th characteristic family and for any given n-tuple $\iota = (\iota_1, \dots, \iota_n) \in \{0,1\}^n$, consider the function $\beta_k^\iota : \mathbb{R} \to [-h,h]$, with support contained in $[\xi_k^-, \xi_k^+]$, $\xi_k^\pm \doteq \pm L/2 - \lambda_k(0)\, T$, defined by setting (see Figure 1)

$$\beta_k^\iota(x) \doteq (-1)^{\iota_\ell}\, \frac{2hn}{L}\cdot\!\left(\frac{L}{2n} - \left|x - \xi_k^- - (2\ell+1)\cdot\frac{L}{2n}\right|\right) \qquad \forall\, x \in \left[\xi_k^- + \frac{\ell\, L}{nN},\ \xi_k^- + \frac{(\ell+1)L}{nN}\right], \tag{22}$$

for all $\ell \in \{0, \dots, n-1\}$.

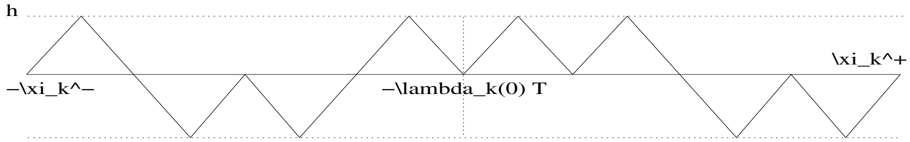


Figure 1: The function β_k^ι for $n = 8$ and $\iota = (1, -1 - 1, 1, 1, 1, -1 - 1)$.

Observe that, if we assume $h \le \min\{d, Lb/(2n)\}$, one has $\beta_k^\iota \in \mathcal{PC}^1_{[d,b]}$ for every n-tuple $\iota = (\iota_1, \dots, \iota_n) \in \{0,1\}^n$. Therefore, for any given N-tuple of n-tuples $(\iota_1, \dots, \iota_N) \in (\{0,1\}^n)^N$, letting $\beta_k^{\iota_k}$, $k = 1, \dots, N$, be maps defined as in (22), we have $(\beta_1^{\iota_1}, \dots, \beta_N^{\iota_N}) \in \mathcal{PC}^{1,N}_{[L,d,b,T]}$. Hence, setting

$$\mathcal{B}_{n,h} \doteq \Big\{ (\beta_1^{\iota_1}, \dots, \beta_N^{\iota_N}) \ \Big|\ \beta_k^{\iota_k} : \mathbb{R} \to [-h, h] \ \text{ defined as in (22) with} \\ \mathrm{Supp}(\beta_k^{\iota_k}) \subset [\xi_k^-, \xi_k^+] \ \ \forall\ k, \quad (\iota_1, \dots, \iota_N) \in (\{0,1\}^n)^N \Big\}, \tag{23}$$

one finds

$$\mathcal{B}_{n,h} \subset \mathcal{PC}^{1,N}_{[L,d,b,T]}. \tag{24}$$

Comparing the definitions (19), (23), it follows from (24) that, for all $n \ge 2$ and $h \le \min\{d, Lb/(2n)\}$, there holds

$$\mathcal{F}_{n,h} \doteq \Big\{ \phi^{\iota_1, \dots, \iota_N}(-\cdot) \ \Big|\ (\iota_1, \dots, \iota_N) \in (\{0,1\}^n)^N \Big\} \subset \mathcal{A}_{[L,d,b,T]}, \tag{25}$$

where we have used the notation $\phi^{\iota_1,\dots,\iota_N} \doteq \phi^{(\beta_1^{\iota_1},\dots,\beta_N^{\iota_N})}$ for a map defined as in Step 2 in connection with the N-tuple $(\beta_1^{\iota_1}, \dots, \beta_N^{\iota_N}) \in \mathcal{B}_{n,h}$.

Step 4: *(A combinatorial result)* Because of (25), in order to establish a lower bound on the ε-entropy $H_\varepsilon\big(\mathcal{A}_{[L,M,b,T]} \mid L^1(\mathbb{R}, \Omega)\big)$, it will be sufficient to provide such an estimate for $H_\varepsilon\big(\mathcal{F}_{n,h} \mid L^1(\mathbb{R}, \Omega)\big)$. Towards this goal, one can prove that, adopting the L^1-norm $\|(\beta_1, \dots, \beta_N)\|_{L^1} \doteq \sum_k \|\beta_k\|_{L^1}$ on the set $\mathcal{B}_{n,h}$ and the usual L^1-norm on the set $\mathcal{F}_{n,h}$, there holds

$$\big\|(\beta_1^{\bar\iota_1}, \dots, \beta_N^{\bar\iota_N}) - (\beta_1^{\iota_1}, \dots, \beta_N^{\iota_N})\big\|_{L^1} \le 2\big\|\phi^{\bar\iota_1,\dots,\bar\iota_N} - \phi^{\iota_1,\dots,\iota_N}\big\|_{L^1}, \tag{26}$$

for all $(\bar\iota_1, \dots, \bar\iota_N), (\iota_1, \dots, \iota_N) \in (\{0,1\}^n)^N$, if we assume h sufficiently small. Next, define

$$\mathcal{C}_n^{\mathcal{B}}(\varepsilon) \doteq \max_{\overline{\beta} \in \mathcal{B}_{n,h}} \#\Big\{\beta \in \mathcal{B}_{n,h} \ \Big|\ \|\beta - \overline{\beta}\|_{L^1} \le \varepsilon\Big\}, \tag{27}$$

(with the L^1-distance on $\mathcal{B}_{n,h}$ defined as above). Notice that, because of (26), any element of an ε-cover of $\mathcal{F}_{n,h}$ contains at most $\mathcal{C}_n^{\mathcal{B}}(4\varepsilon)$ functions of $\mathcal{F}_{n,h}$. Thus, since the cardinality of $\mathcal{F}_{n,h}$ is 2^{nN}, it follows that the number of sets in an ε-cover of $\mathcal{F}_{n,h}$ is at least

$$N_\epsilon(\mathcal{F}_{n,h} \mid L^1(\mathbb{R}, \Omega)) \ge \frac{2^{nN}}{\mathcal{C}_n^{\mathcal{B}}(4\varepsilon)}. \tag{28}$$

Observe now that

$$\left\|(\beta_1^{\bar\iota_1},\dots,\beta_N^{\bar\iota_N})-(\beta_1^{\iota_1},\dots,\beta_N^{\iota_N})\right\|_{L^1}=\frac{Lh}{n}\cdot d\big((\iota_1,\dots,\iota_N),(\bar\iota_1,\dots,\bar\iota_N)\big),\tag{29}$$

where

$$d\big((\iota_1,\dots,\iota_N),(\bar\iota_1,\dots,\bar\iota_N)\big)\doteq\#\Big\{(k,\ell)\in\{1,\dots,N\}\times\{1,\dots,n\}\ \Big|\ (\iota_k)_\ell\neq(\bar\iota_k)_\ell\Big\}.$$

Thus, if we fix an nN-tuple $\bar\iota\doteq(\bar\iota_1,\dots,\bar\iota_N)\in(\{0,1\}^n)^N$, and define the number

$$\mathcal{C}_n^{\mathcal{I}}(\varepsilon)\doteq\#\Big\{\iota\in(\{0,1\}^n)^N\ \Big|\ d(\iota,\bar\iota)\le\varepsilon\Big\},\tag{30}$$

which is independent on the choice of $\bar\iota\in(\{0,1\}^n)^N$, it follows from (29) that

$$\mathcal{C}_n^{\mathcal{B}}(4\varepsilon)=\mathcal{C}_n^{\mathcal{I}}\left(\frac{4n\varepsilon}{Lh}\right).\tag{31}$$

We can now derive an upper bound on $\mathcal{C}_n^{\mathcal{I}}(4n\varepsilon/(Lh))$ performing a standard combinatorial computation of the number of nN-tuples that differ for a given number of entries, and then invoking Hoeffding's inequality ([11, Theorem 2]). In this way we find

$$\mathcal{C}_n^{\mathcal{I}}\left(\frac{4n\varepsilon}{Lh}\right)\le 2^{nN}\cdot\exp\left(-\frac{nN}{2}\left(1-\frac{8\varepsilon}{LhN}\right)^2\right),\tag{32}$$

which, together with (28), (31), yields

$$N_\epsilon(\mathcal{F}_{n,h}\mid L^1(\mathbb{R},\Omega))\ge\left(\frac{nN}{2}\left(1-\frac{8\varepsilon}{LhN}\right)^2\right),\tag{33}$$

for all h sufficiently small and $n\ge2$. In order to derive the largest lower bound on the right-hand side of (33) we maximize the map $\Psi(h,n)\doteq\frac{nN}{2}\left(1-\frac{8\varepsilon}{LhN}\right)^2$ with respect to those h for which (33) holds. Thus, we find that the maximum of $\Psi(h,n)$ is attained for $\overline{n}=NL^2b/(48\varepsilon)$, $\overline{h}=24\varepsilon/(NL)$. Hence, we deduce from (33) that

$$N_\varepsilon(\mathcal{F}_{\overline{n},h_{\overline{n}}}\mid L^1(\mathbb{R},\Omega))\ge\exp\big(\Psi(h_{\overline{n}},\overline{n})\big)=\exp\left(\frac{L^2N^2b}{216}\cdot\frac{1}{\varepsilon}\right),\tag{34}$$

which, in turn, because of (25), yields

$$H_\varepsilon(\mathcal{A}_{[L,d,b,T]}\mid L^1(\mathbb{R},\Omega))\ge\frac{L^2N^2b}{216\ \ln(2)}\cdot\frac{1}{\varepsilon}.\tag{35}$$

Finally, recalling (20), (21), we recover from (35) the lower bound (12).

4. **Temple systems of conservation laws.** In this section we assume that (1) is a strictly hyperbolic system of Temple class, which thus enjoys the following additional properties:

- it is endowed with a coordinates system $w=(w_1,\dots,w_n)$ of Riemann invariants $w_k=W_k(u)$, $u\in\Omega$, associated to each characteristic field r_k;
- the level sets $\{u\in\Omega;\ w_i(u)=\textit{constant}\}$ of every Riemann invariant are hyperplanes.

We shall assume that $W(0)=0$, and that as w ranges within the product set $\Pi\doteq[a_1,b_1]\times\cdots\times[a_N,b_N]$, the corresponding state $u=W^{-1}(w)$ remains inside the domain Ω of the flux function f. We also recall that a characteristic field r_k of a system (1) is said to be *genuinely nonlinear* (GNL) in the sense of Lax if $\nabla\lambda_k(u)\cdot r_k(u)\neq0$ for all $u\in\Omega$, while we say that r_k is *linearly degenerate* (LD) if $\nabla\lambda_k(u)\cdot r_k(u)\equiv0$ for all $u\in\Omega$. As observed in the introduction, the results in [4], [7] show that a Temple system with GNL or LD characteristic families admits a continuous semigroup of entropy weak solutions $\{S_t : \mathcal{D} \to \mathcal{D}\}_{t\geq 0}$, defined on domains $\mathcal{D}$ of L^∞-functions with possibly unbounded variation of the form

$$\mathcal{D} \doteq \Big\{ u \in L^1(\mathbb{R}, \Omega) \;\Big|\; W_k(u(x)) \in [a_k, b_k] \;\text{ for all } x \in \mathbb{R}, \quad k = 1, \dots, N \Big\}. \tag{36}$$

We shall adopt the notation $S_t^w \overline{w} \doteq W(u(t,\cdot))$ for the Riemann coordinates expression of the entropy weak solution solution of (1), with initial data $\overline{u} \doteq W^{-1} \circ \overline{w}$. When all characteristic families are genuinely nonlinear such a semigroup is Lipschitz continuous and the map $w(t,x) \doteq S_t^w \overline{w}(x)$ satisfies the following Oleĭnik-type inequalities on the decay of positive waves:

$$\frac{w_k(t,y) - w_k(t,x)}{y-x} \leq \frac{1}{c\,t} \qquad \forall\; x < y, \quad t > 0, \quad k = 1, \dots, N, \tag{37}$$

for some constant $0 < c \leq \inf\big\{ |\nabla \lambda_k(u) \cdot r_k(u)|\,;\; u \in W^{-1}(\Pi),\, k = 1, \dots, N \big\}$. Relying on the analysis of the evolution of the Riemann coordinates along the characteristics and on the Oleĭnik-type inequalities, we have established in [2] a natural extension to this class of systems of the upper and lower bounds provided by Theorem 2.1 and Theorem 2.2 for scalar conservation laws with strictly convex (or concave) flux. Namely, adopting the norm $\|w\|_{L^1} \doteq \sum_i \|w_i\|_{L^1}$ on the space $L^1(\mathbb{R}, \Pi)$, we have proved the following

Theorem 4.1 ([2])**.** *In the same setting of Theorem 3.1, assume that the system (1) is of Temple class, strictly hyperbolic, and that all characteristic families are genuinely nonlinear or linearly degenerate. Then, given any $L, m, M, T > 0$, and setting*

$$\mathcal{L}_{[L,m,M]} \doteq \Big\{ \overline{w} \in L^1(\mathbb{R}, \Pi) \;\Big|\; \mathrm{Supp}(\overline{w}) \subset [-L, L],\; \|\overline{w}\|_{L^1} \leq m, \|\overline{w}\|_{L^\infty} \leq M \Big\}, \tag{38}$$

for $\varepsilon > 0$ sufficiently small, the following hold.

(i)

$$H_\varepsilon\Big(S_T^w\big(\mathcal{L}^w_{[I,m,M]}\big) \;\big|\; L^1(\mathbb{R}, \Pi) \Big) \geq \frac{N^2 L^2}{T} \cdot \frac{1}{\max\big\{c_6,\, c_7 \frac{NL}{T}\big\}} \cdot \frac{1}{\varepsilon}. \tag{39}$$

where c_6, c_7 are nonegative constants which depend only on the gradient of the eigenvalues $\lambda_i(u)$ of the Jacobian matrix $Df(u)$ and on the corresponding right eigenvectors $r_i(u)$, in a neighbourhood of the origin

(ii) If all characteristic families are genuinely nonlinear, one has

$$H_\varepsilon\Big(S_T^w\big(\mathcal{L}^w_{[I,m,M]}\big) \;\big|\; L^1(\mathbb{R}, \Pi) \Big) \leq \frac{32 N^2 L_T^2}{c\,T} \cdot \frac{1}{\varepsilon}, \tag{40}$$

where $L_T \doteq L + \sqrt{\frac{2mT}{c}} \cdot \sup\Big\{ |\nabla\lambda_k(u) \cdot r_j(u)|\,;\; |W(u)| \leq M,\; k, j = 1, \dots, N \Big\}$, and c is the constant appearing in (37).

Acknowledgments. FA and KTN are partially supported by the European Union Seventh Framework Programme [FP7-PEOPLE-2010-ITN] under grant agreement n.264735-SADCO, by the ERC Starting Grant 2009 n.240385 ConLaws, and by Fondazione CaRiPaRo Project "Nonlinear Partial Differential Equations: models, analysis, and control-theoretic problems".

REFERENCES

[1] F. Ancona, O. Glass and K. T. Nguyen, *Lower compactness estimates for scalar balance laws*, in "Comm. Pure Appl. Math", 65 (2012), 1303–1329.

[2] F. Ancona, O. Glass and K. T. Nguyen, *On compactness estimates for hyperbolic systems of conservation laws*, preprint, 2013.

[3] P. L. Bartlett, S. R. Kulkarni, S. E. Posner, *Covering numbers for real-valued function classes*, in "IEEE Trans. Inform. Theory", 43 (1997), 1721–1724.

[4] S. Bianchini, *Stability of L^∞ solutions for hyperbolic systems with coinciding shocks and rarefactions*, in "SIAM Journal on Mathematical Analysis" **33** (2001), no. 4, 959–981.

[5] S. Bianchini, A. Bressan, *Vanishing viscosity solutions to nonlinear hyperbolic systems*, in "Annals of Mathematics", **161**, 223–342, 2005.

[6] A. Bressan, "Hyperbolic systems of conservation laws" Oxford Lecture Series in Mathematics and its applications 20, Oxford University Press, Oxford, 2000.

[7] A. Bressan, P. Goatin, *Stability of L^∞ solutions of Temple class systems*, in "Differ. Integ. Equat." **13** (10-12), (2000), 1503–1528.

[8] C. M. Dafermos, *Generalized characteristics and the structure of solutions of hyperbolic conservation laws* in "Indiana Univ. Math. J", **26**, 1097–1119, 1977.

[9] C. M. Dafermos, "Hyperbolic conservation laws in continuum physics" Grundlehren Math. Wissenschaften Series, Vol. 325. Second Edition. Springer Verlag, 2005.

[10] C. De Lellis, F. Golse, *A Quantitative Compactness Estimate for Scalar Conservation Laws*, in "Comm. Pure Appl. Math", 58 (2005), 989–998.

[11] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, in "J. Amer. Statist. Assoc.", 58 (1963), 13–30.

[12] L. Hörmander, "Lectures on nonlinear hyperbolic differential equations" Matheamtiques & Applications, Vol. 26. Springer Verlag, Berlin, 1997.

[13] S. N. Kružkov, *First order quasilinear equations with several independent variables*, in "Mat. Sb. (N.S.)", 123 (1970), 228–255. (Russian) English translation in "Math. USSR Sbornik", (1970), 217–243.

[14] P. D. Lax, *Weak solutions of nonlinear hyperbolic equations and their numerical computation*, in "Comm. Pure Appl. Math.", 7 (1954), 159–193.

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

[16] P. D. Lax, *Accuracy and resolution in the computation of solutions of linear and nonlinear equations. Recent advances in numerical analysis*, in "Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis.", (1978). "Publ. Math. Res. Center Univ. Wisconsin, 107–117. Academic Press, New York, 1978."

[17] T. P. Liu, *The Riemann problem for general systems of conservation laws*, in "J. Differential Equations", **18**, 218–234, 1975.

[18] Oleinik O. A., *Discontinuous solutions of non-linear differential equations*, in "Uspehi Mat. Nauk (N.S.)", 12 (1957), 3–73. (Russian) English translation in "Ann. Math. Soc. Trans.", Ser. **2** 26, 95–172.

[19] D. Serre, Systemes de Lois de Conservation. II, Diderot Editeur, 1996.

[20] B. Temple, *Systems of conservation laws with invariant submanifolds*, in "Trans. Amer. Math. Soc." **280** (1983), 781–795.

E-mail address: ancona@math.unipd.it
E-mail address: glass@ceremade.dauphine.fr
E-mail address: khaivycy@gmail.com

ONE-DIMENSIONAL CONSERVATION LAW WITH BOUNDARY CONDITIONS: GENERAL RESULTS AND SPATIALLY INHOMOGENEOUS CASE

Boris Andreianov

Laboratoire de Mathématiques CNRS UMR 6623
Université de Franche-Comté, Besançon, France

Abstract. The note presents the results of the recent work [5] of K. Sbihi and the author on existence and uniqueness of entropy solutions for boundary-value problem for conservation law $u_t + \varphi(u)_x = 0$ (here, we focus on the simplified one-dimensional setting). Then, using nonlinear semigroup theory, we extend these well-posedness results to the case of spatially dependent flux $\varphi(x,u)$.

1. **Introduction.** Consider the general boundary-value problem for one-dimensional conservation law in $Q := (0,T) \times \Omega$ where we choose $\Omega := (-\infty, 0)$:

$$\begin{cases} u_t + \varphi(x,u)_x = 0 & \text{in } \ Q_T := (0,T)\times(-\infty,0) \\ u|_{t=0} = u_0 & \text{in } \ (-\infty,0) \\ \varphi_\nu(u) \in \beta(u) & \text{on } \Sigma := (0,T)\times\{0\}. \end{cases} \tag{$E_{\varphi,\beta}$}$$

Here φ is a regular function of (x,u); $\varphi_\nu(\cdot)$ will denote $\varphi(0,\cdot)$; and β is a maximal monotone graph on $\mathbb{R}$ that encodes the boundary condition. The simplest and best known case is $\beta = \{u^D\} \times \mathbb{R}$, which encodes the Dirichlet condition $u = u^D$ on Σ.

The well-posedness theory of the Cauchy problem associated with the conservation law $u_t + \varphi(x,u)_x = f$ was achieved in the founding work of Kruzhkov [13]. Taking into account the boundary condition is a delicate matter. Indeed, already in the Dirichlet case, the classical work of Bardos, LeRoux and Nédélec [7] states that, for u_0 of bounded variation, there exists a unique entropy solution in Q to the conservation law in ($E_{\varphi,\beta}$) which satisfies a relaxed formulation of the boundary condition; this relaxed formulation is justified, as in [13], by the vanishing viscosity argument. We aim at explaining in which way the general boundary condition $\varphi_\nu(u) \in \beta(u)$ should be relaxed; this is of interest, e.g., for obstacle problems ($\beta = \partial I_{[m,M]}$ where ∂ is the subdifferential and I is the indicator function) and for the zero-flux boundary condition ($\beta = \{0\} \times \mathbb{R}$); the latter condition is particularly important in practice. Notice that our setting provides a nontrivial extension of the result of Bürger, Frid and Karlsen [9] on the zero-flux problem: we do not assume $\varphi(0) = 0 = \varphi(1)$. Thus the first objective of this note is to point out, in a simplified setting, the meaning that can be given to the *formal boundary condition* "$\varphi_\nu(u) \in \beta(u)$ on Σ". We highlight the ideas and results of the recent work [5] of K. Sbihi and the author, where the multi-dimensional problem with spatially homogeneous flux ($\varphi = \varphi(u)$) but variable graphs $\big(\beta_{(t,x)}\big)_{(t,x)\in\Sigma}$ was explored.

2000 *Mathematics Subject Classification.* Primary: 35L65, 35L04; Secondary: 35A01, 35A02.
Key words and phrases. General boundary condition, Entropy solution, Integral solution.
The author is supported by the French ANR project CoToCoLa.

The second objective of the note is to generalize some of the results of [5]. Indeed, the assumption of x-independence of the flux played an important role in the formulation, because it allowed to consider *strong traces* for u on Σ (see [3, 5] for details). In the present note, traces need not exist; the construction we use, proposed in [1], is based upon the nonlinear semigroup techniques (see [6]). It strongly relies on the assumption $N = 1$ and on the t-independence of both φ and β. The semigroup approach allows to bypass as well the usual technical assumption

$$\text{for a.e. } x \in \Omega,\ \varphi(x,\cdot) \text{ is non-affine on any interval } [a,b] \text{ with } a < b. \tag{H1}$$

What we prove is that there exists an entropy solution in $[0,T] \times \Omega$ which verifies well-chosen up-to-the boundary entropy inequalities (see Definition 2.2) involving β. We interpret the information contained in these up-to-the-boundary inequalities as the *effective boundary condition* "$\varphi_\nu(u) \in \widetilde{\mathcal{B}}(u)$". Here, the maximal monotone graph $\widetilde{\mathcal{B}}$ is the projection of β on the graph of the function φ_ν, as shown on Fig. 1.

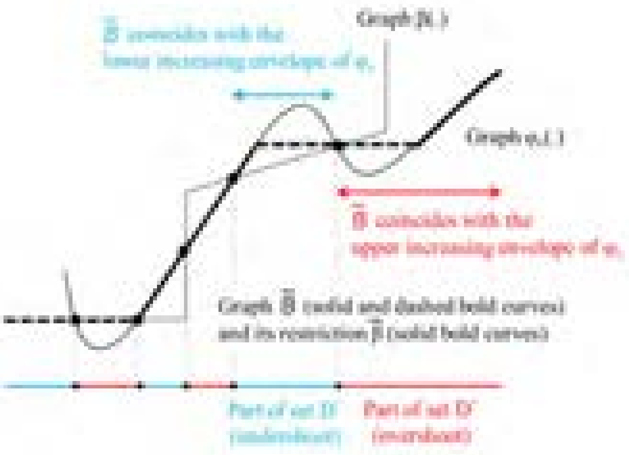

$\widetilde{\mathcal{B}}$ coincides with the lower increasing envelope of φ_ν
Graph $\beta(.)$
Graph $\varphi_\nu(.)$
$\widetilde{\mathcal{B}}$ coincides with the upper increasing envelope of φ_ν
Graph $\widetilde{\mathcal{B}}$ (solid and dashed bold curves) and its restriction $\widetilde{\beta}$ (solid bold curves)
Part of set D^- (undershoot)
Part of set D^+ (overshoot)

FIGURE 1. Construction of the projected graph $\widetilde{\mathcal{B}}$ and of $\widetilde{\beta}$

The main conclusion is: the graph β in the formulation of ($E_{\varphi,\beta}$) should be interpreted as its projection $\widetilde{\mathcal{B}}$. Indeed, the solution in the sense of Definition 2.2 can be attained as the limit of well-established approximation procedures (approximation of β by a kind of Yosida approximation or by "truncations" $\beta^{m,n} := \beta + I_{[-m,n]}$; the vanishing viscosity approximation involving the graph β or its approximates; and the Euler time-implicit discretization). Following Bardos, LeRoux and Nédélec [7], we see these facts as a justification of the notion of solution proposed for ($E_{\varphi,\beta}$).

2. Assumptions, definitions, results.

Definition 2.1. Extend β to a maximal monotone graph from $\overline{\mathbb{R}}$ to $\overline{\mathbb{R}}$ and define the *overshoot set* D^+ and the *undershoot set* D^- by

$$D^+ := \Big\{ z \in \overline{\mathbb{R}} \,|\, \sup \beta(z) \geq \varphi_\nu(z) \Big\}, \quad D^- := \Big\{ z \in \overline{\mathbb{R}} \,|\, \inf \beta(z) \leq \varphi_\nu(z) \Big\}.$$

Further, define the *crossing set* $D^0 := D^+ \cap D^-$. Finally, define $\widetilde{\mathcal{B}}$ on $\mathbb{R}$ as the closest to β maximal monotone graph that contains $\{(z, \varphi_\nu(z)) \,|\, z \in D^0\}$; and define $\widetilde{\beta}$ as the subgraph that $\widetilde{\mathcal{B}}$ and the graph $G\varphi_\nu$ of the function φ_ν have in common.

The "closest" to β graph $\widetilde{\mathcal{B}}$ does exist (see [5]). In fact, $\widetilde{\mathcal{B}}$ is single-valued and continuous, constituted of upper (respectively, lower) increasing envelopes of $G\varphi_\nu$ over connected components of D^+ (resp., D^-), see Fig. 1. It contains portions of $G\varphi_\nu$ complemented by horizontal segments over intervals of $\mathbb{R} \setminus \mathrm{Dom}\widetilde{\beta}$. As to $\widetilde{\beta}$, it is a maximal monotone subgraph of $G\varphi_\nu$. In the Dirichlet case ([7]), the graph $\widetilde{\beta}$ appeared in the work [12] as a way to express the Bardos-LeRoux-Nédélec condition.

Now, write $q^\pm(x,u,k)$ for $\mathrm{sign}\,(u-k)(\varphi(x,u)-\varphi(x,k))$ ("semi-Kruzhkov" entropy fluxes). Following [13] and adapting the boundary approach of Carrillo [10], we set

Definition 2.2. An $L^\infty(Q)$ function u is called entropy solution of problem $(E_{\varphi,\beta})$ if $u(0,\cdot)=u_0$[1] and u verifies the following inequalities[2]:

$$\forall k\in\mathbb{R}\ \ \forall\xi\in\mathcal{D}((0,T)\times\overline{\Omega}),\ \xi\ge 0,\ \text{ such that } \xi|_\Sigma=0 \text{ if } k\in D^\mp$$
$$\int_0^T\!\!\int_\Omega\Big(-(u-k)^\pm\xi_t-q^\pm(x,u,k)\cdot\nabla\xi\Big)+\int_0^T\!\!\int_\Omega \mathrm{sign}^\pm(u-k)\varphi_x(x,k)\,\xi\ \le\ 0 \tag{1}$$

Let us clarify the relation between the *formal boundary condition* "$\varphi_\nu(u)\in\beta(u)$" and the condition contained in Definition 2.2. We claim that, up to technical details

$$\varphi_\nu(u)\in\widetilde{\mathcal{B}}(u)\ \text{ on } \Sigma \tag{2}$$

is the boundary relation entailed by inequalities (1).

Proposition 2.3. *(see [5, Prop. 3.3]) In the case where u admits a strong boundary trace*[3] *γu on Σ, u is an entropy solution in the sense of Definition 2.2 if and only if it verifies the Kruzhkov inequalities with $\xi\in\mathcal{D}((0,T)\times\Omega)$, $\xi\ge 0$, and*

$$(\gamma u)(t)\in Dom\,\widetilde{\beta}\ \ for\ a.e.\ t\in(0,T). \tag{3}$$

Furthermore, in this situation (3) is equivalent to the property

$$\forall k\in D^\pm\quad q^\pm(0,(\gamma u)(t),k)\ge 0\ \ for\ a.e.\ t\in(0,T). \tag{4}$$

Since $\widetilde{\beta}$ is a subgraph of the graph of φ_ν, relation (3) means that $\varphi_\nu(u)\in\widetilde{\beta}(u)$ (so that $\varphi_\nu(u)\in\widetilde{\mathcal{B}}(u)$); therefore we say that (3) (or (2)) is the *effective boundary condition* for problem $(E_{\varphi,\beta})$. Condition (3) was introduced by K. Sbihi in her thesis [15] (see also [3, 4]). For details on the graph $\widetilde{\mathcal{B}}$, on the entropy formulation (1) and its different reformulations, a detailed study of existence and convergence of approximate procedures we refer to the recent paper [5] of Sbihi and the author.

The results of [5] for the x-independent flux $\varphi\in C(\mathbb{R})$ in space dimension one can be summarized as follows. Consider the assumption

$$\exists A>0\ \ \forall z\notin[-A,A]\ \ \mathrm{sign}\,(z)\,\phi_\nu(z)\ \le\ \mathrm{sign}\,z\,\beta(z). \tag{H2}$$

While (H2) is not required for well-posedness, we use it to get L^∞ estimate needed to prove that the vanishing viscosity method converges to the entropy solution in the sense (1). When (H2) is dropped, in order to justify the advent of $\widetilde{\beta}$ we need an additional stage of approximation of β by rapidly growing at infinity graphs $\beta^{m,n}$.

[1] We have $u\in C(0,T;L^1_{loc}(\Omega))$ since by (1), u is a Kruzhkov solution inside Q (see, e.g., [5]).

[2] Note that admissible test functions ξ in (1) are different for the "+" sign and for the "−" sign.

[3] In different contexts, such solutions were called *trace-regular* in [1, 2].

Theorem 2.4. *(compilation of results of [5], one-dimensional case, $\varphi(x,u)\equiv\varphi(u)$)*
(i) (uniqueness, comparison, contraction) If $u,\hat u$ are solutions of ($E_{\varphi,\beta}$) in the sense of Definition 2.2 with initial data $u_0,\hat u_0$ respectively, then for a.e. $t\in(0,T)$

$$\|(u-\hat u)^+(\cdot,t)\|_{L^1}\le\|(u_0-\hat u_0)^+\|_{L^1}. \tag{5}$$

(ii) (existence, construction of solution) Assume (H1). Then for all L^∞ datum u_0 there exists a (unique) solution u of ($E_{\varphi,\beta}$) in the sense of Definition 2.2.

If (H2) holds, then $u=\lim_{\varepsilon\to0}u^\varepsilon$; here u^ε is a weak solution to problem ($E_{\varphi,\beta}$) regularized by the vanishing viscosity term εu_{xx} (with "$\varphi_\nu(u)-\varepsilon u_x\in\beta(u)$" as boundary condition). And, if (H2) does not hold, then $u=\lim_{n,m\to\infty}u^{m,n}$ where $u^{m,n}$ is the vanishing viscosity limit for the boundary graph $\beta^{m,n}=\beta+I_{[-m,n]}$.

A crucial point of the uniqueness proof is that strong boundary traces $\gamma u,\gamma\hat u$ on Σ exist (see [3]). Generalization of the uniqueness result to the multi-dimensional case and space-time dependent graphs β is straightforward, but the boundary condition cannot always be formulated using (1) (e.g., (3) can be used instead). Existence results for this general case involve technical assumptions which main goal is to ensure uniform L^∞ estimates on the approximate solutions. The assumption on nonlinearity of φ can be relaxed if φ is Lipschitz continuous (see [5, Th. 7.1]). Truncations $\beta^{m,n}$ can be replaced by a two-parameter Yosida approximation (see [5, Ex. 6.14]).

Now, let us look at the case where $\varphi=\varphi(x,u)$. Because we focus on understanding the boundary condition, we consider a single boundary point $x=0$ and we avoid a few technical difficulties by taking the following (rather artificial) assumption:

$$\varphi(x,\cdot)\equiv0\ \text{ for } x\le-1. \tag{H3}$$

We also assume

$$\text{both } \varphi \text{ and } \varphi_x \text{ are Lipschitz continuous on } [-1,0]\times\mathbb{R}, \tag{H4}$$

$$\Phi: z\mapsto \max\nolimits_{x\in[-1,1]}\varphi_x(x,z) \text{ is Lipschitz continuous on } [-1,0]\times\mathbb{R}. \tag{H5}$$

Assumptions (H4),(H5) can be relaxed; in particular, the role of (H5) is to ensure an L^∞ estimate on solutions in the situation where (H2) holds.

For x-dependent flux φ, existence of strong boundary traces for Kruzhkov entropy solutions of $u_t+\varphi(x,u)_x=0$ in Q is yet not proved. Despite this obstacle, we will prove the following, which is the main new result of this note.

Theorem 2.5. *Assume (H3),(H4),(H5) hold. Assume $u_0\in L^1(-\infty,0)\cap L^\infty(-\infty,0)$. Then there exists a unique entropy solution of ($E_{\varphi,\beta}$) in the sense of Definition 2.2.*

As in Theorem 2.4, the existence proof justifies the notion of solution (see Remark 1). The key ingredient for the proof of Theorem 2.5 is the *stationary problem*

$$\begin{cases}\hat u+\varphi(x,\hat u)_x=f & \text{in } (-\infty,0)\\ \varphi_\nu(\hat u(0))\in\beta(\hat u(0)).\end{cases} \tag{$S_{\varphi,\beta}$}$$

Problem ($S_{\varphi,\beta}$) is used as a building brick in construction of a solution of ($E_{\varphi,\beta}$) via the time-implicit discretization, and it is essential for the uniqueness proof. To state a notion of solution, consider that $\hat u$ is an entropy solution of ($S_{\varphi,\beta}$) if it is a time-independent solution of ($E_{\varphi,\beta}$) with additional source term $f=g-\hat u$.

3. **Uniqueness, L^1 contraction and comparison proof: the ideas.** Using the Kruzhkov doubling of variables inside $[0,T]\times\Omega$, one gets[4] the *Kato inequality*

$$\int_\Omega \xi(u-\hat{u})^+(\cdot,t) \le \int_\Omega \xi(u_0-\hat{u}_0)^+ - \int_0^t\int_\Omega \nabla\xi\cdot q^+(x,u,\hat{u}) \tag{6}$$

with $\xi\in\mathcal{D}'(\Omega)$, $\xi\ge 0$, and for a.e. t. Here, we wish to let $\xi\to 1$ on Ω. If strong traces $\gamma u,\gamma\hat{u}$ on Σ exist, the last term passes to the limit and it yields the integral of $\operatorname{sign}^+(\gamma u-\gamma\hat{u})\big(\varphi_\nu(\gamma u)-\varphi_\nu(\gamma\hat{u})\big)$ over a part of Σ. Then, due to the characterization (3) and the monotonicity of $\widetilde{\beta}$, this term can be dropped and inequality (5) follows.

Now, in the situation where $\gamma\hat{u}$ exists but γu may not exist, we are still able to make the above arguments work. We use the following hint. Provided the strong trace $\gamma\hat{u}$ exists, the *weak trace* $\gamma_w q^\pm(\cdot,u(\cdot),\hat{u}(\cdot))$ on Σ (see [11]) verifies

$$\big(\gamma_w q^\pm(\cdot,u(\cdot),\hat{u}(\cdot))\big)(t) = \big(\gamma_w q^\pm(\cdot,u(\cdot),k)\big)(t)|_{k=(\gamma\hat{u})(t)}, \quad \text{for a.e. } t\in(0,T). \tag{7}$$

Furthermore, for a.e. t, $k=(\gamma\hat{u})(t)\in\operatorname{Dom}\widetilde{\beta}$ by Proposition 2.3. We also have

Lemma 3.1. *Assume u is an entropy solution of $(E_{\varphi,\beta})$ in the sense of Definition 2.2. Then for all $k\in D^\pm$, we have (respectively)*

$$(\gamma_w q^\pm(\cdot,u(\cdot),k))(t)\ge 0 \ \text{ for a.e. } t\in(0,T), \tag{8}$$

where γ_w denotes the weak boundary trace in the sense of Chen and Frid [11]. Furthermore, the two inequalities in (8) hold simultaneously for all $k\in \mathrm{Dom}\,\widetilde{\beta}$.

Proof. The first claim is straightforward. For the second one consider, e.g., $k\in D^+\cap\operatorname{Dom}\widetilde{\beta}$. Then it is enough to prove $\gamma_w q^-(\cdot,u)\cdot),k)\ge 0$. To this end, take $k_0\in[-\infty,k]$ such that k_0 is the closest to k point in D^-. By definition of $D^\pm$, we have $\varphi_\nu(\kappa)\le\varphi_\nu(k)$ for all $\kappa\in[k_0,k]$; hence $\varphi(x,\kappa)\le\varphi_\nu(k)+\overline{\overline{o}}_{x\to 0}(1)$. Developing the formula for $q(x,u(x),k)$, writing $\operatorname{sign}^-(u(x)-k) = 1\!\!1_{[u(x)\le k_0]}+1\!\!1_{[k_0<u(x)<k]}$, we find $q^-(x,u(x),k)\ge q^-(x,u(x),k_0)+\overline{\overline{o}}_{x\to 0}(1)$. Hence we can apply (8) with $k_0\in D^-$ and deduce that $\gamma_w q^-(\cdot,u(\cdot),k)\ge 0$. Details can be found in [5, Prop. 7.4(i)]. □

Finally, combining (7) and Lemma 3.1, passing to the limit as $\xi\to 1$ in Kato inequality (6) we get the desired result (5) whenever $\hat{u}$ is *trace-regular*, i.e., $\gamma\hat{u}$ exists.

In order to put ourselves in the situation where $\hat{u}$ is trace-regular, we will consider $\hat{u}=\hat{u}(x)$ entropy solution of $(S_{\varphi,\beta})$. Actually, in the preceding argument we only need that the trace of the function $V\varphi_\nu(\hat{u})$ exist, where $V\varphi_\nu : z\mapsto \int_0^z|\varphi_\nu'(s)|\,ds$ is the variation function of φ_ν (also known as the *singular mapping*). We refer to [3, 5] for the use of $V\varphi_\nu$ within the arguments involving the traces of $q^\pm(\cdot,u(\cdot),\hat{u}(\cdot))$. Existence of traces for solutions of $(S_{\varphi,\beta})$ follows, roughly speaking, from the fact that $q(\cdot,\hat{u}(\cdot),k)\in W^{1,1}(-\infty,0)$ for all $k\in\mathbb{R}$. We refer to [1, Lemma 3.1] for the proof in the case where φ_ν has finitely many extrema; the general case is similar.

Fortunately, comparison results concerning two solutions u and $\hat{u}$ of $(E_{\varphi,\beta})$ can be deduced from those concerning one solution $u\in L^1(Q)\cap L^\infty(Q)$ and all possible stationary solutions $v\in L^1(\Omega)\cap L^\infty(\Omega)$: to do this, one exploits the theory of nonlinear semigroups governed by m-accretive operators (details of this approach

[4] If φ is x-dependent, this result is not entirely contained in [13]: see [1, Th. 5.1] for the full argument that relies on the fact that a *local* entropy solution is a vanishing viscosity limit.

can be found in [8, 1, 2]). Roughly speaking, the above arguments prove that an entropy solution u of ($E_{\varphi,\beta}$) is an *integral solution* of the abstract evolution problem

$$\frac{d}{dt}u + \mathcal{A}u \ni 0, \quad u(0) = u_0 \tag{9}$$

where $\mathcal{A}$ is the operator associated with the formal expression $u(\cdot) \mapsto \varphi(\cdot, u(\cdot))_x$ in the entropy sense, as defined in (11) below. Then we apply the general result of uniqueness of an integral solution (see [6, 8] and Theorem 4.2 below).

4. Study of the stationary problem and use of the semigroup theory.

In the space $L^1 = L^1((-\infty, 0))$, consider the following definitions.

Definition 4.1 (elements of the nonlinear semigroup theory, see [6]).
- The *bracket* on L^1 is given by $\big[\, v\,,\, w \,\big] = \int w \operatorname{sign} v + \int w\, 1\!\!1_{[x\,|\,v(x)=0]}$.
- A multi-valued nonlinear operator $\mathcal{A}$ on L^1 is accretive if for all $(v, w), (\hat{v}, \hat{w}) \in \mathcal{A}$ one has $\big[v - \hat{v}, w - \hat{w}\big]_{L^1} \geq 0$. It is called m-accretive if, in addition, the domain $\mathrm{Dom}(I + \lambda\mathcal{A})^{-1}$ of the resolvent of $\mathcal{A}$ equals L^1 for all sufficiently small $\lambda > 0$.
- A function $u \in C([0, T]; L^1)$ is an integral solution of problem (9) if

$$\forall (v, w) \in \mathcal{A} \quad \frac{d}{dt}\|u(t) - v\|_{L^1} \leq \big[u(t) - v, 0 - w\big]_{L^1} \text{ in } \mathcal{D}'((0, T)). \tag{10}$$

The main result associated with these notions is the following (see, e.g., [6]).

Theorem 4.2. *Assume that $\mathcal{A}$ is accretive and its closure, m-accretive; assume $Dom\,\mathcal{A} = L^1$. Then for all $u_0 \in L^1$ there exists a unique integral solution to (9); further, two solutions with different data verify (5). Moreover, the integral solution is obtained by time-explicit discretization method (the Crandall-Liggett formula).*

Now, we apply the result to the operator $\mathcal{A}$ defined by its graph:

$$\begin{aligned}\mathcal{A} := \{(\hat{u}, g) \in L^1 \times L^1 \,|\, \hat{u} \text{ is an entropy solution of } (S_{\varphi,\beta}) \text{ with } f = \hat{u} + g,\\ \text{in particular, } V\varphi_\nu(\hat{u}(\cdot)) \text{ is continuous at } x = 0^-\}.\end{aligned} \tag{11}$$

Proposition 4.3 (properties of the stationary problem ($S_{\varphi,\beta}$)).
(i) The operator $\mathcal{A}$ is accretive on L^1, moreover, its closure is m-accretive on L^1.
(ii) The domain $Dom\,\mathcal{A}$ is dense in L^1.

Proof. The accretivity in (i) follows by rewriting the arguments of the beginning of this section for stationary solutions with strong traces (to be precise, with those of $V\varphi_\nu(u)$). In the place of (5) we find the refined contraction property

$$\|u - \hat{u}\|_{L^1} \leq \int_\Omega \operatorname{sign}(u - \hat{u})(f - \hat{f}) + \int_\Omega 1\!\!1_{[x\,|\,u(x)=\hat{u}(x)]}|f - \hat{f}| \tag{12}$$

which, together with the definition of $\mathcal{A}$, implies its accretivity in L^1.

Further, the m-accretivity of the closure of $\mathcal{A}$ is an existence claim for ($S_{\varphi,\beta}$) (with flux $\lambda\varphi$ and λ small enough) for some L^1-dense set of specific data. E.g., it is enough to prove that the set C^1_c of compactly supported in $(-\infty, 0]$ functions of class C^1 is included in the domain $\mathrm{Dom}(I + \lambda\mathcal{A})^{-1}$ of the resolvent of $\mathcal{A}$, for all $\lambda > 0$ small enough. This claim is proved using vanishing viscosity approximation.

First, we solve $u^\varepsilon + \lambda\varphi(x, u^\varepsilon)_x = \varepsilon u^\varepsilon_{xx} + f$ subject to the boundary condition $\varphi_\nu(u^\varepsilon) + \varepsilon u^\varepsilon_x \in \beta(u^\varepsilon)$ at $x = 0$. Existence of a weak (variational) solution u^ε follows by adapting classical arguments (β can be regularized, then the problem is reduced to a coercive nonlinear elliptic problem in $H^1(-\infty, 0)$ for which a solution can be constructed by a Leray-Schauder argument). Following Carrillo [10], we can get the comparison result analogous to (5), with $\varepsilon > 0$.

Then compactness of the sequence $(u^\varepsilon)_\varepsilon$ should be obtained, and for this we need firstly a uniform L^∞ estimate on the solution. This estimate follows from the comparison of u^ε with constant functions. To see this, observe that $k \in \mathbb{R}^+$ is a solution of $k + \lambda\varphi(x,k)_x = k + \lambda\varphi_x(x,k) \geq 0 = \varepsilon k_{xx}$ for λ small enough (here, assumption (H5) is used). Thus the constant k is a super-solution of the equation inside $(-\infty, 0)$; the delicate point is to ensure that k is a super-solution of our viscosity regularized boundary-value problem with graph β. This is true for $k > A$ provided assumption (H2) holds; thus we temporarily assume (H2). In a similar way, we prove that $k < -A$ is a sub-solution, and by the comparison argument we find a uniform in ε bound in L^∞ on u^ε in terms of A and the right-hand side f.

Following [7], let us get a uniform BV estimate on $(u^\varepsilon)_\varepsilon$. For $v^\varepsilon := u^\varepsilon_x$ we have

$$(1 + \lambda\varphi_{xu}(x,u^\varepsilon))v^\varepsilon + (\lambda\varphi_u(x,u^\varepsilon)v^\varepsilon - \varepsilon v^\varepsilon_x)_x = f_x - \lambda\varphi_{xx}(x,u^\varepsilon). \tag{13}$$

By (H3) and because f is compactly supported, $u^\varepsilon - \varepsilon u^\varepsilon_{xx} = 0$ for $x < -1$; since $\|u^\varepsilon\|_\infty \leq \mathit{const}$, we get $|u^\varepsilon(x)| \leq \mathit{const}\, e^{-|x|}$ for all x. Now, the flux of (13) is $F^\varepsilon := \lambda\varphi_u(\cdot,u^\varepsilon)v^\varepsilon - \varepsilon v^\varepsilon_x$; it verifies $F^\varepsilon(0) = \lambda\varphi_x(0,u^\varepsilon(0)) - \int_{-\infty}^0 (f - u^\varepsilon)(y)\,dy$. By (H4), $|F^\varepsilon(0)| \leq \mathit{const}$ is bounded. Further, $1 + \lambda\varphi_{xu}(\cdot,u^\varepsilon(\cdot)) \geq \frac{1}{2}$ for small enough λ, due to (H4). Now we take a Lipschitz approximation of $\operatorname{sign} v^\varepsilon$ as a test function; a uniform estimate of $\int_{-\infty}^0 |v^\varepsilon| = \|u^\varepsilon_x\|_{L^1}$ follows. Due to the exponential decay of u^ε at $-\infty$, we see that $(u^\varepsilon)_\varepsilon$ admits an accumulation point $u \in L^1$.

Now, it remains to write entropy inequalities for u^ε and pass to the limit. From the weak formulation, using Lipschitz approximations of $\operatorname{sign}^\pm(u^\varepsilon - k)$ as test functions (see, e.g., [10] and [1, Appendix]), for all $k \in \mathbb{R}$ we get

$$\begin{aligned}\int_\Omega \Big((u^\varepsilon - k)^\pm \xi - q^\pm(x,u^\varepsilon,k)\cdot\nabla\xi\Big) + \int_\Omega \operatorname{sign}^\pm(u^\varepsilon - k)\varphi_x(x,k)\xi \\ \leq \int_\Omega \varepsilon|u^\varepsilon - k|_x\xi_x \;-\; \operatorname{sign}^\pm(u^\varepsilon(0) - k)(b^\varepsilon - \varphi_\nu(k))\xi(0)\end{aligned} \tag{14}$$

where $b^\varepsilon \in \beta(u^\varepsilon(0))$ (the last term is a boundary term). Convergence of u^ε and the classical uniform estimate of $\|\varepsilon(u^\varepsilon_x)^2\|_{L^1}$ permit to pass to the limit in all terms except for the last one, which we will bound from above. Consider, e.g., $k \in D^+$. In the "sign$^+$" inequality (14), the monotonicity of β and the choice of k yield

$$-\operatorname{sign}^+(u^\varepsilon(0) - k)(b^\varepsilon - \varphi_\nu(k)) \leq (b(k) - \varphi_\nu(k))^- = 0. \tag{15}$$

Further, we simply impose $\xi(0) = 0$ in the "sign$^-$" inequality (14) and the boundary term vanishes. Thus we arrive to the stationary analogue of inequalities (1) with the adequate choice of ξ. This proves that u is an entropy solution of $(S_{\varphi,\beta})$.

It remains to bypass (H2). This is done by working firstly with truncated graphs $\beta^{m,n}$ that do satisfy (H2). Convergence of the associated solutions $u^{m,n}$ to a limit u is ensured by monotonicity (see [4, 5]). Then, as in [4], one observes that the boundary conditions for graphs $\tilde\beta^{m,n}$ pass to the limit (e.g., if φ_ν is monotone near $\pm\infty$, then $\widetilde{\mathcal{B}}^{m,n}$ coincide with $\widetilde{\mathcal{B}}$ for large enough n, m). To be specific, if $k \in D^{\pm;n,m}$ (the overshoot or undershoot set defined for graph $\beta^{m,n}$) then $k \in D^\pm$ for n, m large enough. Thus entropy inequalities for $u^{m,n}$ yield analogous inequalities for u.

As to the claim (ii), it can be proved by showing that as $\lambda \to 0$, the solution of $u + \lambda\mathcal{A} = f$ converges to f in L^1; see [1] for details corresponding to our case. □

With Theorem 4.2 and Prop. 4.3, we get the uniqueness claim of Theorem 2.5:

Proposition 4.4. *An entropy solution of* ($E_{\varphi,\beta}$) *is an integral solution of* (9) *with* $\mathcal{A}$ *defined by* (11)*. In particular, there exists at most one entropy solution for given datum, and we have* (5) *for entropy solutions* $u, \hat{u}$ *with data* $u_0, \hat{u}_0$*.*

5. **Existence of solution, justification of the effective boundary condition.** In order to prove existence of an entropy solution, we can restrict our attention to $L^1 \cap L^\infty$ data due to assumptions (H3),(H4). Let us give two existence arguments.

Under the genuine nonlinearity assumption (H1) on $\varphi(x,\cdot)$, one can follow closely the existence proof of Proposition 4.3(i), substituting the stationary problem by the evolution problem. Indeed, (H5) (along with (H2)) allows to construct super- and sub-solutions of ($E_{\varphi,\beta}$) under the form $\hat{u}(t,x) = k(t)$. The uniform L^∞ bound along with assumption (H1) ensures compactness of $(u^\varepsilon)_\varepsilon$ in L^1_{loc} (see Panov [14]). In this argument, we do not need Lipschitz regularity of φ_x in (H4).

In general we do not assume (H1); we exploit the existence result for ($S_{\varphi,\beta}$) and the Crandall-Liggett construction of Theorem 4.2. Indeed, in this case the time compactness comes for gratis; one only has to show that the integral solution (also known as the mild solution) coming from Crandall-Liggett formula is also an entropy solution (cf. [6]). This itinerary was taken in the work of K. Sbihi ([15], see also [3]). In the setting of the present note, the proof is much simpler than in [15, 3] since the stability of the entropy formulation (1) by L^1 convergence is evident.

Thus we achieve the following result and complete the proof of Theorem 2.5:

Proposition 5.1. *For all* $u_0 \in L^1((-\infty,0)) \cap L^\infty((-\infty,0))$ *there exists the integral solution to* (9) *which is also the unique entropy solution of* ($E_{\varphi,\beta}$)*. For general* L^∞ *datum* u_0*, there exists a unique entropy solution of* ($E_{\varphi,\beta}$) *obtained as the* $L^1_{loc}(Q)$ *limit of solutions* u_n *with* $L^1 \cap L^\infty$ *data* $u_{0,n}(\cdot) := u_0(\cdot)1\!\!1_{[-n,0]}(\cdot)$*.*

Remark 1. To conclude the note, let us stress that the entropy formulation (1) and the projected graph $\widetilde{\mathcal{B}}$ naturally appeared from the vanishing viscosity approximation of ($E_{\varphi,\beta}$) or ($S_{\varphi,\beta}$): the main arguments here were (14), (15), and Prop. 2.3.

REFERENCES

[1] B. Andreianov, Semigroup approach to conservation laws with discontinuous flux. Springer Proc. in Math. and Stat., G-Q. Chen, H. Holden and K.H. Karlsen, eds., 2013.

[2] B. Andreianov and M. K. Gazibo. *Entropy formulation of degenerate parabolic equation with zero-flux boundary condition.* ZAMP Zeitschr. Angew. Math. Phys. (2013), published online, doi:10.1007/s00033-012-0297-6

[3] B. Andreianov and K. Sbihi, *Strong boundary traces and well-posedness for scalar conservation laws with dissipative boundary conditions.* Hyperbolic problems: theory, numerics, applications (Proc. of the HYP2006 Conference, Lyon), Springer, Berlin, pp.937–945, 2008.

[4] B. Andreianov and K. Sbihi, *Scalar conservation laws with nonlinear boundary conditions*, C. R. Acad. Sci. Paris, Ser. I 345 (2007), pp.431–434.

[5] B. Andreianov and K. Sbihi, *Well-posedness of general boundary-value problems for scalar conservation laws*, Transactions AMS, accepted. Available as hal-00708973 preprint.

[6] F. Andreu-Vaillo, V. Caselles and J.M. Mazón, *Parabolic quasilinear equations minimizing linear growth functionals.* Progress in Mathematics, 223. Birkhäuser, Basel, 2004.

[7] C. Bardos, A.Y. Le Roux, and J.-C. Nédélec, *First order quasilinear equations with boundary conditions, Comm. Partial Diff. Equ.* 4 (1979), no.4, pp.1017–1034.

[8] Ph. Bénilan, P. Wittbold, *On mild and weak solutions of elliptic-parabolic problems*, Adv. Differ. Equ. 1 (1996), pp.1053–1073.

[9] R. Bürger, H. Frid, and K.H. Karlsen, *On the well-posedness of entropy solutions to conservation laws with a zero-flux boundary condition*, J. Math. Anal. Appl. 326 (2007), pp.108–120.

[10] J. Carrillo, *Entropy solutions for nonlinear degenerate problems.* Arch. Ration. Mech. Anal. 147 (1999), no.4, pp.269–361.

[11] G.-Q. Chen and H. Frid, *Divergence-Measure fields and hyperbolic conservation laws,* Arch. Ration. Mech. Anal. 147 (1999), pp.89–118.

[12] F. Dubois and Ph. LeFloch, *Boundary conditions for nonlinear hyperbolic systems of conservation laws*, J. Differ. Equ. 71 (1988), no.1, pp.93–122.

[13] S.N. Kruzhkov, *First order quasilinear equations with several independent variables*, Mat. Sb. 81(123) (1970), pp.228–255.

[14] E.Yu. Panov, *Existence and strong pre-compactness properties for entropy solutions of a first-order quasilinear equation with discontinuous flux*, Arch. Ration. Mech. Anal. 195 (2010), no.2, pp.643–673.

[15] K. Sbihi, *Study of some nonlinear PDEs in L^1 with general boundary conditions* (French, English). PhD Thesis, Strasbourg, France (2006), *http://tel.archives-ouvertes.fr/tel-00110417*

E-mail address: boris.andreianov@univ-fcomte.fr

THE RIEMANN PROBLEM FOR KERR EQUATIONS AND NON-UNIQUENESS OF SELFSIMILAR ENTROPY SOLUTIONS

Denise Aregba-Driollet

Institut de Mathématiques de Bordeaux, UMR 5251
IPB, Univ. Bordeaux, 351 cours de la Libération
33405 Talence, France

Abstract. We solve the Riemann problem for a nonlinear full wave Maxwell system arising in nonlinear optics. This system is hyperbolic, some eigenvalues have non-constant multiplicity and are neither genuinely nonlinear, nor linearly degenerate. In a particular 2×2 reduced case, we are able to exhibit two distinct selfsimilar entropy solutions. We compute the amounts of entropy dissipation and compare them.

1. **Introduction.** In nonlinear optics, the propagation of electromagnetic waves in a crystal can be modelized by the so-called Kerr model, which consists of Maxwell's equations

$$\begin{cases} \partial_t D - \text{curl} H = 0, \\ \partial_t B + \text{curl} E = 0, \end{cases}$$

with $\text{div} D = \text{div} B = 0$ and the constitutive relations

$$\begin{cases} B &= \mu_0 H \\ D &= \mathbf{D}(E) = \varepsilon_0 (1 + \varepsilon_r |E|^2) E. \end{cases}$$

Here μ_0, ε_0 are the free space permeability and permittivity and ε_r is the relative permittivity, see [12] for further details.

The model is a 6×6 system of conservation laws in the unknown $u = (D, H)$:

$$\begin{cases} \partial_t D - \text{curl} H = 0, \\ \partial_t H + \mu_0^{-1} \text{curl}(\mathbf{P}(D)) = 0 \end{cases} \tag{1}$$

where $\mathbf{P}$ is the reciprocal function of $\mathbf{D}$. Denoting

$$q(e) = \varepsilon_0 (e + \varepsilon_r e^3), \qquad e \in \mathbb{R}, \qquad p = q^{-1},$$

we have

$$E = \mathbf{P}(D) = \frac{D}{\varepsilon_0 (1 + \varepsilon_r p^2(|D|))}\,.$$

As proposed in [6] we also introduce the one dimensional model satisfied by solutions $D(x, t) = (0, d(x, t), 0)$, $H(x, t) = (0, 0, h(x, t))$ and $x = x_1 \in \mathbb{R}$. In that framework the solutions of Kerr model 1 satisfy the following p-system:

$$\begin{cases} \partial_t d + \partial_x h &= 0, \\ \partial_t h + \mu_0^{-1} \partial_x p(d) &= 0. \end{cases} \tag{2}$$

2000 *Mathematics Subject Classification.* Primary: 35L65; Secondary: 35A02, 78A60.

Key words and phrases. Systems of conservation laws, Riemann problem, Lax entropy solutions, Liu's criteria, nonlinear Maxwell's equations.

As $p' > 0$ it is strictly hyperbolic but the properties of the function p differ from the ones which appear in the general framework of gas dynamics or viscoelasticity. Here:

$$p(0) = 0, \quad p' > 0,$$

and p is strictly convex on $]-\infty, 0]$, strictly concave on $[0, +\infty[$.

Known existence results for system 1 are related to strong solutions, see [10], [7] and references therein. A first insight into weak solutions, which is also useful in the aim of designing numerical schemes, is the study of the Riemann problem: a direction $\omega \in \mathbb{R}^3$, $|\omega| = 1$, and $u_\pm \in \mathbb{R}^6$ being fixed, one looks for the solution of system 1 with initial data

$$u(x,0) = \left| \begin{array}{ll} u_- & \text{if } x \cdot \omega < 0, \\ u_+ & \text{if } x \cdot \omega > 0. \end{array} \right. \tag{3}$$

This work is devoted to the resolution of this problem and to the link between the solutions for the full model 1 and the ones for the reduced system 2. We point out the fact that we do not suppose that the initial data are divergence free (*ie* that $(D_+ - D_-) \cdot \omega = 0$ and $(H_+ - H_-) \cdot \omega = 0$) because in numerical applications, this condition is not exactly satisfied in general.

The electromagnetic energy is a mathematical entropy [6]. In the reduced 2×2 case it reduces to the classical entropy of the p-system.

The characteristic fields of system 1 have been described in [3], the following proposition summarizes the results:

Proposition 1. [3] *The Kerr system 1 is hyperbolic diagonalizable: for all $\omega \in \mathbb{R}^3$, $|\omega| = 1$, the eigenvalues are given by*

$$\lambda_1 \leq \lambda_2 = -\lambda < \lambda_3 = \lambda_4 = 0 < \lambda_5 = \lambda \leq \lambda_6 = -\lambda_1 \tag{4}$$

where, denoting $c = (\varepsilon_0 \mu_0)^{-\frac{1}{2}}$ the light velocity:

$$\lambda_1^2 = \frac{c^2}{1 + \varepsilon_r |E|^2}, \quad \lambda^2 = c^2 \frac{1 + \varepsilon_r (|E|^2 + 2(E \cdot \omega)^2)}{(1 + \varepsilon_r |E|^2)(1 + 3\varepsilon_r |E|^2)}. \tag{5}$$

and the inequalities in 4 are strict if and only if $\omega \times D \neq 0$.

The characteristic fields 1,3,4,6 are linearly degenerate.

The characteristic fields 2 and 5 are genuinely nonlinear in the open set

$$\Omega(\omega) = \{(D, H) \in \mathbb{R}^6 \; ; \; \omega \times D \neq 0\}.$$

The system is not strictly hyperbolic and the second and fifth characteristic fields are neither genuinely nonlinear, nor linearly degenerate. Hence, Lax' wellknown result [8] does not apply. Here for $|u_+ - u_-|$ small enough, we construct a "Lax' solution" and prove that such a construction is unique.

In the 2×2 case, the system 2 is strictly hyperbolic but similarly to the 6×6 case, the characteristic fields are genuinely nonlinear only in the domain $\Omega_1 = \{(d, h) \in \mathbb{R}^2 \; ; \; d \neq 0\}$. Nevertheless in that case we can construct the solution by using multiple waves, following Wendroff [13] and Liu [9].

System 2 being a particular case of 1, we find Lax' solutions of 1 which are also weak solutions of 2, but they are different from the "Liu's solutions". Moreover, we shall prove that the electromagnetic energy is dissipated by both solutions, so that there exists (at least) two selfsimilar entropy solutions of the Riemann problem for system 2.

A first study of the Riemann problem can be found in [4]: the problem is solved for a reduced 4×4 system and it is assumed that $D \cdot \omega$ is identically zero. Related numerical schemes are constructed in 1D and 2D transverse electric configurations.

In [3] we studied Kerr shocks and related shock profiles provided by the Kerr-Debye model, which is a hyperbolic quasilinear relaxation approximation of Kerr model. We studied Lax and Liu's admissibility criteria and proved that only Lax shocks give rise to Kerr-Debye shock profiles.

In [7], Godunov's scheme, which requires the solution of the Riemann problem, was implemented for a a two-dimensionnal transverse electric configuration. Actually this case reduces to the one of the p-system 2. Liu's solution was implemented. The results were found to coincide with the ones obtained by a Kerr-Debye relaxation scheme.

The plan of the paper is the following. In section 2 we determine the simple waves and the wave functions. In section 3 we construct the solution of the Riemann problem. Section 4 is devoted to the 2×2 case.

2. Wave functions.

If $u_- \neq u_+$ are connected by a k-Lax shock or a k-rarefaction wave or a k-contact discontinuity, u_- and u_+ are said to be connected by a k-wave. A plane discontinuity σ, u_+, u_- is a weak solution u of 1 such that $u(x,t) = u_-$ if $x \cdot \omega < \sigma t$, $u(x,t) = u_+$ else. All the plane discontinuities have been studied in [3]. The centred rarefaction waves are computed in [1]. The results are the following.

Proposition 2. Contact discontinuities. *Stationary contact discontinuities are characterized by*

$$\omega \times [H] = 0, \quad \omega \times [E] = 0. \tag{6}$$

The divergence free ones are constant.

A discontinuity σ, u_+, u_- is a contact discontinuity associated to λ_1 or λ_6 if and only if

$$\begin{cases} |E_+| = |E_-| \\ \sigma^2 = c^2(1+\epsilon_r|E_+|^2)^{-1} = c^2(1+\epsilon_r|E_-|^2)^{-1} \end{cases}$$

and

$$\begin{cases} \omega \cdot [D] = 0 \\ [H] = \sigma\omega \times [D]. \end{cases}$$

Moreover the only discontinuities satisfying Rankine-Hugoniot conditions and such that $|E_-| = |E_+|$ are the above contact discontinuities.

The shocks and rarefactions are related to the second and sixth characteristic fields. We recall that a discontinuity σ, u_-, u_+ is a Lax' k-shock if

$$\lambda_k(u_+) < \sigma < \lambda_{k+1}(u_+), \quad \lambda_{k-1}(u_-) < \sigma < \lambda_k(u_-).$$

Those inequalities are entropy conditions and also ensure that one can construct the solution of the Riemann problem as a superposition of simple waves. In our case, defining

$$f(d,d_0) = \frac{c^2\, d}{1+\epsilon_r p^2\left(\sqrt{d_0^2+d^2}\right)}, \qquad (d,d_0) \in \mathbb{R}^2,$$

we express the shocks by using the function S defined as

$$S(d_1,d_2,d_0) = \left((f(d_2,d_0) - f(d_1,d_0))(d_2-d_1)\right)^{\frac{1}{2}}.$$

The rarefaction waves are obtained *via* the integral curves of the eigenvectors of the system. ζ being a unitary vector orthogonal to ω, we use the function R_ζ defined by

$$R_\zeta(d_1, d_2, d_0) = \int_{d_1}^{d_2} \lambda(d_0\omega + s\zeta)ds, \quad d_1 \leq d_2\,, \quad d_0 \in \mathbb{R}.$$

Finally let ϕ_ζ be the function defined for $d_1 \geq 0$, $d_2 \geq 0$ and $d_0 \in \mathbb{R}$ by

$$\phi_\zeta(d_1, d_2, d_0) = \begin{cases} S(d_2, d_1, d_0) & if \quad d_1 \geq d_2\,, \\ -R_\zeta(d_1, d_2, d_0) & if \quad d_1 < d_2\,. \end{cases}$$

Proposition 3. *ϕ_ζ is a decreasing C^1 function with respect to d_2 and for all $d \geq 0$, $d_0 \in \mathbb{R}$:*

$$\phi_\zeta(d, 0, d_0) = \frac{cd}{\sqrt{1 + \epsilon_r p^2(\sqrt{d_0^2 + d^2})}}\,, \qquad \lim_{d_2 \to +\infty} \phi_\zeta(d, d_2, d_0) = -\infty.$$

The 2 and 5 waves are characterized as follows, see [1], [3] for the proof:

Proposition 4. *If $u_- \neq u_+$ are connected by a 2 or a 5 wave, then $D_- \neq D_+$ and $D_- \cdot \omega = D_+ \cdot \omega$. Moreover $\omega \times (\omega \times D_-)$ and $\omega \times (\omega \times D_+)$ are colinear.*

Reciprocally, let us consider u_- and u_+ such that $D_- \neq D_+$ and $D_- \cdot \omega = D_+ \cdot \omega = d_0$. If $\omega \times D_+ \neq 0$, we set $\overline{D} = D_+$. Else, $\omega \times D_- \neq 0$ and we set $\overline{D} = D_-$. We define ζ by

$$\zeta = -\frac{\omega \times (\omega \times \overline{D})}{|\omega \times (\omega \times \overline{D})|}\,.$$

u_- and u_+ are connected by a 2-wave if there exist two distinct nonnegative real numbers d_-, d_+ such that

$$D_\pm = d_0\omega + d_\pm\zeta, \quad H_+ = H_- + \phi_\zeta(d_-, d_+, d_0)\omega \times \zeta\,.$$

u_- and u_+ are connected by a 5-wave if there exist two distinct nonnegative real numbers d_-, d_+ such that

$$D_\pm = d_0\omega + d_\pm\zeta, \quad H_+ = H_- + \phi_\zeta(d_+, d_-, d_0)\omega \times \zeta\,.$$

3. **Solution of the full wave Riemann problem.** Suppose that $u_\pm = (D_\pm, H_\pm)$ and $\omega \in R^3$, $|\omega| = 1$, are given. We look for intermediate states u_1, u_*, u_{**}, u_2 such that:

- u_- and u_1 are connected by a 1-contact discontinuity,
- u_1 and u_* are connected by a 2-wave,
- u_* and u_{**} are connected by a stationary contact discontinuity,
- u_{**} and u_2 are connected by a 5-wave,
- u_2 and u_+ are connected by a 6-contact discontinuity.

In the following we shall denote $d_0^\pm = D_\pm \cdot \omega$.

Suppose that a solution exists. For the 1 and 6 contact discontinuities, the following conditions have to be fulfilled:

$$\begin{cases} D_1 \cdot \omega = D_- \cdot \omega = d_0^-\,, & |D_1| = |D_-|, \\ D_2 \cdot \omega = D_+ \cdot \omega = d_0^+, & |D_2| = |D_+|, \end{cases} \tag{7}$$

$$\begin{cases} H_1 - H_- = \sigma_-\,\omega \times (D_1 - D_-), \\ H_+ - H_2 = \sigma_+\,\omega \times (D_+ - D_2). \end{cases} \tag{8}$$

where

$$\sigma_\pm = \pm c\,(1 + \epsilon_r|E_\pm|)^{-\frac{1}{2}}\,.$$

For the 2 and 5 waves we know that D_1, D_*, ω are coplanar and D_2, D_{**}, ω are coplanar. Moreover $[D]\cdot\omega = 0$. There exist unitary vectors ζ_1, ζ_2, orthogonal to ω such that

$$D_1 = d_0^- \omega + d_1\zeta_1, \quad D_* = d_0^- \omega + d_*\zeta_1$$

and

$$D_2 = d_0^+ \omega + d_2\zeta_2, \quad D_{**} = d_0^+ \omega + d_{**}\zeta_2$$

and d_1, d_*, d_{**}, d_2 are non negative.

The stationary contact discontinuity is defined by conditions 6. One has

$$E_* = e_0^* \omega + e_*\zeta_1, \quad E_{**} = e_0^{**}\omega + e_{**}\zeta_2\,,$$

where

$$e_* = \frac{d_*}{\epsilon_0(1+\epsilon_r p^2(|D_*|))}, \quad e_{**} = \frac{d_{**}}{\epsilon_0(1+\epsilon_r p^2(|D_{**}|))}\,.$$

Therefore $e_*\,\omega\times\zeta_1 = e_{**}\,\omega\times\zeta_2$. Hence either $e_* = e_{**} = 0$ or those quantities are both positive and $\zeta_1 = \zeta_2$. The first case occurs if and only if $\omega\times D_* = \omega\times D_{**} = 0$. In the second case we have $e_* = e_{**}$, which also reads as

$$f(d_*, d_0^-) = f(d_{**}, d_0^+). \tag{9}$$

First case: $\omega\times D_* = \omega\times D_{**} = 0$. In that case, $D_* = d_0^-\omega$, $D_{**} = d_0^+\omega$. u_1 and u_* are the left and right states of a 2-shock propagating with speed

$$\sigma_2 = -\sqrt{\frac{f(d_1,d_0^-) - f(0,d_0^-)}{d_1}} = \sigma_-\,.$$

In the same way, u_{**} and u_2 are the left and right states of a 5-shock propagating with speed σ_+. Consequently the contact discontinuities merge with the shocks. Let us denote

$$V = \omega\times(H_+ - H_- - \omega\times(\sigma_+D_+ - \sigma_-D_-))\,. \tag{10}$$

Conditions 6, 8 and Rankine-Hugoniot conditions on the shocks imply that $V=0$ and

$$H_* = H_- - \omega\times\sigma_-D_-\,, \quad H_{**} = H_+ - \omega\times\sigma_-D_+\,. \tag{11}$$

If $D_-\times\omega = 0$ then $u_- = u_*$. Else u_- and u_* are connected by a 2-Lax shock.

In the same way, if $D_+\times\omega = 0$ then $u_+ = u_{**}$, else u_+ and u_{**} are connected by a 5-Lax shock.

Second case: $D_*\times\omega \neq 0$ and $D_{**}\times\omega\neq 0$. In this case, $\zeta_1 = \zeta_2 = \zeta$ and

$$D_1 = d_0^-\omega + d_1\zeta, \quad D_2 = d_0^+\omega + d_2\zeta, \tag{12}$$

$$D_* = d_0^-\omega + d_*\zeta, \quad D_{**} = d_0^+\omega + d_{**}\zeta, \tag{13}$$

with $d_1\geq 0$, $d_*>0$, $d_{**}>0$, $d_2\geq 0$. Let us denote

$$d = D\cdot\zeta, \quad h = H\cdot(\omega\times\zeta).$$

By 7-8:

$$\begin{cases} d_1 = |\omega\times(\omega\times D_-)|, & d_2 = |\omega\times(\omega\times D_+)|, \\ h_1 = h_- + \sigma_-(d_1 - d_-) & h_2 = h_+ + \sigma_+(d_2 - d_+). \end{cases} \tag{14}$$

As $u_*, u_{**}\in\Omega(\omega)$, one can define the 2 and 5-wave curves:

$$H_* - H_1 = \phi_\zeta(d_1, d_*, d_0^-)\omega\times\zeta\,, \quad H_2 - H_{**} = \phi_\zeta(d_2, d_{**}, d_0^+)\omega\times\zeta\,. \tag{15}$$

By 6, $h_* = h_{**}$ and

$$h_* = h_1 + \phi_\zeta(d_1, d_*, d_0^-) = h_2 - \phi_\zeta(d_2, d_{**}, d_0^+).$$

Therefore, using 9, we see that d_* and d_{**} are solution of the two by two system

$$\left\{ \begin{array}{l} f(d_*, d_0^-) = f(d_{**}, d_0^+)\,, \\ \\ h_1 + \phi_\zeta(d_1, d_*, d_0^-) = h_2 - \phi_\zeta(d_2, d_{**}, d_0^+). \end{array} \right. \tag{16}$$

As ϕ_ζ is decreasing and d_*, d_{**} are positive:

$$\phi_\zeta(d_1, d_*, d_0) + \phi_\zeta(d_2, d_{**}, d_0) < \phi_\zeta(d_1, 0, d_0^-) + \phi_\zeta(d_2, 0, d_0^+) = \sigma_+ d_2 - \sigma_- d_1.$$

This inequality is useful to determine ζ. As a matter of fact, using 8, we have also

$$\left\{ \begin{array}{l} H_* = H_- + \sigma_- \omega \times (D_1 - D_-) + \phi_\zeta(d_1, d_*, d_0^-)\omega \times \zeta, \\ H_{**} = H_+ - \sigma_+ \omega \times (D_+ - D_2) - \phi_\zeta(d_2, d_{**}, d_0^+)\omega \times \zeta\,. \end{array} \right.$$

Again by 6:

$$V = \left(\sigma_+ d_2 - \sigma_- d_1 - \phi_\zeta(d_1, d_*, d_0^-) - \phi_\zeta(d_2, d_{**}, d_0^+)\right) \zeta.$$

Therefore $V \neq 0$ and

$$\zeta = \frac{V}{|V|}. \tag{17}$$

Those results can be summarized as follows.

Proposition 5. *Consider u_-, u_+ such that the Riemann problem for system 1 has a solution which is a superposition of simple waves. Let V be the vector defined in 10. Only the following two cases occur:*

1) $V = 0$, u_- and u_* are connected by a 2-Lax shock propagating with velocity σ_-, u_+ and u_{**} are connected by a 5-Lax shock propagating with velocity σ_+, $D_* = (D_- \cdot \omega)\omega$, $D_{**} = (D_+ \cdot \omega)\omega$, H_* and H_{**} are given by 11.

2) $V \neq 0$, ζ is defined by 17, u_1 and u_2 are determined by conditions 12, 14 and u_*, u_{**} are determined by 13, 15 and the solution of system 16.

Sufficient conditions are as follows.

Theorem 3.1. *Let u_-, u_+ be a Riemann data for system 1 in the direction ω. There exists $\eta > 0$ such that if $|(D_- - D_+) \cdot \omega| < \eta$ then the Riemann problem has a unique solution in the class of the functions which are superpositions of simple waves. Let V be the vector defined in 10.*

If $V = 0$, then the solution is the superposition of a 2-Lax' shock, a stationary contact discontinuity and a 5-Lax' shock.

If $V \neq 0$, then the solution is the superposition of a 1-contact discontinuity, a 2-wave (Lax' shock or rarefaction), a stationary contact discontinuity, a 5-wave (Lax' shock or rarefaction) and a 6-contact discontinuity.

Moreover we can construct the solution in every case.

To prove this theorem, we remark that if $d_0^+ = d_0^-$, then system 16 can be solved. For the general case we use the implicit function theorem. We refer to [1] for the detailed proof.

4. **Non-uniqueness of selfsimilar entropy solutions.** Here we consider the Riemann problem 1-3 with $\omega = (1,0,0)^T$, $D_\pm = (0, D_{2,\pm}, 0)$, $H_\pm = (0,0,H_{3,\pm})$. As $(D_+ - D_-)\cdot\omega = (H_+ - H_-)\cdot\omega = 0$, the stationary contact discontinuity is trivial. We have $V = (0,v,0)$ and if $v \neq 0$ then $\zeta = (0, \mathrm{sgn}(v), 0)$. Hence Lax' solution is of the form $(0, D_2, 0)$, $(0,0,H_3)$ and (D_2, H_3) is a weak solution of the Riemann problem for the p-system 2.

Reciprocally, weak solutions (d,h) of 2 give solutions $(0,d,0)$, $(0,0,h)$ of 1. Following [9] (see also [13]) one can compute a weak solution of the Riemann problem for the p-system by using multiple waves. In that case the shocks satisfy Liu's (E) condition, which generalize Lax' conditions when p has inflexion points. This solution differs from the first one.

As an example, we computed both solutions in a particular case. The results for a fixed time t are depicted in Figure 1. The Lax' solution of 1-3 consists of a 1-contact discontinuity, a 2-shock, a 5-shock and a 6-contact discontinuity. Liu's solution consists of a shock propagating to the left and a shock propagating to the right. Observe that in that case the sign of d can change through the shock, while Lax' shock condition applied to 1 imply that the sign of $d = D_2$ is constant.

To be more precise, we denote $-\mu_1 = \mu_2 = \sqrt{\mu_0^{-1} p'(d)}$ the eigenvalues of 2. Let (d_-, h_-) a fixed left state. The set of right states (d_+, h_+) connected to (d_-, h_-) by a Liu's one-shock is parametrized by $d \in [d_-, d_*(d_-)]$ if $d_- < 0$, by $d \in [d_*(d_-), d_-]$ if $d_- > 0$, where d_* is defined by

$$d_*(d) = q(-\frac{1}{2}p(d)).$$

Symmetrically, the set of left states (d_-, h_-) connected to a given right state (d_+, h_+) by a Liu's two-shock is parametrized by $d \in [d_+, d_*(d_+)]$ if $d_+ < 0$, and by $d \in [d_*(d_+), d_+]$ if $d_+ > 0$.

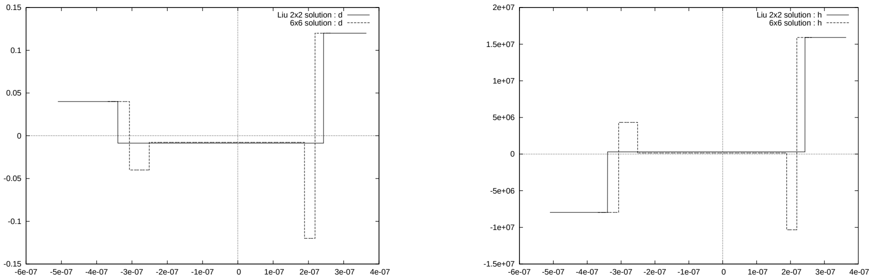


FIGURE 1. Two solutions of the p-system for the same initial data. Left: d, right: h.

At this point, we have two distinct solutions of the Riemann problem for 2, or equivalently for 1. Moreover, we can prove that both dissipate the physical entropy, namely the electromagnetic energy. For those solutions, this entropy is just the wellknown entropy of the p-system:

$$\eta(d,h) = E(d) + \frac{1}{2}\mu_0 h^2, \quad E(d) = \frac{1}{2}\epsilon_0(e^2 + \frac{3\epsilon_r}{2}e^4), \quad e = p(d)$$

with entropy flux $Q(d,h) = eh$. Contact discontinuities and rarefactions conserve the entropy. For the shocks a straightforward calculation leads to the following, see [1] for details.

Proposition 6. Entropy dissipation for the 2×2 system

- *Lax' and Liu's shocks satisfy the entropy dissipation property:*
$$[Q(d,h)] - \sigma[\eta(d,h)] = -\frac{\epsilon_0 \epsilon_r}{4} \sigma [e]^2 \; [e^2] \leq 0.$$
- *Let (d_-, h_-) be a fixed left state for a Liu's 1-shock. The entropy dissipation rate is a decreasing function of $|d_+ - d_-|$ over the interval $[0, |d^*(d_-) - d_-|]$.*
- *Let (d_+, h_+) be a fixed right state for a Liu's 2-shock. The entropy dissipation rate is a decreasing function of $|d_+ - d_-|$ over the interval $[0, |d^*(d_+) - d_+|]$.*

We have solved the Riemann problem for the full wave Kerr system and we have proved the non-uniqueness of selfsimilar entropy solutions. Which is the physical solution? Liu's solution is more dissipative than the other one. 1D and 2D numerical experiments with a physically relevant relaxation model, the Kerr-Debye system, lead to Liu's solutions see [2], [7], but this is possibly due to numerical viscosity: it is wellknown that contact discontinuities are not easy to catch numerically. On another hand the full wave model should be the more realistic one. Further investigations are in progress.

REFERENCES

[1] D. Aregba-Driollet, *Godunov schemes for Kerr equations*, in preparation.

[2] D. Aregba-Driollet and C. Berthon, *Numerical approximation of Kerr-Debye equations*, preprint 2009.

[3] D. Aregba-Driollet and B. Hanouzet, *Kerr-Debye relaxation shock profiles for Kerr equations*, Commun. Math. Sci., **9** (2011), 1-31.

[4] A. de La Bourdonnaye, *High-order scheme for a nonlinear Maxwell system modelling Kerr effect*, J. Comput. Phys., **160** (2000), 500–521.

[5] A. Bressan, "Hyperbolic systems of conservation laws. The one-dimensional Cauchy problem," Oxford Lecture Series in Mathematics and its Applications, 20, Oxford University Press, Oxford, 2000.

[6] G. Carbou and B. Hanouzet. *Relaxation approximation of some nonlinear Maxwell initial-boundary value problem*, Commun. Math. Sci., **4** (2006), 331–344.

[7] M. Kanso, "Sur le modèle de Kerr-Debye pour la propagation des ondes électromagnétiques," Ph.D thesis, Université Bordeaux 1, 2012.

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

[9] T.-P. Liu, *The Riemann problem for general 2×2 conservation laws*, Trans. Amer. Math. Soc., **199** (1974), 89–112.

[10] R. Racke. "Lectures on nonlinear evolution equations. Initial value problems," Aspects of Mathematics, E19. Friedr. Vieweg and Sohn, Braunschweig, 1992.

[11] D. Serre "Systèmes de lois de conservation I. and II.," Diderot, Paris, 1996. Cambridge University Press, Cambridge, 1999 for the english translation ("Systems of conservation laws I. and II.")

[12] Y.-R. Shen, "The Principles of Nonlinear Optics," Wiley Interscience, 1994.

[13] B. Wendroff, *The Riemann problem for materials with nonconvex equations of state. I. Isentropic flow*, J. Math. Anal. Appl., **38** (1972), 454–466.

E-mail address: `aregba@math.u-bordeaux1.fr`

PENALTY METHODS FOR EDGE PLASMA TRANSPORT IN A TOKAMAK

Thomas Auphan, Philippe Angot and Olivier Guès

Aix Marseille Université, CNRS, Centrale Marseille, LATP
UMR 7353, 13453 Marseille, France

Abstract. The volume penalty method belongs to the immersed boundary methods, which is used for the numerical simulation of boundaries in PDE problems. This paper focuses on the mathematical aspect of penalization for quasilinear hyperbolic problems with non characteristic boundary. A penalty method which does not generate any boundary layer is proposed, with an application to edge plasma transport for a tokamak.

1. **Introduction.** The solution of partial differential equation using approximate numerical scheme is now a very common game for scientists, useful for a wide range of areas such as fluid mechanics. One issue is the treatment of the boundary conditions. The usual way to deal with the boundary conditions is to use a body-fitted mesh. The volume penalization methods give another approach: the physical domain, also called the original domain, is included in a larger domain with a simple shape. The system of equations is then extended to the fictitious domain so that, at the boundary of the original domain, the conditions are approximately recovered.

The main advantages of penalty methods is the possibility to use non body-fitted meshes and efficient solvers such as pseudo spectral methods [8]. One problem is the error added by the penalization and, in some case, the generation of a boundary layer next to the interface [4, 5]. In previous works [1, 2], a numerical comparison between two numerical methods has been provided. This paper proposes a more theoretical study of the penalization for more general quasilinear hyperbolic problems.

This is the first theoretical result about the penalization of a general quasilinear hyperbolic problem with non characteristic and maximally strictly dissipative boundary conditions. The second section gives an example of application of this penalization, whereas the third one contains the theoretical result. The last section proposes an idea of extension in the case of an obstacle with two opposites sides in contact with the original domain.

2. **An example from plasma fusion.** This section shows quickly how a penalty method can be applied for the numerical simulation of the edge plasma transport in a tokamak, for more details see [2]. A tokamak is an apparatus to study plasma and the fusion reaction. The goal is to perform the nuclear fusion reaction by magnetic confinement. For this civil application, the fusion by magnetic confinement is the

2000 *Mathematics Subject Classification.* Primary: 35L60, 65N85; Secondary: 65N15.

Key words and phrases. Nonlinear, hyperbolic system, boundary value problem, penalization method, plasma transport.

The authors are supported by the ANR ESPOIR and the FR-FCM.

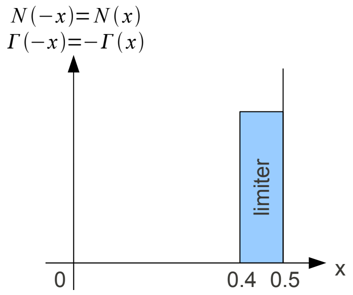


FIGURE 1. Representation of the computational domain.

most advanced technology, compared to inertial fusion. The wall of the tokamak can have a complicated shape, with obstacles such as the limiter.

The plasma next to the wall, is called edge plasma, can be modeled by the fluid approximation. Due to the strong magnetic field, the plasma transport occurs essentially along the magnetic field lines (here, the curvilinear coordinate x). The area next to the wall of the tokamak where the magnetic field lines are interrupted by the limiter is called the *scrape-off* layer: this is the main interest of this section. The boundary condition are given by the Bohm criterion. The toy model considered is composed of two equations: the mass conservation and the momentum conservation. N stands for the plasma density, Γ the plasma momentum and $M = \Gamma/N$ the Mach number and all the quantities are dimensionless. The system writes ($M_0 = 1 - \eta$, with $\eta << 1$):

$$\left\{\begin{array}{l} \partial_t N + \partial_x \Gamma = S_N \\ \partial_t \Gamma + \partial_x \left(\dfrac{\Gamma^2}{N} + N \right) = S_\Gamma \\ \begin{pmatrix} M_0 & -1 \end{pmatrix} \begin{pmatrix} N(t,0) \\ \Gamma(t,0) \end{pmatrix} = 0 \end{array}\right.$$

Notice that the system is very similar to shallow water equations. The penalization appears naturally with the new unknowns expressed below, because the Dirichlet boundary condition becomes homogeneous:

$$\tilde{u}(t,\mathbf{x}) = \ln\left(N(t,\mathbf{x})\right)$$

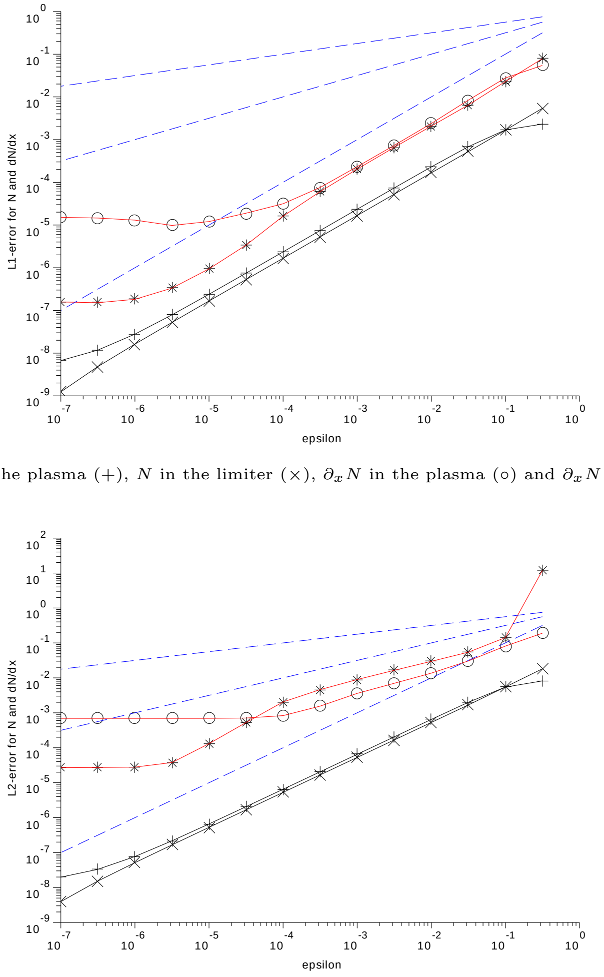
$$\tilde{v}(t,\mathbf{x}) = \frac{\Gamma(t,\mathbf{x})}{N(t,\mathbf{x})} - M_0$$

Hence, only $\tilde{v}$ is affected by the boundary condition.

Finally, the penalization obtained thanks to the results presented in the section 3 is:

$$\left\{\begin{array}{l} \partial_t N + \partial_x \Gamma = S_N \\ \partial_t \Gamma + \partial_x \left(\dfrac{\Gamma^2}{N} + N \right) + \frac{\chi}{\varepsilon} \left(\dfrac{\Gamma}{M_0} - N \right) = S_\Gamma \end{array}\right.$$

Where χ is the characteristic function of the limiter, *i.e.* $\chi = 1$ in the limiter and $\chi = 0$ elsewhere. According to the Figure 2, the main advantage of this method is the absence of spurious boundary layer: the error due to the penalization decreases with an optimal rate when the penalization parameter ε tends to 0. The main drawback is due to the fact that only the incoming fields of the hyperbolic system are penalized. So, at the boundary of the computational domain, it is necessary to



L^1 error for N in the plasma ($+$), N in the limiter ($\times$), $\partial_x N$ in the plasma ($\circ$) and $\partial_x N$ in the limiter ($*$)

L^2 error for N in the plasma ($+$), N in the limiter ($\times$), $\partial_x N$ in the plasma ($\circ$) and $\partial_x N$ in the limiter ($*$)

FIGURE 2. Errors for N and $\partial_x N$ in L^1 and L^2 norms with the boundary layer free penalization. The dashed lines represent the curves $\varepsilon^{1/4}, \varepsilon^{1/2}$ and ε. The computations were made using a VF Roe scheme with non conservative variables with extensions up to the second order (see [6]). The mesh step is $\delta x = 10^{-5}$ and the computational domain is described in the Figure 1. For more details, see [1, 2].

provide transparent boundary condition, which is not easy. Besides, non compatible initial boundary condition may generates artifact, see for instance, the L^2 error of the Figure 2 and the numerical results of [2]. The manufactured solution used for this example was (with $M_0 = 0.9$):

$$N(t,x) = \exp\left(\frac{-x^2}{0.16(t+1)}\right) \qquad \Gamma(t,x) = M_0 \sin\left(\frac{\pi x}{0.8}\right) \exp\left(\frac{-x^2}{0.16(t+1)}\right)$$

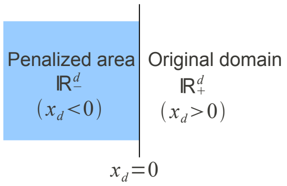


FIGURE 3. The space domain for the problem of the section 3

3. The penalty method for a quasilinear non characteristic hyperbolic problem.

This section provides a penalty method for general quasilinear non characteristic hyperbolic problem in a d-dimensional space. The issues of singularities and of the compatibility of the initial condition are not the point of this paper: so, the solution $\mathbf{u}$ of the problem considered is supposed to be null in the past and local in time. Let us now write the hyperbolic boundary value problem that is studied in this section:

$$\left\{\begin{array}{ll} \partial_t \mathbf{u}(t,\mathbf{x}) + \sum_{j=1}^{d} \bar{\mathbf{A}}_j(\mathbf{a}(t,\mathbf{x}),\mathbf{u}(t,\mathbf{x}))\partial_j \mathbf{u}(t,\mathbf{x}) = \bar{\mathbf{f}}(\mathbf{a}(t,\mathbf{x}),\mathbf{u}(t,\mathbf{x})) & \\ \qquad\qquad\qquad\qquad\qquad\qquad (t,\mathbf{x}) \in]-T_0,T[\times\mathbb{R}^d_+ & \\ \mathbf{\Theta}(\mathbf{a}(t,\mathbf{x}',0),\mathbf{u}(t,\mathbf{x}',0)) = \mathbf{0} \qquad (t,\mathbf{x}') \in]-T_0,T[\times\mathbb{R}^{d-1} & \\ \mathbf{u}_{|t<0} = \mathbf{0} & \end{array}\right. \tag{1}$$

In this paper, $\mathbf{x} = (x_1,\ldots,x_d) = (\mathbf{x}',x_d)$ stands for the space variable. The space domain is represented in the Figure 3.

To ensure the well posedness of the hyperbolic problem, the coefficients of (1) satisfies the following hypotheses:

1. $\mathbf{a} :]-T_0,T[\times\mathbb{R}^d \to \mathbb{R}^{N'}$ is in $H^\infty(]-T_0,T[\times\mathbb{R}^d)$.
2. $\bar{\mathbf{f}} : \mathbb{R}^{N'}\times\mathbb{R}^N \to \mathbb{R}^N$ is indefinitely differentiable and, for all $t<0, \mathbf{x}\in\mathbb{R}^d, \bar{\mathbf{f}}(\mathbf{a}(t,\mathbf{x}),\mathbf{0}) = \mathbf{0}$.
3. $\mathbf{\Theta} : \mathbb{R}^{N'}\times\mathbb{R}^N \to \mathbb{R}^p$ is indefinitely differentiable and for all $(\mathbf{y},\mathbf{U}) \in \mathbb{R}^{N'}\times\mathbb{R}^N$, $\nabla_{\mathbf{u}}\mathbf{\Theta}\,(\mathbf{y},\mathbf{U})$ has a constant rank p.
4. For all $j\in\{1,\ldots,d\}$, $\bar{\mathbf{A}}_j : \mathbb{R}^{N'}\times\mathbb{R}^N \to \mathcal{M}_N(\mathbb{R})$ is indefinitely differentiable.
5. There exists a symmetrizer $\mathbf{S}(\mathbf{y},\mathbf{U})$ such that, for all $(\mathbf{y},\mathbf{U}) \in \mathbb{R}^{N'}\times\mathbb{R}^N$:
   - $\mathbf{S}(\mathbf{y},\mathbf{U})$ symmetric and positive definite, uniformly in $(\mathbf{y},\mathbf{U})$ when $\mathbf{U}$ is in a neighborhood $\mathcal{U}\subset\mathbb{R}^N$ of $\mathbf{0}$ and $\mathbf{y}$ in a neighborhood $\mathcal{Z}\subset\mathbb{R}^{N'}$ of $\mathbf{0}$. This means that there exists $\bar{e}>0$ such that, for all $(\mathbf{y},\mathbf{U})\in\mathcal{Z}\times\mathcal{U}$, and for all $\mathbf{W}\in\mathbb{R}^N$, $\langle\mathbf{S}(\mathbf{y},\mathbf{U})\mathbf{W},\mathbf{W}\rangle \geq \bar{e}\|\mathbf{W}\|^2$, where $\langle,\rangle$ and $\|.\|$ are respectively the euclidean scalar product and norm on $\mathbb{R}^N$.
   - For all $j\in\{1,\ldots,d\}$, $\mathbf{S}(\mathbf{y},\mathbf{U})\bar{\mathbf{A}}_j(\mathbf{y},\mathbf{U})$ is symmetric.

The problem is supposed to be non characteristic, *i.e.* for all $(\mathbf{y},\mathbf{U})\in\mathcal{Z}\times\mathcal{U}$ such that $\mathbf{\Theta}(\mathbf{y},\mathbf{U})=\mathbf{0}$, the matrix $\bar{\mathbf{A}}_d(\mathbf{y},\mathbf{U})$ is invertible. The boundary conditions are maximally strictly dissipative: For all $\mathbf{y}\in\mathcal{Z}$, if there exists $\mathbf{U}\in\mathbb{R}^N$ such that $\mathbf{\Theta}(\mathbf{y},\mathbf{U})=\mathbf{0}$, the quadratic form have the following properties:

- $\exists\bar{\mu}>0, \forall\mathbf{y}\in\mathbb{R}^{N'}, \forall\mathbf{W}\in\ker\nabla_{\mathbf{u}}\mathbf{\Theta}(\mathbf{y},\mathbf{0}), \langle\mathbf{S}(\mathbf{y},\mathbf{U})\bar{\mathbf{A}}_d(\mathbf{y},\mathbf{U})\mathbf{W},\mathbf{W}\rangle \leq -\bar{\mu}\|\mathbf{W}\|^2$.
- Besides $\dim\ker\nabla_{\mathbf{u}}\mathbf{\Theta}(\mathbf{y},\mathbf{0})$ is maximal for the property above.

One can assert there exists a finite time $\theta>0$ such that the original problem (*cf.* equation (1)) admits a solution $\mathbf{u}$ in $H^\infty(]-T_0,\theta[\times\mathbb{R}^d)$. For a proof, see [7, 9].

To provide the penalty method, it is simpler to reformulate the problem in order to have an homogeneous Dirichlet boundary condition for the p first unknowns of the problem, and avoid the issue of the nonlinear boundary condition $\mathbf{\Theta}(\mathbf{a},\mathbf{u}) = \mathbf{0}$. The following lemma shows that such a suitable change of unknown exists.

Lemma 3.1. *There exists $\mathcal{Q} \subset \mathcal{U}, \mathcal{V}$, two neighborhoods of $\mathbf{0} \in \mathbb{R}^N$ and $\mathcal{Y} \subset \mathcal{Z}$ a neighborhood of $\mathbf{0} \in \mathbb{R}^{N'}$ satisfying: there exists $\mathbf{H} \in \mathcal{C}^\infty(\mathcal{Y} \times \mathcal{V}, \mathcal{Q})$ such that, for all $\mathbf{y} \in \mathcal{Y}$, $\mathbf{H}(,\mathbf{y},.)$ is a $\mathcal{C}^\infty$-diffeomorphism from $\mathcal{V}$ to $\mathcal{Q}$ and such that*

$$\forall \mathbf{U} \in \mathcal{Q}, \forall \mathbf{y} \in \mathcal{Y}, \mathbf{\Theta}(\mathbf{y},\mathbf{U}) = \mathbf{0} \Longleftrightarrow V_1 = V_2 = \cdots = V_p = 0$$

Where $\mathbf{V} \in \mathbb{R}^N$ is such that $\mathbf{U} = \mathbf{H}(\mathbf{y},\mathbf{V})$ and $(V_1,\ldots,V_N) = \mathbf{V}$.

Henceforth, the function $\mathbf{a}$ is assumed to be valued in the neighborhood $\mathcal{Y}$, *i.e.*, $\forall (t,\mathbf{x}) \in]-T_0,T[\times\mathbb{R}^d, \mathbf{a}(t,\mathbf{x}) \in \mathcal{Y}$.

In order to simplify the notations, the dependence of the functions and matrices on $(t,\mathbf{x})$ and $\mathbf{a}(t,\mathbf{x})$ is now implicit. So, for instance, $\partial_j\left(\bar{\mathbf{A}}_j(\mathbf{u})\right)$ represents

$$\nabla_{\mathbf{a}}\bar{\mathbf{A}}_j(\mathbf{a}(t,\mathbf{x}),\mathbf{u}(t,\mathbf{x}))\cdot\partial_j\mathbf{a}(t,\mathbf{x}) + \nabla_{\mathbf{u}}\bar{\mathbf{A}}_j(\mathbf{a}(t,\mathbf{x}),\mathbf{u}(t,\mathbf{x}))\cdot\partial_j\mathbf{u}(t,\mathbf{x}).$$

The matrix $\mathbf{P}$ is defined as the projection on the linear subspace $\mathbb{R}^p \times \{0\}^{N-p}$. With the unknown $\mathbf{v}$, the boundary condition becomes $\mathbf{P}\mathbf{v} = \mathbf{0}$. For the new unknown $\mathbf{v}$, the system writes (the parameter function $\mathbf{a}$ is understood):

$$\left\{\begin{array}{ll} \nabla_{\mathbf{v}}\mathbf{H}(\mathbf{v})\,\partial_t\mathbf{v}+\displaystyle\sum_{j=1}^{d}\bar{\mathbf{A}}_j\left(\mathbf{H}(\mathbf{v})\right)\nabla_{\mathbf{v}}\mathbf{H}(\mathbf{v})\partial_j\mathbf{v}=\bar{\mathbf{f}}\left(\mathbf{H}(\mathbf{v})\right) & \text{in }]-T_0,T[\times\mathbb{R}^d_+ \\ \mathbf{P}\mathbf{v}_{|x_d=0} = \mathbf{0} & \text{in }]-T_0,T[\times\mathbb{R}^{d-1} \end{array}\right. \tag{2}$$

The system is then multiplied on the left by $\nabla_{\mathbf{v}}\mathbf{H}(\mathbf{v})^\top\mathbf{S}\left(\mathbf{H}(\mathbf{v})\right)$:

$$\left\{\begin{array}{ll} \mathbf{A}_0(\mathbf{v})\,\partial_t\mathbf{v}+\sum_{j=1}^{d}\mathbf{A}_j(\mathbf{v})\partial_j\mathbf{v} = \mathbf{f}(\mathbf{v}) & \text{in }]-T_0,T[\times\mathbb{R}^d_+ \\ \mathbf{P}\mathbf{v}_{|x_d=0} = \mathbf{0} & \text{in }]-T_0,T[\times\mathbb{R}^{d-1} \\ \mathbf{v}_{|t<0} = \mathbf{0} & \text{in }]-T_0,0[\times\mathbb{R}^d_+ \end{array}\right. \tag{3}$$

In this new formulation, the functions $\mathbf{A}_j$ and $\mathbf{f}$ are:

$$\mathbf{A}_0(\mathbf{v}) = \nabla_{\mathbf{v}}\mathbf{H}(\mathbf{v})^\top\mathbf{S}\left(\mathbf{H}(\mathbf{v})\right)\nabla_{\mathbf{v}}\mathbf{H}(\mathbf{v})$$

$$\mathbf{A}_j(\mathbf{v}) = \nabla_{\mathbf{v}}\mathbf{H}(\mathbf{v})^\top\mathbf{S}\left(\mathbf{H}(\mathbf{v})\right)\bar{\mathbf{A}}_j(\mathbf{v})\nabla_{\mathbf{v}}\mathbf{H}(\mathbf{v})$$

$$\mathbf{f}(\mathbf{v}) = \nabla_{\mathbf{v}}\mathbf{H}(\mathbf{v})^\top\mathbf{S}\left(\mathbf{H}(\mathbf{v})\right)\left(\bar{\mathbf{f}}\left(\mathbf{H}(\mathbf{v})\right) - \nabla_{\mathbf{a}}\mathbf{H}(\mathbf{v})\cdot\partial_t\mathbf{a} - \sum_{j=1}^{d}\bar{\mathbf{A}}_j(\mathbf{v})\nabla_{\mathbf{a}}\mathbf{H}(\mathbf{v})\cdot\partial_j\mathbf{a}\right)$$

According to the properties on $\mathbf{S}\left(\mathbf{H}(\mathbf{v})\right)$ and $\nabla_{\mathbf{v}}\mathbf{H}(\mathbf{v})$, we can assert that $\mathbf{A}_0(\mathbf{y}, \mathbf{V})$, is uniformly positive definite regarding $(\mathbf{y},\mathbf{V})$, where $\mathbf{y} \in \mathcal{Y}$ and $\mathbf{V}$ such that $\mathbf{H}(\mathbf{y},\mathbf{V}) \in \mathcal{Q}$. Hence, there exists $e > 0$ (independent from $\mathbf{V}$) such that, for all $\mathbf{y} \in \mathcal{Y}$ and for all $\mathbf{W} \in \mathbb{R}^N$, $\langle\mathbf{A}_0(\mathbf{y},\mathbf{V})\mathbf{W},\mathbf{W}\rangle \geq e\|\mathbf{W}\|^2$.

Let us remind that the assumption of the maximally strictly dissipative boundary condition is invariant by the change of unknown. Besides, for the reformulated problem (3), the property of maximally strictly dissipative boundary conditions means: For all $\mathbf{V} \in \mathbb{R}^N$ such that $\mathbf{P}\mathbf{V} = \mathbf{0}$, the quadratic form have the following properties:

- $\exists \mu > 0, \forall \mathbf{W} \in \ker\mathbf{P}, \forall \mathbf{y} \in \mathcal{Y}, \langle\mathbf{A}_d(\mathbf{y},\mathbf{V})\mathbf{W},\mathbf{W}\rangle \leq -\mu\|\mathbf{W}\|^2$
- $N-p$ is the number of strictly negative eigenvalues of $\mathbf{A}_d(\mathbf{y},\mathbf{V})$ with multiplicity. Thus, with multiplicity, there are p strictly positive eigenvalues.

Let us now introduce the following penalized system, which is the main concern of the paper

$$\begin{cases} \mathbf{A}_0(\mathbf{v}_\varepsilon)\,\partial_t\mathbf{v}_\varepsilon + \sum_{j=1}^d \mathbf{A}_j(\mathbf{v}_\varepsilon)\partial_j\mathbf{v}_\varepsilon + \dfrac{\chi}{\varepsilon}\mathbf{P}\mathbf{v}_\varepsilon = \mathbf{f}(\mathbf{v}_\varepsilon) & \text{in }]-T_0,T[\times\mathbb{R}^d \\ \mathbf{v}_{\varepsilon\,|t<0} = \mathbf{0} & \text{in }]-T_0,0[\times\mathbb{R}^d \end{cases} \tag{4}$$

where $\chi = 0$ in $]-T_0,T[\times\mathbb{R}^d_+$ and $\chi = 1$ elsewhere.

Notice that the boundary condition of the reformulated problem (3) is $\mathbf{P}\mathbf{v}_{|x_d=0} = \mathbf{0}$ and the penalization term added in the penalized system (4) simply writes $\dfrac{\chi}{\varepsilon}\mathbf{P}\mathbf{v}_\varepsilon$. So, when ε tends to 0, from the formal point of view, one recovers the boundary condition $\mathbf{P}\mathbf{v}_{\varepsilon|x_d=0} \approx \mathbf{0}$. The main result of this paper is the theorem 3 (see below) which ensures that the penalized system (4) is well-posed and provides an estimation of the error due to the penalization.

Main Theorem. *Under the assumptions on the coefficients explained above, one can assert that there exists a finite time $T \leq \theta$ and $\varepsilon_0 > 0$ such that, for all $\varepsilon \in]0,\varepsilon_0]$, the penalized problem (4) has a unique solution $\mathbf{v}_\varepsilon \in H^1(]-T_0,T[\times\mathbb{R}^d) \cap W^{1,\infty}(]-T_0,T[\times\mathbb{R}^d)$.*

Besides, $\mathbf{v}_\varepsilon$ is smooth on each side of the interface $x_d = 0$, i.e.*, $\mathbf{v}_{\varepsilon|x_d>0} \in H^\infty(]-T_0,T[\times\mathbb{R}^d_+)$ and $\mathbf{v}_{\varepsilon|x_d<0} \in H^\infty(]-T_0,T[\times\mathbb{R}^d_-)$.*

Moreover, for all $s \in \mathbb{N}$, the following estimate holds as ε goes to 0:

$$\|\mathbf{v} - \mathbf{v}_\varepsilon\|_{H^s(]-T_0,T[\times\mathbb{R}^d_+)} = \mathcal{O}(\varepsilon)$$

This theorem gives a simple and incomplete penalization for the reformulated problem. One could give a formulation for the original problem (1) but this is not very convenient to use: it is easier to change the unknown so that the penalization matrix $\mathbf{P}$ appears naturally, and then return to the original independent variables.

The optimal convergence rate of the penalization error estimate $\|\mathbf{v} - \mathbf{v}_\varepsilon\|_{H^s(]-T_0,T[\times\mathbb{R}^d_+)}$ is an evidence of the absence of the absence of boundary layer. This the first theoretical result about a boundary-layer-free penalty method for general quasilinear hyperbolic value problem.

3.1. **Sketch of the proof of the main theorem.** The complete proof of the main theorem is written in [3]. In the proof, Ω_T^+ stands for the original domain $(]-T_0,T[\times\mathbb{R}^d_+)$ and Ω_T represents $]-T_0,T[\times\mathbb{R}^d$.

First, one provides an asymptotic expansion of the solution of the penalized problem (4) $\mathbf{v}_{a|x_d>0}(t,\mathbf{x}) = \sum_{n=0}^M \varepsilon^n \mathbf{V}^{n,+}(t,\mathbf{x})$ and $\mathbf{v}_{a|x_d<0} = \sum_{n=0}^M \varepsilon^n \mathbf{V}^{n,-}(t,\mathbf{x})$, where M is a sufficiently large integer. The absence of fast variables of the form ${}^{x_d}/_{\varepsilon^\alpha}$ in the terms $\mathbf{V}^{n,\pm}$ is a first evidence of the absence of boundary layer. Notice that $\mathbf{V}^{0,+}$ is the exact solution of the reformulated hyperbolic problem (3), *i.e.* without penalization. Another important point for the asymptotic expansion is the definition of the time T of the existence of the solution. Besides, T appears to be less or equal to θ which is the time of the existence of solution of the non penalized problem (3).

In order to obtain the exact solution, $\mathbf{v}_\varepsilon$, of (4), the following form is considered:

$$\mathbf{v}_\varepsilon = \mathbf{v}_a + \varepsilon\mathbf{w}$$

The goal is now to build $\mathbf{w}$, which is the solution of:

$$\begin{cases} \mathbf{A}_0(\mathbf{v}_a+\varepsilon\mathbf{w})\partial_t\mathbf{w}+\sum_{j=1}^d \mathbf{A}_j(\mathbf{v}_a+\varepsilon\mathbf{w})\partial_j\mathbf{w}-\mathbf{B}(\varepsilon\mathbf{w})\mathbf{w}+\frac{1}{\varepsilon}\chi\mathbf{P}\mathbf{w}=-\varepsilon^{M-1}\mathbf{R}_\varepsilon \\ \qquad\qquad\qquad (t,\mathbf{x})\in\Omega_T \\ \mathbf{w}_{|t<0}=\mathbf{0} \end{cases} \tag{5}$$

Where $\mathbf{R}_\varepsilon$ is a corrective term for the equation satisfied by $\mathbf{v}_a$ and $\mathbf{B}(\varepsilon\mathbf{w})\mathbf{w}$ satisfies:

$$(\mathbf{A}_0(\mathbf{v}_a+\varepsilon\mathbf{w})-\mathbf{A}_0(\mathbf{v}_a))\partial_t\mathbf{v}_a+\sum_{j=1}^d(\mathbf{A}_j(\mathbf{v}_a+\varepsilon\mathbf{w})-\mathbf{A}_j(\mathbf{v}_a))\partial_j\mathbf{v}_a=-\varepsilon\mathbf{B}(\mathbf{v}_a,\nabla\mathbf{v}_a,\varepsilon\mathbf{w})\mathbf{w}$$

To simplify the notations in the equations, $\mathbf{B}(\varepsilon\mathbf{w})$ stands for $\mathbf{B}(\mathbf{v}_a,\nabla\mathbf{v}_a,\varepsilon\mathbf{w})$.

To prove its existence, $\mathbf{w}$ is approximated by the sequence $(\mathbf{w}^k)_{k\in\mathbb{N}}$ generated by a Picard's iterative scheme:

$$\mathbf{w}^0=\mathbf{0}$$

$$\begin{cases} \mathbf{A}_0(\mathbf{v}_a+\varepsilon\mathbf{w}^k)\partial_t\mathbf{w}^{k+1}+\sum_{j=1}^d \mathbf{A}_j(\mathbf{v}_a+\varepsilon\mathbf{w}^k)\partial_j\mathbf{w}^{k+1}-\mathbf{B}(\varepsilon\mathbf{w}^k)\mathbf{w}^{k+1} \\ \qquad +\dfrac{\chi}{\varepsilon}\mathbf{P}\mathbf{w}^{k+1}=-\varepsilon^{M-1}\mathbf{R}_\varepsilon \qquad\qquad \text{in } \Omega_T \\ \mathbf{w}^{k+1}_{|t<0}=\mathbf{0} \end{cases}$$

This sequence is expected to converge toward $\mathbf{w}$ in $L^2(\Omega_T)$ and then in $H^\infty(\Omega_T^+)$, in $H^\infty(\Omega_T^-)$ and in $H^1(\Omega_T)$.

For any λ sufficiently large, it is possible to prove this energy estimate:

$$\|\mathbf{w}^{k+1}\exp(-\lambda t)\|_{L^2(\Omega_T)}+\frac{1}{\sqrt{\varepsilon}}\|\chi\mathbf{P}\mathbf{w}^{k+1}\exp(-\lambda t)\|_{L^2(\Omega_T)}\leq\frac{C(R)}{\sqrt{\lambda}}\varepsilon^M\|\mathbf{R}_\varepsilon\exp(-\lambda t)\|_{L^2(\Omega_T)}$$

Where $R>0$ is such that $\|\mathbf{w}^k\|_\infty+\|\nabla\mathbf{w}^k\|_\infty\leq R$ and $C(R)$ is a constant which does not depends on $\mathbf{w}^{k+1},\mathbf{w}^k,\lambda,\varepsilon$.

The next step is the extension of this estimate to the tangential derivatives of $\mathbf{w}^{k+1}$, which enables one to show that, for a sufficiently large fixed value of λ, $\mathbf{w}^{k+1}$ and $\nabla\mathbf{w}^{k+1}$ are bounded independently from $\mathbf{w}^k$. Then, the sequence $(\mathbf{w}^k)$ converges for the L^2 norm, toward $\mathbf{w}$, the solution of (5). Finally, the solution $\mathbf{v}_\varepsilon=\mathbf{v}_a+\varepsilon\mathbf{w}$ of the penalized hyperbolic problem (4) is obtained. The error estimate comes from the equality: $\mathbf{v}_{\varepsilon|x_d>0}-\mathbf{v}=\sum_{n=1}^M\varepsilon^n\mathbf{V}^{n,+}+\varepsilon\mathbf{w}$.

4. **Idea for the extension to a two sides obstacle.** In some cases, the obstacle has two opposites sides in contact with the original domain. Applying carelessly the penalization presented above can create interferences and singularities inside the limiter: to avoid such an issue, one can multiply the flux term by a smooth function $x_d\mapsto\alpha(x_d)$ which is null in an area at the center of the obstacle and equal to 1 in the original domain and in a neighborhood of the interface. The Figure 4 gives a possible obstacle and the shape of the function α.

As α is regular, the well posedness of the system is still guaranteed.

5. **Conclusion.** A general recipe for the penalization of a nonlinear hyperbolic problem is proposed with numerical tests for a one dimensional problem. This penalty method has two advantages:

- After the change of unknown, it is a natural penalization for the reformulated problem (3).
- The error due to the penalization, $\|\mathbf{v}-\mathbf{v}_{\varepsilon|x_d>0}\|_{H^s}$, has an optimal rate of convergence, *i.e.* $\mathcal{O}(\varepsilon)$.

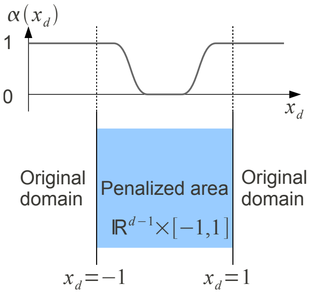


FIGURE 4. The space domain for the two sides limiter with a schematic plot of the function α

Acknowledgments. This work has been funded by the ANR ESPOIR (Edge Simulation of the Physics Of ITER Relevant turbulent transport) and the *Fédération nationale de Recherche sur la Fusion par Confinement Magnétique* (FR-FCM).The authors thank Guillaume Chiavassa, Guido Ciraolo and Philippe Ghendrih for fruitful discussions.

REFERENCES

[1] Ph. Angot, Ph. Auphan, and O. Guès. Penalty methods for the hyperbolic system modelling the wall-plasma interaction in a tokamak. In *Finite Volumes for Complex Applications VI - Problems & Perspectives*, volume 1, pages 31–38. Springer, June 2011.

[2] Ph. Angot, Ph. Auphan, and O. Guès. An optimal penalty method for an hyperbolic system modeling the edge plasma transport in a tokamak. *Submitted*, 2012.

[3] T. Auphan. Penalization for non-linear hyperbolic system. *Advances in Differential Equations*, 2013. Accepted.

[4] G. Carbou and P. Fabrie. Boundary layer for a penalization method for viscous incompressible flow. *Differential Equations*, 8(12):1453–1480, 2003.

[5] B. Fornet and 0. Guès. Penalization approach of semi-linear symmetric hyperbolic problems with dissipative boundary conditions. *Discrete and Continuous Dynamical Systems*, 23(3):827 – 845, 2009.

[6] T. Gallouët, J-M. Hérard, and N. Seguin. Some approximate Godunov schemes to compute shallow-water equations with topography. *Computers and Fluids*, 32(4):479 – 513, 2003.

[7] 0. Guès. Problème mixte hyperbolique quasi-linéaire caractéristique. *Communications in Partial Differential Equations*, 15:595–654, 1990.

[8] L. Isoardi, G. Chiavassa, G. Ciraolo, P. Haldenwang, E. Serre, Ph. Ghendrih, Y. Sarazin, F. Schwander, and P. Tamain. Penalization modeling of a limiter in the tokamak edge plasma. *Journal of Computational Physics*, 229(6):2220 – 2235, 2010.

[9] J. B. Rauch and F. J. III Massey. Differentiability of solutions to hyperbolic initial-boundary value problems. *Trans. Amer. Math. Soc.*, 189:303–318, 1974.

E-mail address: thomas.auphan@univ-amu.fr
E-mail address: philippe.angot@univ-amu.fr
E-mail address: olivier.gues@cmi.univ-amu.fr

ON THE DOI MODEL FOR THE SUSPENSIONS OF ROD-LIKE MOLECULES IN COMPRESSIBLE FLUIDS

Hantaek Bae

Department of Mathematics
University of California, Davis, USA

Konstantina Trivisa

Department of Mathematics
& Institute for Physical Science and Technology
University of Maryland, College Park, USA

Abstract. Polymeric fluids arise in many practical applications in biotechnology, medicine, chemistry, industrial processes and atmospheric sciences. In this article, we investigate the Doi model for the suspensions of rod-like molecules in a compressible fluid. This model describes the interaction between the orientation of rod-like polymer molecules on the microscopic scale and the macroscopic properties of the fluid in which these molecules are contained. Prescribing arbitrarily initial data in suitable spaces we establish the global-in-time existence of a weak solution to our model. The proof relies on the construction of an approximate sequence of solutions and the establishment of compactness.

1. **Introduction.** The evolution of rod-like molecules in both compressible and incompressible fluids is of great scientific interest with a variety of applications in science and engineering. The present article deals with the Doi model for the suspension of rod-like molecules in a dilute regime. This model describes the interaction between the orientation of rod-like polymer molecules on the microscopic scale and the macroscopic properties of the fluid in which these molecules are contained. More precisely, the macroscopic flow leads to a change of the orientation and, in the case of flexible particles, to a change in shape of the suspended microstructure; this process in turn yields the production of a fluid stress.

We now derive the system of equations. A smooth motion of a body in continuum mechanics is described by a family of one-to-one mappings $X(t,\cdot):\Omega\to\Omega,\ \ t\in I$. The curve $X(t,x)$ represents the trajectory of a particle occupying a position $x\in\Omega$ at time t and this curve is determined by a velocity field $u: I\times\Omega\to\mathbb{R}^3$ through

$$\frac{\partial}{\partial t}X(t,x)=u\left(t,X(t,x)\right),\quad X(0,a)=a.$$

Then, the conservation of mass can be formulated as follows:

$$\frac{d}{dt}\int_{X(t,B)}\rho(t,x)dx=0,\quad B\subset\Omega,$$

2000 *Mathematics Subject Classification.* Primary: 35D30, 35L15.

Key words and phrases. Doi model, suspensions of rod-like molecules, fluid-particle interaction model, compressible Navier-Stokes equations, Fokker-Planck-type equation.

The second author is partially supported by the National Science Foundation under the awards DMS 0807815 and DMS 1109397.

where ρ is represents the fluid density. This equation is equivalent to

$$\frac{d}{dt}\int_B \rho(t,x)dx + \int_{\partial B} \rho(t,x)\left[u(t,x)\cdot \hat{n}\right] dS = 0,$$

where $\hat{n}$ is the unit outer normal vector on $\partial\Omega$. If ρ is smooth, one can use Green's theorem to deduce the following continuity equation:

$$\rho_t + \nabla\cdot(u\rho) = 0. \tag{1}$$

We next obtain the equation of motion by applying Newton's second law of motion:

$$\frac{d}{dt}\int_{X(t,B)} (\rho u)(t,x)dx = \int_{X(t,B)} \rho(t,x)F(t,x)dx + \int_{\partial X(t,B)} \mathbf{t}\left(t,x,\hat{n}\right) dS,$$

where F is an external force and a vector $\mathbf{t}$ is a traction. (From now, we take $F=0$ for the simplicity of the argument.) This equation is equivalent to

$$\frac{d}{dt}\int_B (\rho u)(t,x)dx + \int_{\partial B} (\rho u)(t,x)\left[u(t,x)\cdot\hat{n}\right] dS = \int_{\partial B} \mathbf{t}(t,x,\hat{n})dS. \tag{2}$$

By the fundamental laws of Cauchy in the continuum mechanics, $\mathbf{t}$ can be expressed by a a symmetric stress tensor $\mathbb{T}(t,x)$; $\mathbf{t}(t,x,\hat{n}) = \mathbb{T}(t,x)\hat{n}$. Therefore, (2) becomes

$$\frac{d}{dt}\int_B (\rho u)(t,x)dx + \int_{\partial B} (\rho u)(t,x)\left[u(t,x)\cdot\hat{n}\right] dS = \int_{\partial B} \mathbb{T}(t,x)\hat{n}dS. \tag{3}$$

By applying Green's lemma to (3), we have

$$(\rho u)_t + \nabla\cdot(\rho u\otimes u) = \nabla\cdot\mathbb{T}, \quad \text{where} \quad (\nabla\cdot\mathbb{T})_i = \sum_{j=1}^{3}\frac{\partial \mathbb{T}_{ij}}{\partial x_j}. \tag{4}$$

The stress tensor $\mathbb{T}$ of a general fluid obeys Stokes' law: $\mathbb{T} = \mathbb{S} - p\mathbb{I}_{3\times3}$, where p is the pressure and $\mathbb{S}$ is the stress tensor. Therefore,

$$(\rho u)_t + \nabla\cdot(\rho u\otimes u) + \nabla p = \nabla\cdot\mathbb{S}. \tag{5}$$

We now define $\mathbb{S}$ and p to our model. We assume that p only depends on ρ;

$$p = a\rho^\gamma, \quad \gamma > 3/2. \tag{6}$$

The stress tensor $\mathbb{S}$ consists of two parts: $\mathbb{S} = \mathbb{S}_1 + \mathbb{S}_2$, where $\mathbb{S}_1$ is the viscous stress tensor generated by the fluid

$$\mathbb{S}_1 = \mu\big(\nabla u + (\nabla u)^t\big) + \lambda(\nabla\cdot u)\mathbb{I}_{3\times3}, \tag{7}$$

and $\mathbb{S}_2$ is the macroscopic symmetric stress tensor derived from the orientation of the rods at the molecular level. We assume that the stress tensor $\mathbb{S}_2$ is given by an expansion

$$\mathbb{S}_2(x,t) = \sigma(x,t) + \sigma^{(1)}(x,t) + \sigma^{(2)}(x,t), \quad \text{where}$$

$$\sigma(t,x) = \int_{S^2}(3\tau\otimes\tau - \mathbb{I}_{3\times3})f(t,x,\tau)d\tau,$$

$$\sigma^{(1)}(t,x) = -\left[\int_{S^2}\gamma_{ij}^{(1)}(\tau)f(t,x,\tau)d\tau\right]\mathbb{I}_{3\times3},$$

$$\sigma^{(2)}(t,x) = -\left[\int_{S^2}\int_{S^2}\gamma_{ij}^{(2)}(\tau_1,\tau_2)f(t,x,\tau_1)f(t,x,\tau_2)d\tau_1 d\tau_2\right]\mathbb{I}_{3\times3}.$$

This, and more general expansions for $\mathbb{S}_2$ are encountered in the polymer literature (cf. Doi and Edwards [9]). We also refer the reader to the articles by Constantin et al [6], [7], where a general class of stress tensors is presented in the context of incompressible fluids. The structure coefficients in the expansion $\gamma_{ij}^{(1)}, \gamma_{ij}^{(2)}$ are in general smooth and (t, x, f) independent. Assuming for simplicity that $\gamma_{ij}^{(1)}(\tau) = \gamma_{ij}^{(2)}(\tau_1, \tau_2) = 1$ and denoting $\eta(t, x) = \int_{S^2} f(t, x, \tau) d\tau$, $\mathbb{S}_2$ takes the form

$$\mathbb{S}_2(x, t) = \sigma(x, t) - \eta \mathbb{I}_{3\times 3} - \eta^2 \mathbb{I}_{3\times 3}. \tag{8}$$

By substituting (6), (7) and (8) to (5), the equation of motion becomes

$$(\rho u)_t + \nabla \cdot (\rho u \otimes u) - \mu \Delta u - \lambda \nabla(\nabla \cdot u) + a \nabla \rho^\gamma + \nabla \eta^2 = \nabla \cdot \sigma - \nabla \eta. \tag{9}$$

We note that η and σ depend on the distribution f, which is described by a compressible Fokker-Plank equation,

$$f_t + \nabla \cdot (uf) + \nabla_\tau \cdot \big(P_{\tau^\perp}(\nabla u \tau) f\big) - D_\tau \Delta_\tau f - D \Delta f = 0, \tag{10}$$

where $P_{\tau^\perp}(\nabla_x u \tau) = \nabla_x u \tau - (\tau \cdot \nabla_x u \tau)\tau$ is the projection of $\nabla u \tau$ on the tangent space of S^2 at $\tau \in S^2$. With ∇_τ and Δ_τ we denote the gradient and the Laplace operator on the unit sphere, while ∇ and Δ represent the gradient and the Laplacian operator in $\mathbb{R}^3$. The second term $\nabla \cdot (uf)$ in (10) describes the change of f due to the displacement of the center of mass of the rods due to macroscopic advection, while the term $\nabla_\tau \cdot \big(P_{\tau^\perp}(\nabla u \tau) f\big)$ is a drift-term on the sphere representing the shear-forces acting on the rods. The term $D_\tau \Delta_\tau f$ and $D \Delta f$ represent the rotational and translation diffusion due to Brownian motion and these effect causes the rods to change their orientation spontaneously. To obtain a closed system of equations, we finally need the equation of η; by integrating (10) over S^2, we have

$$\eta_t + \nabla \cdot (u \eta) - D \Delta \eta = 0. \tag{11}$$

In sum, after normalizing all the constants by 1, we have the following system of equations:

$$\rho_t + \nabla \cdot (\rho u) = 0, \tag{12a}$$

$$(\rho u)_t + \nabla \cdot (\rho u \otimes u) - \Delta u - \nabla(\nabla \cdot u) + \nabla \rho^\gamma + \nabla \eta^2 = \nabla \cdot \sigma - \nabla \eta, \tag{12b}$$

$$f_t + \nabla \cdot (uf) + \nabla_\tau \cdot (P_{\tau^\perp}(\nabla_x u \tau) f) - \Delta_\tau f - \Delta_x f = 0, \tag{12c}$$

$$\eta_t + \nabla \cdot (u \eta) - \Delta \eta = 0, \tag{12d}$$

$$\sigma(t, x) = \int_{S^2} (3\tau \otimes \tau - \mathbb{I}_{3\times 3}) f(t, x, \tau) d\tau. \tag{12e}$$

In this paper, we deal with the above system on a bounded domain $\Omega \subset \mathbb{R}^3_x$. Since viscous fluids are believed to adhere completely to a rigid boundary, we impose Dirichlet boundary conditions to u, f, and η:

$$u = 0, \quad f = 0, \quad \text{and} \quad \eta = 0 \text{ on } \partial\Omega.$$

Related results on the Doi model for the suspension of rod-like molecules in *incompressible* fluids have been studied by many authors. We refer the reader to Constantin [6, 7, 8], Lions and Masmoudi [15, 16], Masmoudi [17] and Otto and Tzavaras [20] for results on related models on the whole space. In [1], the authors treat the Doi model for an incompressible fluid within a bounded domain and establish results on the global existence of solutions. For *compressible* models, related results have been presented in a series of articles. We refer the reader to Carrillo et al [3, 4, 5], Goudon et al [12, 13], and Mellet and Vasseur [18, 19], where asymptotic, analytical and numerical results on related fluid-particle interaction models are discussed. What distinguishes the model presented in this article, besides the general type of the stress tensor under consideration, is the fact that, unlike other models, the Fokker-Planck-type equation considered here takes into consideration the presence of the *shear forces* acting on the rods as well as the Brownian effects. This new element yields a new equation for the entropy induced by the probability density function f at the microscopic level.

Since the definition of a weak solution and the main result are rather complicated, they are stated in Section 2, and we provide the outline of the proof in Section 3.

2. **Definition of weak solution and main result.** The notion of weak solution follows from the energy identity; we multiply (12b) by u and integrate over Ω;

$$\begin{aligned}&\frac{d}{dt}\int_\Omega\Big[\frac{\rho|u|^2}{2}+\frac{\rho^\gamma}{\gamma-1}+\eta^2\Big]dx+\int_\Omega\Big[|\nabla u|^2+|\nabla\cdot u|^2+2|\nabla\eta|^2\Big]dx\\&=-\int_\Omega\nabla u:\sigma dx+\int_\Omega(\nabla\cdot u)\eta dx.\end{aligned}\tag{13}$$

To obtain the energy identity, we need to remove the right-hand side of (13). To this end, we introduce an entropy ψ at the microscopic level; $\psi(t,x)=\displaystyle\int_{S^2}(f\ln f)(t,x,\tau)d\tau$. Then, ψ satisfies

$$\begin{aligned}&\psi_t+\nabla\cdot(u\psi)-\Delta\psi+4\int_{S^2}|\nabla_\tau\sqrt{f}|^2d\tau+4\int_{S^2}|\nabla\sqrt{f}|^2d\tau\\&=\nabla u:\sigma-(\nabla\cdot u)\eta.\end{aligned}\tag{14}$$

Integrating (14) over Ω, we obtain

$$\frac{d}{dt}\int_\Omega\psi dx+4\int_\Omega\int_{S^2}\left(|\nabla_\tau\sqrt{f}|^2+|\nabla\sqrt{f}|^2\right)d\tau dx=\int_\Omega\left(\nabla u:\sigma-(\nabla\cdot u)\eta\right)dx.\tag{15}$$

By adding (15) to (13), we have

$$\begin{aligned}&\frac{d}{dt}\int_\Omega\Big[\frac{\rho|u|^2}{2}+\frac{\rho^\gamma}{\gamma-1}+\eta^2+\psi\Big]dx+4\int_\Omega\int_{S^2}|\nabla_\tau\sqrt{f}|^2d\tau dx\\&+4\int_\Omega\int_{S^2}|\nabla\sqrt{f}|^2d\tau dx+\int_\Omega\Big[|\nabla u|^2+|\nabla\cdot u|^2+2|\nabla\eta|^2\Big]dx=0.\end{aligned}\tag{16}$$

We now define a weak solution of the system (12) based on (16). Let $\gamma>\frac{3}{2}$ and Ω be a C^1 bounded domain. Assume that initial data $\{\rho_0,u_0,f_0,\eta_0\}$ satisfy

$$\begin{aligned}&\rho_0\in L^1\cap L^\gamma(\Omega),\quad \rho_0u_0=m_0\in L^{\frac{2\gamma}{\gamma+1}}(\Omega),\quad f_0\in L^1(\Omega\times S^2),\quad \eta_0\in L^2(\Omega),\\&\frac{m_0^2}{\rho_0}\in L^1(\Omega)\quad\text{for}\quad\rho_0\neq0,\quad\frac{m_0^2}{\rho_0}=0\quad\text{for}\quad\rho_0=0.\end{aligned}\tag{17}$$

Definition 2.1. We say the set $\{\rho,u,f,\eta,\sigma\}$ is a weak solution of (12) if

(i) ρ,u,f,η,σ satisfy

$$\begin{aligned}&\rho\in L^\infty(0,T;L^\gamma(\Omega)),\quad\nabla u\in L^2(0;T;L^2(\Omega)),\\&\rho|u|^2\in L^\infty(0,T;L^1(\Omega)),\quad\rho u\in C_w([0,T];L^{\frac{2\gamma}{\gamma+1}}(\Omega)),\\&\eta\in L^\infty(0,T;L^2(\Omega))\cap L^2(0,T;\dot{H}^1(\Omega)),\quad f\ln f\in L^\infty(0,T;L^1(\Omega\times S^2))\\&\nabla_\tau\sqrt{f}\in L^2(\Omega\times S^2\times(0,T)),\quad\nabla\sqrt{f}\in L^2(\Omega\times S^2\times(0,T)),\end{aligned}$$

(ii) (12a) holds in the sense of renormalized solutions, i.e.,

$$b(\rho)_t + \nabla\cdot(b(\rho)u) + \Big(b^{'}(\rho)\rho - b(\rho)\Big)\,\nabla\cdot u = 0 \tag{18}$$

holds in the sense of distributions for any $b \in C^1$, $|b^{'}(z)z| + |b(z)| \leq C\ \forall z \in \mathbb{R}$,
(iii) (12b), (12c), and (12d) hold in the sense of distributions,
(iv) and $\{\rho, u, f, \eta, \sigma\}$ satisfy the following energy inequality:

$$\begin{aligned}
&\int_\Omega \Big[\frac{\rho|u|^2}{2} + \frac{\rho^\gamma}{\gamma-1} + \eta^2 + \psi\Big](t)dx + 4\int_0^t\int_\Omega\int_{S^2} |\nabla_\tau\sqrt{f}|^2 d\tau dx dt\\
&+ 4\int_0^t\int_\Omega\int_{S^2}|\nabla\sqrt{f}|^2 d\tau dx dt + \int_0^t\int_\Omega \Big[|\nabla u|^2 + |\nabla\cdot u|^2 + 2|\nabla\eta|^2\Big]dxdt\\
&\leq \int_\Omega \Big[\frac{\rho_0|u_0|^2}{2} + \frac{\rho_0^\gamma}{\gamma-1} + \eta_0^2 + \psi_0\Big]dx.
\end{aligned} \tag{19}$$

Remark 1. (1) The central difficulty in showing the existence of a weak solution in the theory of compressible fluids is typically the dependence of the pressure on ρ^γ. From the a priori estimate, we have $\rho \in L^\infty(0,T;L^\gamma(\Omega))$, which is not enough to pass to the limit to $\nabla\rho^\gamma$ in the sense of distributions. The issue is resolved by showing that ρ satisfies a better integrability condition in the renormalized form (18) ([10, 14]). Note that in the present context the suspension stress tensor depends on the density of the particles in a nonlinear way as well: $\nabla\eta^2$. This term can be easily handled from the regularity of η: $\eta \in L^2(0,T;H^1(\Omega))$.
(2) The additional difficulties in the present context involve the presence of two nonlinear terms in the equation of f. For $\chi \in C_c^\infty(\Omega\times S^2)$

$$\begin{aligned}
&\int_\Omega\int_{S^2} \nabla\cdot(u^{(n)}f^{(n)})\chi d\tau dx = -\int_\Omega u_i^{(n)}\Big[\int_{S^2}\partial_{x_i}\chi f^{(n)}d\tau\Big]dx,\\
&\int_\Omega\int_{S^2}\nabla_\tau\cdot\Big(P_{\tau^\perp}(\nabla_x u^{(n)}\tau)f^{(n)}\Big)\,\chi d\tau dx = -\int_\Omega \frac{\partial u_i^{(n)}}{\partial x_j}\Big[\int_{S^2}\tau_j f^{(n)}\frac{\partial\chi}{\partial\tau_i}d\tau\Big]dx.
\end{aligned} \tag{20}$$

To pass to the limit in (20), we need to show that $\displaystyle\int_{S^2}\partial_{x_i}\chi f^{(n)}d\tau$ and $\displaystyle\int_{S^2}\tau_j f^{(n)}\frac{\partial\chi}{\partial\tau_i}d\tau$ converge strongly in $L^2(0,T;L^2(\Omega))$.

Theorem 2.2. *Let $\gamma > \frac{3}{2}$ and Ω be a C^1 bounded domain. Assume that initial data $\{\rho_0, u_0, f_0, \eta_0\}$ satisfy (17). Then, there exists a weak solution $\{\rho, u, f, \eta, \sigma\}$ of the system (12) satisfying (17) at $t = 0$. Moreover, $\rho \in L^p(\Omega\times(0,T))$, $\quad p = \frac{5}{3}\gamma - 1$.*

3. **Outline of Proof.** In order to prove the existence of a weak solution, we first establish compactness of an approximate sequence of solutions $\{\rho^n, u^n, f^n, \eta^n, \sigma^n\}_{n\geq1}$, which is stated in Proposition 1 below. Then, we apply Proposition 1 to an approximate sequence of solutions constructed in Section 3.2 to complete the proof of the existence of a weak solution. We note that compared to approximating schemes used for macroscopic fluid equations in [5], [11] and [14], the scheme presented here is designed to deal with microscopic variables, too. For details of proofs, see [2].

3.1. **Compactness.** We begin with the compactness. Assume that the energy inequality (19) holds for a sequence $\{\rho^n, u^n, f^n, \eta^n, \sigma^n\}_{n\geq1}$. Then, we can obtain various estimates of $\{\rho^n, u^n, f^n, \eta^n, \sigma^n\}_{n\geq 1}$. The energy inequality implies directly the following bounds

$$\begin{aligned} &\rho^n|u^n|^2 \in L^\infty(0,T;L^1(\Omega)), \quad \rho^n \in L^\infty(0,T;L^\gamma(\Omega)),\\ &\nabla u^n \in L^2(0,T;L^2(\Omega)), \quad \eta^n \in L^\infty(0,T;L^2(\Omega)) \cap L^2(0,T;H^1(\Omega)). \end{aligned} \tag{21}$$

We can combine these bounds to obtain other bounds. First, by expressing $\rho^n u^n$ as $\rho^n u^n = \sqrt{\rho^n}\cdot\sqrt{\rho^n}u^n$, and using $\sqrt{\rho^n} \in L^\infty(0,T;L^{2\gamma}(\Omega))$, we have

$$\rho^n u^n \in L^\infty(0,T;L^{\frac{2\gamma}{\gamma+1}}(\Omega)). \tag{22}$$

By the entropy dissipation from (16) and the embedding $\dot{H}^1 \subset L^6$,

$$\sqrt{f^n} \in L^2\big(0,T;L^2(\Omega)L^6(S^2) \cap L^6(\Omega)L^2(S^2)\big),$$

which implies that

$$f^n \in L^1\big(0,T;L^1(\Omega)L^3(S^2) \cap L^3(\Omega)L^1(S^2)\big) \subset L^1(0,T;L^2(\Omega\times S^2)). \tag{23}$$

We finally estimate σ^n. Since $|\sigma^n(t,x)| \leq 3\displaystyle\int_{S^2} f^n(t,x,\tau)d\tau = 3\eta^n(t,x)$,

$$\sigma^n \in L^1(0,T;L^3(\Omega)) \cap L^\infty(0,T;L^2(\Omega)), \tag{24}$$

where the first space is derived from (23) and the second bound is from $\eta^n \in L^\infty(0,T;L^2(\Omega))$. We next estimate the derivative of σ;

$$\begin{aligned} |\nabla\sigma^n(t,x)| &\leq 3\int_{S^2}|\nabla f^n(t,x,\tau)|d\tau \leq C\sqrt{\int_{S^2}|\nabla\sqrt{f^n}|^2 d\tau}\sqrt{\int_{S^2}(\sqrt{f^n})^2 d\tau}\\ &= \sqrt{\int_{S^2}|\nabla\sqrt{f^n}|^2 d\tau}\sqrt{\eta^n} \in L^2(0,T;L^{\frac{4}{3}}(\Omega)). \end{aligned}$$

Proposition 1 (Compactness)**.** *Extracting a subsequence, using the same notation* $\{\rho^n, u^n, f^n, \eta^n, \sigma^n\}_{n\geq 1}$*, the limit functions satisfy the following statements.*

(1) $\rho^n \rightharpoonup \rho$ *in* $L^\gamma(\Omega\times(0,T))$, $\rho \in L^\infty(0,T;L^1\cap L^\gamma(\Omega))$,
(2) $\sqrt{\rho^n}u^n \rightharpoonup \sqrt{\rho}u$ *in* $L^2(0,T;L^2(\Omega))$, $\sqrt{\rho}u \in L^\infty(0,T;L^2(\Omega))$
(3) $u^n \rightharpoonup u$ *in* $L^2(0,T;H^1(\Omega))$, $\sqrt{\rho^n} \rightharpoonup \sqrt{\rho}$ *in* $L^{2\gamma}(\Omega\times(0,T))$
(4) $\rho^n u^n \rightharpoonup \rho u$ *in* $L^{\frac{2\gamma}{\gamma+1}}(\Omega\times(0,T))$, $\rho u \in L^\infty(0,T;L^{\frac{2\gamma}{\gamma+1}}(\Omega))$
(5) $\rho^n u^n_i u^n_j \rightharpoonup \rho u_i u_j$ *in the sense of measures,* $\rho u_i u_j$ *is a bounded measure.*
(6) η^n *converges strongly to* η *in* $L^2(\Omega\times(0,T))$*, and* σ^n *converges strongly to* σ *in* $L^2(\Omega\times(0,T))$.
(7) $\rho^n(\eta^n)^2$ *converges to* $\rho\eta^2$ *in the sense of distributions.*
(8) ρ *and* u *solve (12a) in the sense of renormalized solutions.*
(9) If in addition we assume that ρ^n_0 *converges to* ρ_0 *in* $L^1(\Omega)$,

$$\rho^n \to \rho \quad \textit{in} \quad L^1(\Omega\times(0,T)) \cap C([0,T];L^p(\Omega)) \quad \textit{for all} \quad 1\leq p<\gamma. \tag{25}$$

(10) Finally, we have the following strong convergence:
(i) $\rho_n u^n \to \rho u$ *in* $L^p(0,T;L^r(\Omega))$ *for all* $1\leq p<\infty$, $1\leq r<2\gamma/(\gamma+1)$,
(ii) $u^n \to u$ *in* $L^p(\Omega\times(0,T))\cap\{\rho>0\}$ *for all* $1\leq p<2$,
(iii) $u^n \to u$ *in* $L^2(\Omega\times(0,T))\cap\{\rho\geq\delta\}$ *for all* $\delta>0$,
(iv) $\rho^n u^n_i u^n_j \to \rho u_i u_j$ *in* $L^p(0,T;L^1(\Omega))$ *for all* $1\leq p<\infty$.

3.2. **Approximate sequence of a solution.** We now construct an approximate sequence of solutions to (12), which consists of three parts.

We first regularize $\rho\frac{d}{dt} + \rho u\cdot\nabla$. This step proves Theorem 2.2 by applying Proposition 1 to the sequence of solutions of the following system in the limit $\epsilon \to 0$;

$$\begin{aligned}
&\rho_t + \nabla\cdot(\rho u) = 0,\\
&(\rho_\epsilon u)_t + \nabla\cdot((\rho u)_\epsilon \otimes u) - \Delta u - \nabla(\nabla\cdot u) + \nabla\rho^\gamma + \nabla\eta^2 = \nabla\cdot\sigma_\epsilon - \nabla\eta_\epsilon,\\
&f_t + \nabla\cdot(u_\epsilon f) + \nabla_\tau\cdot(P_{\tau^\perp}(\nabla_x u_\epsilon \tau)f) - \Delta_\tau f - \Delta f = 0,\\
&\eta_t + \nabla\cdot(u_\epsilon\rho) - \Delta\rho = 0.
\end{aligned} \tag{26}$$

The existence of a solution to (26) however requires additional damping.

The second step is to add nonlinear damping to the equation of ρ and η, which provides the existence of solutions of (27) by taking the limit $\delta \to 0$ to the system

$$\begin{aligned}
&\rho_t + \nabla\cdot(\rho u) + \delta\rho^q = 0,\\
&(\rho_\epsilon u)_t + \nabla\cdot((\rho u)_\epsilon \otimes u) - \Delta u - \nabla(\nabla\cdot u) + \nabla[\rho^\gamma + \eta^2] + \delta[(\rho^q)_\epsilon + (\eta^m)_\epsilon]u\\
&= \nabla\cdot\sigma_\epsilon - \nabla\eta_\epsilon,\\
&f_t + \nabla\cdot(u_\epsilon f) + \nabla_\tau\cdot(P_{\tau^\perp}(\nabla_x u_\epsilon \tau)f) - \Delta_\tau f - \Delta f = 0,\\
&\eta_t + \nabla\cdot(\eta u) - \Delta\eta + \delta\eta^m = 0,
\end{aligned} \tag{27}$$

where $q > \gamma + 1$ and $m > 3$, with $m \geq q$. The role of damping is to increase integrability of ρ, η and to show that $\rho \geq C > 0$. These properties will be used to show the existence of solutions to (27) by truncating ρ^γ and η^2.

The final step is truncation of ρ^γ and η^2, which shows the existence of solutions to (27) by taking the limit $N \to \infty$ to the following system of equations

$$\begin{aligned}
&\rho_t + \nabla\cdot(\rho u) + \delta\rho^q = 0,\\
&(\rho_\epsilon u)_t + \nabla\cdot((\rho u)_\epsilon \otimes u) - \Delta u - \nabla(\nabla\cdot u) + \nabla T(N) + \delta[(\rho^q)_\epsilon + (\eta^m)_\epsilon]u\\
&= \nabla\cdot\sigma_\epsilon - \nabla\eta_\epsilon,\\
&f_t + \nabla\cdot(u_\epsilon f) + \nabla_\tau\cdot(P_{\tau^\perp}(\nabla_x u_\epsilon \tau)f) - \Delta_\tau f - \Delta f = 0,\\
&\eta_t + \nabla(\eta u) - \Delta\eta + \delta\eta^m = 0,
\end{aligned} \tag{28}$$

where $T(N) = (\rho\wedge N)^\gamma + (\eta\wedge N)^2$ and $f\wedge N = \min\{f, N\}$. Since ρ is bounded and strictly positive, we can obtain a parabolic equation of u by dividing (12b) by ρ. By truncating ρ^γ and η^2 in the equation of u, we can show that $\nabla u \in L^\infty$, which implies the global existence of solutions to (28). By taking the limit $N \to \infty$, we show the existence of solutions to (27).

4. Concluding remarks.

The present article is part of a research program whose objective is the investigation of general models for polymeric fluids in both compressible and incompressible fluids and in domains with complex geometries. Investigating their asymptotic behavior over bounded domains are also of the goals of this program. We also remark that the investigation of singular limits of complex fluids for compressible flows over bounded domains is of great scientific interest, physically relevant and presents new challenges in the analysis. Unlike the cases involving the whole domain or exterior domains where acoustic waves are damped due to dispersive effects of the wave equation, the main obstacle in the treatment of bounded domains is the persistency of the fast waves over these domains. Therefore in general one can only expect weak convergence of the solutions. There are situations where strong convergence can be achieved due to the interaction of acoustic waves with the boundary of the domain. This phenomenon has been observed for both asymptotic behavior of fluid equations and hydrodynamic limits of kinetic equations. It is therefore natural to ask whether similar phenomena happen for models of polymeric fluids.

Acknowledgments. H.B. gratefully acknowledges the support provided by the Center for Scientific Computation and Mathematical Modeling (CSCAMM) at the University of Maryland where this research was performed.

REFERENCES

[1] H. Bae and K. Trivisa, *On the Doi model for the suspensions of rod-like molecules: Global-in-time existence*, To appear in Commun. Math. Sci., (2012).

[2] H. Bae and K. Trivisa, *On the Doi model for the suspensions of rod-like molecules in compressible fluids*, To appear in Math. Models Methods Appl. Sci., **22** (2012).

[3] J. A. Carrillo and T. Goudon, *Stability and Asymptotic Analysis of a Fluid-Particle Interaction Model*, Comm. Partial Differential Equations, **31** (2006), 1349–1379.

[4] J. A. Carrillo, T. Goudon, and P. Lafitte, *Simulation of fluid and particles flows: asymptotic preserving schemes for bubbling and flowing regimes*, J. Comput. Phys., **227** (2008), 7929–7951.

[5] J. Carrillo, T. Karper and K. Trivisa, *On the dynamics of a fluid-particle interaction model: the bubbling regime*, Nonlinear Analysis: Theory, Methods & Applications, **74** (2011) 2778–2801.

[6] P. Constantin, *Nonlinear Fokker-Planck Navier-Stokes systems*, Commun. Math. Sci., **3** (2005), no.4, 531–544.

[7] P. Constantin, C. Fefferman, E.S. Titi and A. Zarnescu, *Regularity of coupled two-dimensional nonlinear Fokker-Planck and Navier-Stokes systems*, Comm. Math. Phys., **270** (2007), no.3, 789–811.

[8] P. Constantin and N. Masmoudi, *Global well-posedness for a Smoluchowski equation coupled with Navier-Stokes equations in 2D*, Comm. Math. Phys., **278** (2008), no.1, 179–191.

[9] M. Doi and S.F. Edwards, *The theory of polymer dynamics*, Oxford University press, (1986).

[10] E. Feireisl, *Dynamics of viscous compressible fluids.* Oxford University Press, (2003).

[11] E. Feireisl, *On compactness of solutions to the compressible isentropic Navier-Stokes equations when the density is not square integrable*, Comment. Math. Univ. Carolin., **42** (2001), no.1, 83–98.

[12] Th. Goudon, P.-E. Jabin, and A. Vasseur, *Hydrodynamic limit for the Vlasov-Navier-Stokes equations. I. Light particles regime*, Indiana Univ. Math. J., **53** (2004), no.6, 1495–1515.

[13] Th. Goudon, P.-E. Jabin, and A. Vasseur, *Hydrodynamic limit for the Vlasov-Navier-Stokes equations. II. Fine particles regime*, Indiana Univ. Math. J., **53** (2004), no.6, 1517–1536.

[14] P.L. Lions, *Mathematical topics in fluid dynamics, Vol.2, Compressible models*, Oxford Science Publication, (1998).

[15] P.L. Lions and N. Masmoudi, *Global solutions for some Oldroyd models of non-Newtonian flows*, Chinese Ann. Math. Ser. B., **21** (2000), no.2, 131–146.

[16] P.L. Lions and N. Masmoudi, *Global solutions of weak solutions to some micro-macro models*, C.R. Math. Sci. Paris, **345** (2007), no.1, 15–20.

[17] N. Masmoudi, *Global Existence of weak solutions to the FENE Dumbbell model of polymeric flows*, Preprint (2010).

[18] A. Mellet and A. Vasseur, *Asymptotic analysis for a Vlasov-Fokker-Planck/compressible Navier-Stokes system of equations*, Comm. Math. Phys., **281** (2008), 573–596.

[19] A. Mellet and A. Vasseur, *Global weak solutions for a Vlasov-Fokker-Planck/Navier-Stokes system of equations*, Math. Models Methods Appl. Sci., **17** (2007), 1039–1063.

[20] F. Otto and A. Tzavaras, *Continuity of velocity gradients in suspensions of rod-like molecules*, Comm. Math. Phys., **277** (2008), no.3, 729–758.

E-mail address: hantaek@math.ucdavis.edu
E-mail address: trivisa@math.umd.edu

MULTI-SCALE TISSULAR-CELLULAR MODEL FOR WOUND HEALING

Luís Almeida

CNRS, UMR 7598, Laboratoire Jacques-Louis Lions
and UPMC Univ Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions
F-75005, Paris, France

Patrizia Bagnerini

Dipartimento di Ingegneria Meccanica
Energetica, Gestionale e dei Trasporti
Università degli Studi di Genova
P.le Kennedy-Pad D, 16129 Genova, Italy

Abstract. In previous works we developed continuous mathematical models for wound healing and dorsal closure in Drosophila embryos. In this paper we extend this study to the case of non convex wounds in Drosophila pupal epithelium where the sign of the local curvature of the boundary plays an important role in determining the type of acto-myosin contractile structure that is formed. Moreover, we propose a multi-scale model where we combine the previous continuous approach with a cellular-level model that also takes into account interfacial tension between cells. We therefore minimize an extended energy functional so that the junctions of the cells are moved through successive configurations in order to obtain a new mechanical equilibrium. We apply this model to study some simple situations of cell sorting and the movement of genetic clones.

1. Introduction.

Extension of an epithelial membrane to close a hole is a very widespread process both in morphogenesis and in tissue repair. Contraction of actin structures (in one, two or three dimensions) plays an important role in many cellular and tissue movements, both at a multicellular tissue level and at a cellular (and even intracellular) one: from muscle contraction to cell crawling and the contractile ring in cytokinesis. In the [2], [3], [1] we proposed various mathematical models for simulating the contraction of an actin cable structure attached to an external epithelial tissue in different applications such as wound healing or dorsal closure in Drosophila Melanogaster (fruit fly) embryos.

The present work is a natural sequel of these works, but here we are more concerned with wound healing in the pupal stage of Drosophila. An interesting feature of this stage is that the epidermis is under considerably less tension than in embryos. This enables us to make holes which do not become convex while they open, differently to embryo experiments where the the strong pull of the external epidermis renders the hole convex during the opening phase.

Key words and phrases. Actin cable, Wound healing, Dorsal closure, Movement of epithelia, Forces in embryogenesis.

The authors thank the ANR project REGENR for support.

In particular, in collaboration with A. Jacinto's lab (CEDOC, Universidade Nova de Lisboa), we generated C-shaped wounds, where part of the boundary is concave and the rest convex (see figure 1). Observing the healing of this type of wound, we realized that the parts of the boundary where the curvature is positive and those where it is negative don't behave in the same way (see also the discussion in [5]). We propose here (in section 2) to extend the previous model to describe and simulate this phenomenon. Then, in section 3, we couple the continuum model with a cellular one, where the epidermal tissue is described by a two-dimensional network of cells interacting through their common boundaries.

There are many different cell types in multicellular organisms that result from differentiation from embryo stem cells (through specific gene expression) as part of tissue specialization mechanisms: during morphogenesis cells are separated into compartments that are essential for proper assembly of the body's organs. Inside these compartments, once differentiated, cells are usually committed to a particular lineage and cells belonging to each compartment stay together and do not mix with those from other compartments. Cell sorting is therefore an interesting and open problem. Two main hypothesis have been formulated: first, cell segregation at compartment boundaries could be based on *differential cell adhesion* or affinity (i.e. cell populations might develop distinct adhesive properties which prevent intermingling), but molecules involved in these processes have still not been completely identified. The second hypothesis is the *differential interfacial tension* i.e. cells in contact with neighboring cells of a different type, increase the tension at the interfaces and contract the corresponding surfaces.

Motivated by these problems and by some experiences in collaboration with A. Jacinto (work in progress), we consider the tissue formed by a group of cells of two different types, for instance wild type ones and others that are mutant for a certain gene. In section 4, in order to take into account the difference in the interfacial tension between cells of different types, we introduce in the continuum-cellular model an extended energy functional. The junctions of the cells are then moved through successive configurations minimizing this functional in order to obtain a new mechanical equilibrium at each step.

2. **C-shaped wounds in Drosophila pupae.**

In the few minutes that follow a laser (or a mechanical) circular wounding of the epidermis filamentous actin and myosin II concentrate inside the adjacent cells and give rise to a local acto-myosin cable anchored to the adherens junctions that bind the cell to its neighbors. The result is a continuous, supracellular acto-myosin cable that encircles the wound and reduces the wound's perimeter by a purse-string mechanism. In the (not yet published) experiences in collaboration with A. Jacinto's lab, in the C-shaped wounds on pupae, the acto-myosin cable seems to form mainly on the part of the boundary where the curvature is positive. For this reason, in our model the contractile actin cable term is given by $\max(\kappa, 0)$, where κ represents the curvature.

In [4], the authors considered the question of arbitrarily shaped wounds from a theoretical point of view: denoting by $E(t)$ (a subset of $\mathbb{R}^2$) the position of the wound at time t, they study $C^{1,1}$ solutions of the formal geometric equation

$$V(x) = \begin{cases} \kappa(x) & \text{if } x \in \Omega \\ \max(\kappa(x), 0) & \text{if } x \in \partial\Omega \end{cases} \tag{1}$$

where κ denotes the curvature of a closed curves $\partial E(t) \subset \mathbb{R}^2$ (boundaries of the sets $E(t)$) and V the normal inward velocity (i.e. pointing inside $E(t)$). They prove that the solution $E(t)$ of the mean curvature flow with obstacle is contained at each time t in the corresponding solution of the unconstrained mean curvature flow starting from the same set $E(0)$. In the context of our original biological problem and the proposed models, this result indicates that the strategy of assembling an acto-myosin cable only in the positive curvature part of the boundary of the wound (or hole) allows us close it in a more efficient way than if we had assembled the cable all around the boundary.

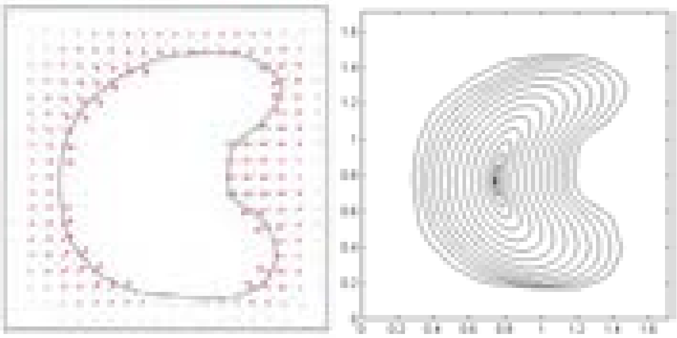


FIGURE 1. The initial wound is a C-shaped curve in the domain [0, 1.7]x[0, 1.7]. We show in the left part of the figure the vector field u_i solution of problem (2) at the initial step $i = 1$ with $f_{cable} = 0.05$ and $f_{pull} = 0$. In the right part we show the successive curve positions.

Flowing only by the positive part of the curvature corresponds (see [4]) to the special case of problem (1) where the obstacle is taken to be the initial position of the boundary. Another nice feature of this type of flow having a natural biological interpretation is the fact that, during the wound closure phase, the epidermis advances without ever retreating, even locally or temporarily, from a region it has occupied (which would not be always true for a full curvature flow corresponding to having assembled a cable all around the wound closure).

In adult wounds, the main closure mechanism is lamellipodial crawling, i.e. the cells in the first rows extend lamellipodia (which are essentially two dimensional actin structures) that attach to extracellular matrix and pull the epithelium forward into the wounded area. Beneath, at the dermal level, activated fibroblasts proliferate and give rise to the granulation tissue which actively contracts. Both these contributions will be taken into account in the model by introducing in the equation (2) the term $f_{pull}\,\mathbf{n}$, i.e. an active pull (of the lamellipodia or the connective tissue) on the leading edge that moves it inwards to close the hole.

For epidermal wounds and the morphogenetic movements that we consider, the time scale of the closure is very long (hours) while the space scale is very small (cell characteristic length - of the order of a few μm). Therefore, it is reasonable to, in a first approximation, neglect inertial forces, and assume that the dynamic process of wound closure is a succession of static equilibria, i.e. to do a quasi-static approximation.

Taking into account the previous discussion, we propose the following quasistatic model: we consider as simulation domain a rectangle M, which contains the wound at time step i, denoted by W_i. Let D_i be the part of the domain occupied by the epidermis i.e. $D_i = M \setminus W_i$ and $\omega_i = \partial W_i$ the boundary of the wound ("the leading edge"). The acto-myosin cable tension gives rise to a force that is proportional to the curvature. This term will be described by a normal force which is proportional to the local curvature of the leading edge at each point. It points towards the interior of the wound W_i at the points of positive curvature and towards the exterior at points of negative curvature. We assume that at each time step i, the corresponding displacement field $\mathbf{u}_i$ satisfies

$$\begin{cases} -\Delta \mathbf{u}_i = \mathbf{0} \ \text{ in } \ D_i, \\ \mathbf{u}_i = \mathbf{0} \ \text{ on } \partial M, \\ \dfrac{\partial \mathbf{u}_i}{\partial n} = f_{cable} \ \max(\kappa, 0)\, \mathbf{n} + f_{pull}\, \mathbf{n} \ \text{ on } \omega_i \ . \end{cases} \tag{2}$$

where $\mathbf{n}$ is the external unit normal to ∂D_i at each point, κ the curvature, f_{cable} the function associated with the intensity of the cable tension at each point along the leading edge ω_i and f_{pull} is function describing the intensity of the inwards pull (of the filopodia/lamellipodia or the connective tissue). We use bold face letters for $\mathbf{u}_i$, $\mathbf{n}$ and $\mathbf{0}$ to make it clear that all these quantities are vectors. We choose Dirichlet homogeneous boundary condition on the boundary of the rectangle M, since, differently from embryos, in the pupae stage epidermis is not under a significant tension and the tissue can be considered at equilibrium.

The algorithm is the following. First we compute the solution of problem (2) in the domain $D_i = M \backslash W_i$ by using finite element methods on a triangular mesh (using Comsol Multiphysics software). We obtain in this way a displacement field $\mathbf{u}_i$. Then, we extended it to the domain inside the inner boundary ω_i by solving

$$\begin{cases} -\Delta \mathbf{u}_i^{int} = 0 \ \text{ in } \ W_i, \\ \mathbf{u}_i^{int} = \mathbf{u}_i \ \text{ on } \omega_i. \end{cases} \tag{3}$$

We obtain in this way an extension (the harmonic extension) of the original vector field $\mathbf{u}_i$ (which for simplicity we will still denote by $\mathbf{u}_i$) defined on rectangular domain M. In order to perform the evolution of contour ω_i, we use level set methods. They consist in implicitly representing the front ω_i as the zero level set of a function $\Phi : \mathbb{R}^2 \times \mathbb{R}^+ \to \mathbb{R}$, solution of the Hamilton-Jacobi equation (HJ)

$$\begin{cases} \partial_t \Phi(x,t) + \mathbf{u}_i(x) \cdot \nabla \Phi(x,t) = 0 \ \text{ in } \ M \times [0,T], \\ \Phi(x,0) = \Phi_i(x) \ \text{ in } \ M, \end{cases} \tag{4}$$

where $\mathbf{u}_i$ (solution of our original problem extended using (3)) gives the direction of front propagation and ∇ denotes the spatial gradient. For the first step, the function $\Phi_1(x)$ is obtained by computing the signed distance to the initial contour ω_1 (positive at the interior of ω_1), whereas for the following contours, $\Phi_i(x)$ is the solution of (4) computed at the previous time step $i-1$.

We use an Eulerian method instead of a particle (Lagrangian) method since changes of topology are naturally handled and surfaces automatically merge and separate. We solve the HJ problem (4) on a regular cartesian grid by using an upwind second order Essentially Non-Oscillatory (ENO) scheme in space and a second order total variation diminishing Runge-Kutta scheme in time. The value of $\mathbf{u}_i$ in the regular grid is computed by interpolating $\mathbf{u}_i$ on a triangular mesh. The level set methods are implemented by using the Matlab toolbox of I. M. Mitchell ([10]).

Since we are doing a quasistatic analysis, the time scale is free for us to fix and thus the numeric values of coefficients f_{cable} and f_{pull} are not physically significant

- just the relative values of the different coefficients make a difference in the simulations (up to rescaling time). To have an idea of size and dynamics of the real situation, small wounds with a diameter of a few tens of μm close in a couple of hours. We choose a time interval T in problem (2) (to pass from the time step i to the time step $i+1$) equal to 0.1 and a spatial discretization step equal to 0.1 both in x and y directions.

The numerical experiment shown in figure 1 corresponds to the successive positions of the boundary ω_i with constant cable tension $f_{cable}(q,i) = 0.2, \forall i = 1 \dots N$, $\forall q \in \omega_i$ and null lamellipodial force $f_{pull} = 0$. We notice that, for simplicity, we do not consider a time dependence (where time is the index i as described before) of the values of the different terms, but the model allows such dependence.

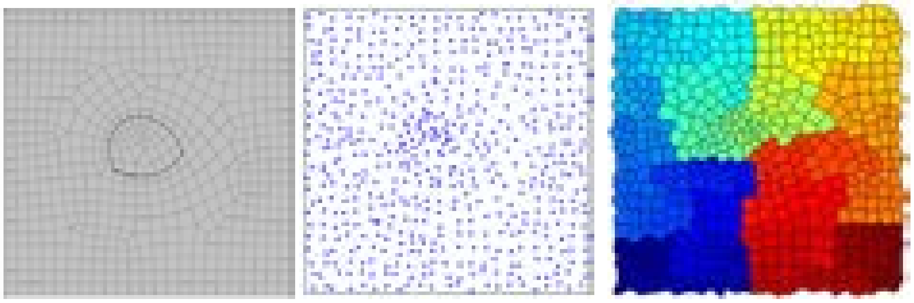


FIGURE 2. Generation of the initial configuration: from the left to the right we show resp. the quadrilateral mesh, the barycenter of the element of the mesh after jiggle and the Voronoi diagram of these points.

3. **Multiscale continuum-cellular model.** The model presented above is a macroscopic model that does not take into account the positions of the individual cells constituting each of the tissues considered. It provides a description at a tissue-level space scale and, being a continuous model, has no ambition of describing the changes in geometry of individual cells - for such a detailed description the geometry of each cell and its neighbors should also play an important role.

Therefore, we decided to couple the continuum model with a cellular one. When the number of cells in the tissue is big, people often concentrate on a small group of cells, not being able to follow in detail all the cells in the system. In this situation, the macroscopic models can be useful for providing reasonable boundary or asymptotic conditions for the cellular-level studies. Moreover, the displacement field $\mathbf{u}_i$ solution of problem (2) can be used to move the junctions of the cells of the tissue. Like in [6, 9, 7], we represent the epidermal tissue as a two-dimensional network of cells discretized as polygons and interacting through their common boundaries (modeled as straight lines connecting vertices).

To perform simulations and test the model, we need to generate a certain number of initial cell configurations. To achieve this, we first generate a quadrilateral mesh in the domain and we compute the barycenters of each elements of the mesh. Next, we jiggle the barycenters of a small random quantity and then we construct the Voronoi triangulation of these points. We choose this procedure to take advantage of the capabilities of mesh generators to obtain elements of the desired size, stretched in one direction, etc. Let $P(i)$ be the vertices of the cells of the epidermal tissue included in D_i at time step i. We displace the set of points $P(i)$ (belonging to ω_i) by using the vector field $\mathbf{u}_i$ in order to obtain a new set of points $P(i+1)$ belonging to D_{i+1}, i.e. the vertices of the cells at the following time. Let $\mathbf{p}_j(i) = (x_j(i), y_j(i)),\ j = 1, \dots N$ be the coordinates of the points of the set $P(i)$ at time step i. We compute the coordinates $\mathbf{p}_j^{i+1}$ of the points in $P(i+1)$ at time step $i+1$ by solving (with a fourth order Runge-Kutta scheme) for each of them, the following boundary value problem (we will be solving this problem N times, with a different initial condition for each j)

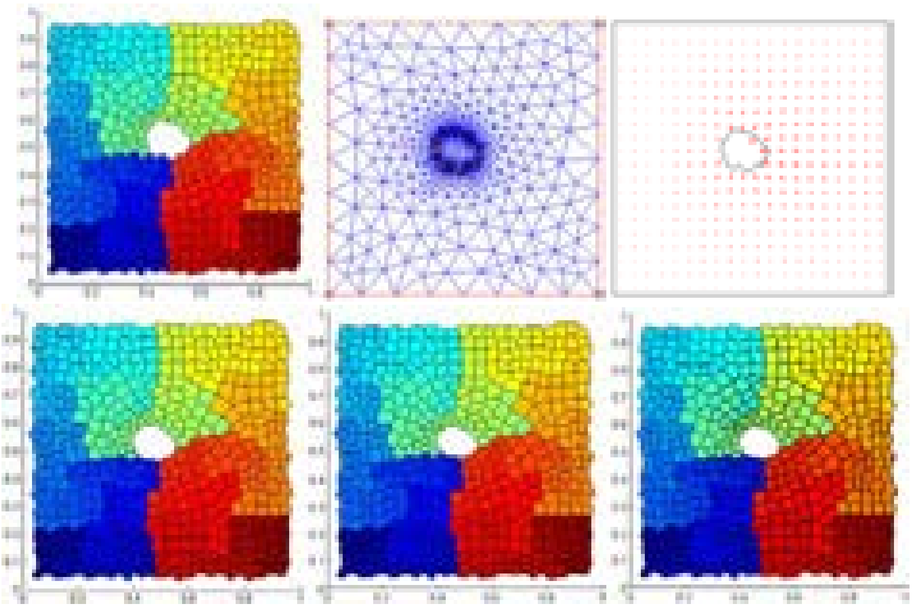


FIGURE 3. Coupling of continuum-cellular model: first row, left to right: the initial configuration, the mesh used to compute the solution of (2) and the corresponding displacement field $\mathbf{u}_i$ at time step i.; second row, left to right: the cells at two successive equilibrium time step and the difference of cells between two successive times.

$$\begin{cases} \mathbf{p}'(t) = \mathbf{u}_i \ \text{ in } \ [0,T] \\ \mathbf{p}(0) = \mathbf{p}_j(i) = (x_j^i, y_j^i). \end{cases} \tag{5}$$

4. **Cell sorting and genetic clones.** We already studied cell segregation in [8]. By studying the cellular tension in Drosophila Dorsal Closure (a stage of embryo development), we realized that there were some cells at the segment boundaries (which we called *mixer* or *chameleon* cells) with a very peculiar and previously not described behavior: they change their genetic identity making a transdifferentiation. Consequently, they do not respect compartment boundaries and give rise to unexpected cell rearrangements at the leading edge.

Motivated by this work and by some experiences on wound healing on Drosophila pupae in collaboration with A. Jacinto (currently in progress), we consider in the tissue a group of cells of different type, i.e. for instance, mutant for a certain gene.

Another possible application is the clones in the wing imaginal disc. Like in the segmentation of Drosophila embryos epithelia, wing disc contains a compartment boundary that separates anterior (A) from posterior (P) cells. This compartment boundary is under the control of the secreted protein Hedgehog (Hh), even thought the precise mechanism remains poorly understood. In [6] they generate cells that lost the ability to transduce the Hh signal (becoming mutant for that gene). Clones in the posterior (P) compartment have wiggly borders with their neighbors since neither cells of the clone nor cells in its neighborhood transduce the Hh signal (cells marked "off" in figure 4). In contrast, clones (arrowhead) have smooth borders when situated in the anterior (A) compartment, since cells surrounding the clone respond to the Hh signal (cells marked "on" in figure 4) and therefore clones try to minimize their surface contact with the neighboring cells. The large clone in the middle is of anterior origin and has taken up a position in the posterior segment. Aiming at studying cell sorting at compartment boundaries, we introduce in the continuum-cellular model the possibility to modify the interfacial tension between cells of different types.

The algorithm is the following. At each time step i, we compute the solution of the continuum model, i.e. the displacement field $\mathbf{u}_i$. Second, we move the junctions of the cells by solving the system of ordinary differential equation (5). Then, we take into account the difference in the interfacial tension between cells of different types, leading the cells to go through successive configurations in order to obtain a new mechanical equilibrium at each step. Stationary and stable network configurations satisfy a mechanical force balance, i.e. at each junction, the sum of forces vanish. We describe these force balances as local minima of an energy function. So, we compute a new stable network configuration by minimizing the following functional (P represents the cell configuration which is defined by the set of the vertices of all the cells, and α is the cell index):

$$F(P) = \sum_{\alpha=1}^{N_C} \sum_{<mn>} \Lambda_{mn} l_{<mn>} + \sum_{\alpha=1}^{N_C} \frac{K}{2}(A_\alpha - \bar{A}_\alpha)^2 + \sum_{\alpha=1}^{N_C} \frac{\Gamma}{2} L_\alpha^2, \tag{6}$$

The first term describes the contributions due to the line tension along each of the edges $< mn >$ (between vertices m and n) of each cell α, the second one to cell area elasticity, and the third one the elasticity of the cell perimeter. The parameter N_C is the number of cells where we perform minimization (it can be a subset of the total cells of the tissue), $l_{<mn>}$ is the edge length, L_α the perimeter of cell α, $\bar{A}_\alpha$ the preferred area (the actual area of the cell at time step i) and Λ_{mn} the line tension on edge $l_{<mn>}$ depending on the gradient in the tension. We obtain in this way a new cell configuration at time step $i+1$. The minimization of the functional F of (6) is performed using the Matlab function fminsearch. In figure 4 we show the result of the simulation. As expected, the clone becomes round, since it minimizes its contact surface with its neighboring cells (which are genetically different).

REFERENCES

[1] Almeida, L.; Bagnerini, P.; Habbal, A.; Noselli, S. and Serman, F. *A mathematical model for dorsal closure*, J Theor Biol, **268** (2011), 105-119.

[2] Almeida, L.; Bagnerini, P.; Habbal, A.; Noselli, S. and Serman, F. *Tissue repair modeling*, Singularities in nonlinear evolution phenomena and applications, Ed. Norm., Pisa, **9** (2009), 27-46.

[3] Almeida, L.; Bagnerini, P. and Habbal, A. *Modeling actin cable contraction*, Comput. Math. Appl., **64** (2012), 310-321.

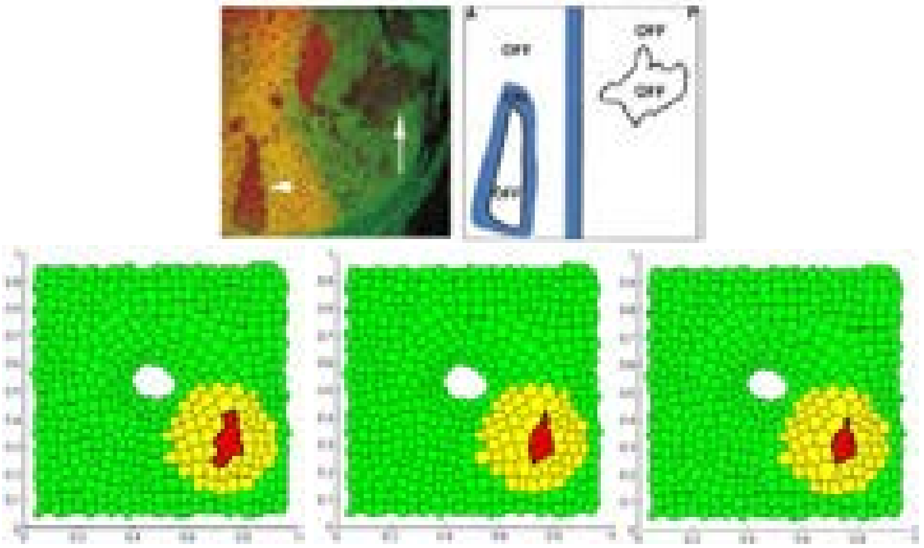


FIGURE 4. First row: the image is taken from [6]. Clones in the posterior (P) compartment have wiggly borders with their neighbors since neither cells of the clone nor cells in its neighborhood transduce the Hh signal (cells marked ”off”). In contrast, clones (arrowhead) have smooth borders when situated in the anterior (A) compartment, since cells surrounding the clone respond to the Hh signal (cells marked ”on”) and therefore try to minimize their surface contact with the neighboring cells. Second row, left to right: the simulation of the movement of the clone at successive time steps.

[4] Almeida, L.; Chambolle, A. and Novaga, M.. *Implicit scheme for mean curvature flow with obstacles*, Annales de l'Institut Henri Poincare (C) Non Linear Analysis, **29** (2012), 667 – 681.

[5] Almeida, L.; Demongeot, J.. *Predictive power of ”a minima” models in biology*, Acta Bioth., **60** (2012), 3–19.

[6] C Dahmann and K Basler. *Opposing transcriptional outputs of hedgehog signaling and engrailed control compartmental cell sorting at the drosophila a/p boundary*, Cell, **100(4)** (2000), 411-422.

[7] R. Farhadifar, J. Röper, B. Aigouy, S. Eaton, and F. Jülicher. *The infuence of cell mechanics, cell-cell interactions, and proliferation on epithelial packing*, Curr Biol, **17(24)** (2007), 2095-2104.

[8] Gettings, M.; Serman, F.; Rousset, R.; Bagnerini, P.; Almeida, L. and Noselli, S. *JNK signalling controls remodelling of the segment boundary through cell reprogramming during Drosophila morphogenesis*, PLoS Biol, **8** (2010).

[9] T. Lecuit and P. Lenne. *Cell surface mechanics and the control of cell shape, tissue patterns and morphogenesis*, Nat Rev Mol Cell Biol, **8(8)** (2007), 633-644.

[10] Mitchell, I. M.. *The flexible, extensible and efficient toolbox of level set methods*, J. Sci. Comput., **35** (2008), 300-329.

E-mail address: `luis@ann.jussieu.fr`
E-mail address: `bagnerini@dime.unige.it`

LOW MACH NUMBER LIMITS TO THE NAVIER-STOKES-SMOLUCHOWSKI SYSTEM

Joshua Ballew

Department of Mathematics
University of Maryland, College Park
College Park, MD, 20904, USA

Abstract. This article presents a general dimensionless scaling of the Navier-Stokes-Smoluchowski system describing interactions between particles and a compressible fluid. Two low Mach number limits are investigated. The first limit is a low stratification limit for which the Froude number is scaled as the square root of the Mach number; the second is a strong stratification limit for which the Froude and Mach numbers are scaled the same. We see that as the Mach number goes to zero in the low stratification case, the solutions to the system converge in appropriate spaces to constant mass densities and weakly to a velocity field satisfying the incompressibility condition. For the strong stratification case, we see for an external force depending only on the vertical coordinate that the solutions converge to densities depending only on the vertical component and a velocity field satisfying the anelastic condition. Finally, we investigate bounds and convergences for the strong stratification case supporting the formal calculations.

1. **Introduction.** The state of fluid-particle-interaction flows is characterized by the following macroscopic variables: the total mass density $\varrho(t,x)$, the velocity field $\mathbf{u}(t,x)$, and the density of particles dispersed in the mixture $\eta(t,x)$, which depend on the Eulerian spatial coordinate $x \in \Omega \subset \mathbb{R}^3$ and on time $t \in (0,\infty)$. The governing equations express the conservation of mass, the balance of momentum, and the balance of particle densities often referred to as the *Smoluchowski equation*:

$$\partial_t \varrho + \operatorname{div}_x (\varrho \mathbf{u}) = 0 \tag{1.1}$$

$$\begin{aligned} &\partial_t (\varrho \mathbf{u}) + \operatorname{div}_x (\varrho \mathbf{u} \otimes \mathbf{u}) + \nabla_x \left(p_F(\varrho) + \frac{D}{\zeta}\eta \right) \\ &= \mu \triangle_x \mathbf{u} + \lambda \nabla_x \operatorname{div}_x \mathbf{u} - (\eta + \beta\varrho) \nabla_x \Phi \end{aligned} \tag{1.2}$$

$$\partial_t \eta + \operatorname{div}_x (\eta (\mathbf{u} - \zeta \nabla_x \Phi)) - D \triangle_x \eta = 0 \tag{1.3}$$

where $p_F(\varrho) = a\varrho^\gamma$ for some $a > 0$, $\gamma > \frac{3}{2}$, and $\beta \neq 0$. We also assume a bounded $C^{2,\nu}$ spatial domain Ω. The fluid is also assumed to be Newtonian so that the stress tensor is given by

$$\mathbb{S} = \mu(\nabla_x \mathbf{u} + \nabla_x \mathbf{u}^T) + \lambda \operatorname{div}_x \mathbf{u} \mathbb{I}.$$

Also, the viscosity coefficients μ and λ, the drag coefficient ζ, and the dispersion coefficient D are assumed to be constant, and Φ is a given external potential that is

2000 *Mathematics Subject Classification.* Primary: 35Q30, 76N99; Secondary: 46E35.

Key words and phrases. Fluid-particle interaction, low Mach number, low stratification singular limit, strong stratification singular limit, compressible and viscous fluid.

taken to be nonnegative. The system (1.1)-(1.3) is supplemented by the following boundary and initial conditions:

$$\mathbf{u} = D\nabla_x \eta \cdot \mathbf{n} + \zeta\eta\nabla_x\Phi \cdot \mathbf{n} = 0 \text{ on } (0,T)\times\partial\Omega \tag{1.4}$$

$$0 \le \varrho(0,x) = \varrho_0 \in L^\gamma(\Omega) \tag{1.5}$$

$$(\varrho\mathbf{u})(0,x) = \mathbf{m}_0 \in L^{6/5}(\Omega;\mathbb{R}^3) \tag{1.6}$$

$$0 \le \eta(0,x) = \eta_0 \in L^2(\Omega). \tag{1.7}$$

We define the energy

$$\mathcal{E}(t) := \int_\Omega \frac{1}{2}\varrho|\mathbf{u}|^2 + \frac{a}{\gamma-1}\varrho^\gamma + \frac{D}{\zeta}\eta\ln\eta + (\beta\varrho+\eta)\Phi \mathrm{d}x(t) \tag{1.8}$$

and require that

$$\frac{\mathrm{d}E}{\mathrm{d}t} + \int_\Omega \mu|\nabla_x\mathbf{u}|^2 + \lambda|\mathrm{div}_x\mathbf{u}|^2 + \left|\frac{2D}{\sqrt{\zeta}}\nabla_x\sqrt{\eta} + \sqrt{\eta}\nabla_x\Phi\right|^2 \mathrm{d}x \le 0.$$

In addition, we require that the spatial domain Ω and external potential Φ obey the following hypotheses, called the *confinement hypotheses*:

Definition 1.1. Let $\Omega \subset \mathbb{R}^3$ be a $C^{2,\nu}$ domain with $\nu > 0$ and $\Phi : \Omega \to \mathbb{R}_0^+$ with $\inf_{x\in\Omega}\Phi(x) = 0$. (Ω,Φ) satisfies the **Confinement Hypotheses (HC)** if and only if

- If Ω is bounded, Φ is bounded and Lipschitz continuous on $\overline{\Omega}$.
- If Ω is unbounded, $\Phi \in W^{1,\infty}_{\text{loc}}(\Omega)$, $e^{-\Phi/2} \in L^1(\Omega)$ and

  $$|\Delta_x\Phi(x)| \le c_1|\nabla_x\Phi(x)| \le c_2\Phi(x)$$

  for $|x|$ greater than some large R.

In [4] it is shown using an artificial pressure and time-discretization approximation that a *renormalized weak solution* exists. In [3], a weak-strong uniqueness result is shown on the NSS system; that is, if there is a weak solution of a certain regularity class, the the weak solution is unique.

The rest of the paper is dedicated to examining certain approximations to the compressible NSS system in the form of singular limits for bounded spatial domains Ω. In particular, we look at conditions for which the speed of the fluid flow is small compared to the speed of sound in the fluid, also known as the low Mach number case. Under a low stratification condition of the scaling of the system, the solutions converge to a solution of the mathematically simpler incompressible fluid model as the Mach number approaches zero. In the strong stratification case, the solutions will converge to functions obeying the anelastic condition, if we assume that the external force depends only on the vertical component of position, physically realized for buoyancy and gravity near the surface of the earth or other similar body. Both of these problems involve using bounds from the energy inequality for the systems to provide estimates that allow us to show the convergence of the solutions. These techniques are motivated by the work in [5, 6, 7, 8].

2. **Dimensionless Scaling.** For each parameter α (time, length, mass, density, pressure, etc.), we define a reference value α_{ref} and then define the dimesionless value

$$\alpha' := \frac{\alpha}{\alpha_{\text{ref}}}.$$

By using the chain rule and basic differentiation properties, the NSS system in terms of the dimensionless parameters and values becomes (with the prime marks omitted)

$$\mathrm{Sr}\partial_t \varrho + \mathrm{div}_x(\varrho \mathbf{u}) = 0 \tag{2.9}$$

$$\begin{aligned} &\mathrm{Sr}\partial_t(\varrho\mathbf{u}) + \mathrm{div}_x(\varrho\mathbf{u}\otimes\mathbf{u}) + \frac{1}{\mathrm{Ma}^2}\nabla_x\left(a\varrho^\gamma + \mathrm{Pc}\frac{D}{\zeta}\eta\right) \\ &= \frac{1}{\mathrm{Re}}(\mu\Delta_x\mathbf{u} + \lambda\nabla_x\mathrm{div}_x\mathbf{u}) - \frac{1}{\mathrm{Fr}^2}(\beta\varrho + \mathrm{Dc}\eta)\nabla_x\Phi \end{aligned} \tag{2.10}$$

$$\mathrm{Sr}\partial_t\eta + \mathrm{div}_x(\eta\mathbf{u}) - \mathrm{Za}\mathrm{div}_x(\zeta\eta\nabla_x\Phi) - \mathrm{Da}D\Delta_x\eta = 0 \tag{2.11}$$

with the scaled energy inequality

$$\begin{aligned} &\mathrm{Sr}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \frac{\mathrm{Ma}^2}{2}\varrho|\mathbf{u}|^2 + \frac{a}{\gamma-1}\varrho^\gamma + \mathrm{Pc}\frac{D\eta}{\zeta}\ln\eta + \frac{\mathrm{Ma}^2}{\mathrm{Fr}^2}(\beta\varrho + \mathrm{Dc}\eta)\Phi\mathrm{d}x \\ &+ \int_\Omega \mathrm{PcDa}D^2\frac{|\nabla_x\eta|^2}{\zeta\eta} + 2\mathrm{Za}D\nabla_x(\eta)\cdot\nabla_x\Phi + \frac{\mathrm{Za}^2}{\mathrm{Da}}\zeta\eta|\nabla_x\Phi|^2\mathrm{d}x \\ &+ \int_\Omega \frac{\mathrm{Ma}^2}{\mathrm{Re}}\mathbb{S}(\nabla_x\mathbf{u}):\nabla_x\mathbf{u}\mathrm{d}x \le 0 \end{aligned}$$

with the unitless coefficients defined in the following table.

Sr:= $\frac{L_{ref}}{\mathbf{u}_{ref}t_{ref}}$	Ma:= $\frac{\mathbf{u}_{ref}}{\sqrt{p_{F_{ref}}/\varrho_{ref}}}$	Re:= $\frac{\varrho_{ref}\mathbf{u}_{ref}L_{ref}}{\mu_{ref}}$
Fr:= $\frac{\mathbf{u}_{ref}}{\sqrt{L_{ref}f_{ref}}}$	Za:= $\frac{\zeta_{ref}f_{ref}}{\mathbf{u}_{ref}}$	Da:= $\frac{D_{ref}}{L_{ref}\mathbf{u}_{ref}}$
Pc:= $\frac{p_{P_{ref}}}{p_{F_{ref}}}$	Dc:= $\frac{\eta_{ref}}{\varrho_{ref}}$.	

Table 2.1: Definitions of the Dimensionless Parameters

3. **Low Stratification Limit.** The scaled low stratification system we consider for each fixed $\varepsilon > 0$ is

$$\partial_t\varrho_\varepsilon + \mathrm{div}_x(\varrho_\varepsilon\mathbf{u}_\varepsilon) = 0 \tag{3.12}$$

$$\begin{aligned} &\varepsilon^2[\partial_t(\varrho_\varepsilon\mathbf{u}_\varepsilon) + \mathrm{div}_x(\varrho_\varepsilon\mathbf{u}_\varepsilon\otimes\mathbf{u}_\varepsilon)] + \nabla_x\left(a\varrho_\varepsilon^\gamma + \frac{D}{\zeta}\eta_\varepsilon\right) \\ &= \varepsilon^2(\mu\Delta_x\mathbf{u}_\varepsilon + \lambda\nabla_x\mathrm{div}_x\mathbf{u}_\varepsilon) - \varepsilon(\beta\varrho_\varepsilon + \eta_\varepsilon)\nabla_x\Phi \end{aligned} \tag{3.13}$$

$$\partial_t\eta_\varepsilon + \mathrm{div}_x(\eta_\varepsilon\mathbf{u}_\varepsilon) - \varepsilon\mathrm{div}_x(\zeta\eta_\varepsilon\nabla_x\Phi) - D\Delta_x\eta_\varepsilon = 0 \tag{3.14}$$

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \frac{\varepsilon^2}{2}\varrho_\varepsilon|\mathbf{u}_\varepsilon|^2 + \frac{a}{\gamma-1}\varrho_\varepsilon^\gamma + \frac{D\eta_\varepsilon}{\zeta}\ln\eta_\varepsilon + \varepsilon(\beta\varrho_\varepsilon + \eta_\varepsilon)\Phi\mathrm{d}x \\ &+ \int_\Omega D^2\frac{|\nabla_x\eta_\varepsilon|^2}{\zeta\eta_\varepsilon} + 2\varepsilon D\nabla_x\eta_\varepsilon\cdot\nabla_x\Phi + \varepsilon^2\zeta\eta_\varepsilon|\nabla_x\Phi|^2\mathrm{d}x \\ &+ \int_\Omega \varepsilon^2\mathbb{S}(\nabla_x\mathbf{u}_\varepsilon):\nabla_x\mathbf{u}_\varepsilon\mathrm{d}x \le 0. \end{aligned} \tag{3.15}$$

To rigorously derive the limit for the low stratification case, we begin by noting that from the results of [4], for each $\varepsilon > 0$, we have solutions $\{\varrho_\varepsilon, \mathbf{u}_\varepsilon, \eta_\varepsilon\}$ in the following sense:

Definition 3.1. We say that $\{\varrho_\varepsilon, \mathbf{u}_\varepsilon, \eta_\varepsilon\}$ is a *renormalized weak solution to the scaled low stratification NSS system* if and only if

- $\varrho_\varepsilon \geq 0$ and $\mathbf{u}_\varepsilon$ form a renormalized solution of the scaled continuity equation, i.e.,
$$\begin{aligned}&\int_0^T \int_\Omega B(\varrho_\varepsilon)\partial_t\varphi + B(\varrho_\varepsilon)\mathbf{u}_\varepsilon \cdot \nabla_x\varphi - b(\varrho_\varepsilon)\mathrm{div}_x\mathbf{u}_\varepsilon\varphi \mathrm{d}x\mathrm{d}t\\ &= -\int_\Omega B(\varrho_0)\varphi(0,\cdot)\mathrm{d}x \qquad (3.16)\end{aligned}$$
where $b \in L^\infty \cap C[0,\infty)$, $B(\varrho) := B(1) + \int_1^\varrho \frac{b(z)}{z^2}\mathrm{d}z$.
- The scaled momentum balance holds in the sense of distribution.
- $\eta_\varepsilon \geq 0$ is a weak solution of the scaled Smoluchowski equation.
- The scaled energy inequality (3.15) is satisfied.

We next define the *low stratification target system.*

Definition 3.2. $\{\varrho^{(1)}, \overline{\mathbf{u}}, \eta^{(1)}\}$ solve the *low stratification target system* if and only if
$$\mathrm{div}_x\overline{\mathbf{u}} = 0 \text{ weakly on } (0,T)\times\Omega,$$
$$\begin{aligned}&\int_0^T \int_\Omega \overline{\varrho}\overline{\mathbf{u}} \cdot \partial_t\mathbf{v} + \overline{\varrho}\overline{\mathbf{u}} \otimes \overline{\mathbf{u}} : \nabla_x\mathbf{v}\mathrm{d}x\mathrm{d}t\\ &= \int_0^T \int_\Omega (\mu\nabla_x\overline{\mathbf{u}} - (\beta r + s)\nabla_x\Phi)\cdot\mathbf{v}\mathrm{d}x\mathrm{d}t - \int_\Omega \overline{\varrho}\overline{\mathbf{u}}\cdot\mathbf{v}(0,\cdot)\mathrm{d}x,\end{aligned}$$
for any divergence-free text function $\mathbf{v}$ and
$$r = -\frac{1}{a\gamma\overline{\varrho}^{\gamma-1}}\left[(\beta\overline{\varrho} + \overline{\eta})\Phi + \frac{D}{\zeta}s\right]$$
weakly where $\overline{\varrho}$ and $\overline{\eta}$ are uniform fluid and particle densities, respectively, with the same total masses as the initial data.

We are now in a position to state the main theorem of this section.

Theorem 3.3. *Let (Ω, Φ) satisfy the confinement hypothesis and for each $\varepsilon > 0$, assume $\{\varrho_\varepsilon, \boldsymbol{u}_\varepsilon, \eta_\varepsilon\}$ is a solution of the low stratification system in the sense of Definition 3.1. Assume the initial data can be expressed as follows:*
$$\varrho_\varepsilon(0,\cdot) = \varrho_{\varepsilon,0} = \overline{\varrho} + \varepsilon\varrho^{(1)}_{\varepsilon,0},\ \boldsymbol{u}_\varepsilon(0,\cdot) = \boldsymbol{u}_{\varepsilon,0},\ \text{and } \eta_\varepsilon(0,\cdot) = \eta_{\varepsilon,0} = \overline{\eta} + \varepsilon\eta^{(1)}_{\varepsilon,0}.$$
where $\overline{\varrho}, \overline{\eta}$ are the spatially uniform densities on Ω. Assume also that as $\varepsilon \to 0$,
$$\varrho^{(1)}_{\varepsilon,0} \rightharpoonup \varrho^{(1)}_0,\ \boldsymbol{u}_{\varepsilon,0} \rightharpoonup \overline{\boldsymbol{u}}_0, \eta^{(1)}_{\varepsilon,0} \rightharpoonup \eta^{(1)}_0$$
weakly-$*$ in $L^\infty(\Omega)$ or $L^\infty(\Omega;\mathbb{R}^3)$ as the case may be. Then up to a subsequence and letting $q := \min\{\gamma, 2\}$,
$$\varrho_\varepsilon \to \overline{\varrho} \text{ in } C([0,T];L^1(\Omega)) \cap L^\infty(0,T;L^q(\Omega))$$
$$\eta_\varepsilon \to \overline{\eta} \text{ in } L^2(0,T;L^2(\Omega))$$
$$\boldsymbol{u}_\varepsilon \to \overline{\boldsymbol{u}} \text{ weakly in } L^2(0,T;W^{1,2}(\Omega;\mathbb{R}^3))$$
and
$$\varrho^{(1)}_\varepsilon = \frac{\varrho_\varepsilon - \overline{\varrho}}{\varepsilon} \to \varrho^{(1)} \text{weakly-}* \text{ in } L^\infty(0,T;L^q(\Omega))$$
$$\eta^{(1)}_\varepsilon = \frac{\eta_\varepsilon - \overline{\eta}}{\varepsilon} \to \eta^{(1)} \text{ weakly in } L^2(0,T;L^2(\Omega))$$
where $\{\varrho^{(1)}, \overline{\boldsymbol{u}}, \eta^{(1)}\}$ solve the target system mentioned previously.

Proof. For the proof, the reader may consult [2]. □

4. **Strong Stratification Limit.** The formal calculations for the strong stratification limit as $\varepsilon \to 0$ follow the same procedure as for the low stratification limit. We use the following scaling: Ma is taken to be a small parameter $\varepsilon > 0$, Za and Da are taken to be ε^{-1}, Fr is taken to be ε, and other parameters are taken to be of order 1. We also assume that $\Phi = gx_3$ where g is a constant (gravity/buoyancy). Thus, the scaled NSS system becomes

$$\partial_t \varrho_\varepsilon + \mathrm{div}_x(\varrho_\varepsilon \mathbf{u}_\varepsilon) = 0 \tag{4.17}$$

$$\begin{aligned}&\varepsilon^2[\partial_t(\varrho_\varepsilon \mathbf{u}_\varepsilon) + \mathrm{div}_x(\varrho_\varepsilon \mathbf{u}_\varepsilon \otimes \mathbf{u}_\varepsilon)] + \nabla_x \left(a\varrho_\varepsilon^\gamma + \frac{D}{\zeta}\eta_\varepsilon \right) \\ &= \varepsilon^2(\mu\Delta_x \mathbf{u}_\varepsilon + \lambda\nabla_x \mathrm{div}_x \mathbf{u}_\varepsilon) - (\beta\varrho_\varepsilon + \eta_\varepsilon)\nabla_x \Phi\end{aligned} \tag{4.18}$$

$$\varepsilon\,[\partial_t \eta_\varepsilon + \mathrm{div}_x(\eta_\varepsilon \mathbf{u}_\varepsilon)] - \mathrm{div}_x(\zeta\eta_\varepsilon\nabla_x\Phi) - D\Delta_x\eta_\varepsilon = 0 \tag{4.19}$$

$$\begin{aligned}&\varepsilon\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \frac{\varepsilon^2}{2}\varrho_\varepsilon|\mathbf{u}_\varepsilon|^2 + \frac{a}{\gamma-1}\varrho_\varepsilon^\gamma + \frac{D\eta_\varepsilon}{\zeta}\ln\eta_\varepsilon + (\beta\varrho_\varepsilon + \eta_\varepsilon)\Phi \mathrm{d}x \\ &+ \varepsilon\int_\Omega \varepsilon^2 \mathbb{S}(\nabla_x \mathbf{u}_\varepsilon) : \nabla_x \mathbf{u}_\varepsilon \mathrm{d}x + \int_\Omega \left| D\frac{\nabla_x \eta_\varepsilon}{\sqrt{\zeta\eta_\varepsilon}} + \sqrt{\zeta\eta_\varepsilon}\nabla_x\Phi\right|^2 \mathrm{d}x \le 0.\end{aligned} \tag{4.20}$$

Now, assuming $\{\varrho_\varepsilon, \mathbf{u}_\varepsilon, \eta_\varepsilon\}$ have the following expansions

$$\varrho_\varepsilon = \tilde{\varrho} + \sum_{i=1}^\infty \varepsilon^i \varrho_\varepsilon^{(i)}$$

$$\eta_\varepsilon = \tilde{\eta} + \sum_{i=1}^\infty \varepsilon^i \eta_\varepsilon^{(i)}$$

$$\mathbf{u}_\varepsilon = \tilde{\mathbf{u}} + \sum_{i=1}^\infty \varepsilon^i \mathbf{u}_\varepsilon^{(i)}$$

we substitue into (4.17)-(4.20) and formally obtain the target system

$$g\tilde{\eta} = -\frac{D}{\zeta}\frac{\mathrm{d}\tilde{\eta}}{\mathrm{d}x_3}$$

$$\frac{\mathrm{d}}{\mathrm{d}x_3}[a\tilde{\varrho}^\gamma] = -\beta g\tilde{\varrho}$$

$$\mathrm{div}_x(\tilde{\varrho}\tilde{\mathbf{u}}) = 0$$

$$\tilde{\varrho}\partial_t\tilde{\mathbf{u}} + \mathrm{div}_x(\tilde{\varrho}\tilde{\mathbf{u}} \otimes \tilde{\mathbf{u}}) + \nabla_x\Pi = \mu\Delta_x\tilde{\mathbf{u}} + \lambda\nabla_x\mathrm{div}_x\tilde{\mathbf{u}} - \left(\beta\varrho^{(2)} + \eta^{(2)}\right)\nabla_x\Phi.$$

For the strong stratification scaling, we have the following weak formulation:

Definition 4.1. We say that $\{\varrho_\varepsilon, \mathbf{u}_\varepsilon, \eta_\varepsilon\}$ form a *renormalized weak solution to the scaled strong stratification NSS system* if and only if

- $\varrho_\varepsilon \ge 0$ and $\mathbf{u}_\varepsilon$ form a renormalized solution of the scaled continuity equation, i.e.,

$$\begin{aligned}&\int_0^T \int_\Omega B(\varrho_\varepsilon)\partial_t\varphi + B(\varrho_\varepsilon)\mathbf{u}_\varepsilon \cdot \nabla_x\varphi - b(\varrho_\varepsilon)\mathrm{div}_x\mathbf{u}_\varepsilon\varphi \mathrm{d}x\mathrm{d}t \\ &= -\int_\Omega B(\varrho_0)\varphi(0,\cdot)\mathrm{d}x\end{aligned} \tag{4.21}$$

  where $b \in L^\infty \cap C[0,\infty)$, $B(\varrho) := B(1) + \int_1^\varrho \frac{b(z)}{z^2}\mathrm{d}z$.

- The scaled momentum balance holds in the sense of distributions.
- $\eta_\varepsilon \geq 0$ is a weak solution of the scaled Smoluchowski equation.
- The scaled energy inequality (4.20) is satisfied.

Note that for this scaling, we assume that $\Phi = gx_3$, where x_3 is the vertical coordinate, and g is a constant greater than zero. We also define the target system.

Definition 4.2. $\{\tilde{\varrho}, \tilde{\mathbf{u}}, \tilde{\eta}, \varrho^{(2)}, \eta^{(2)}\}$ solve the *strong stratification target system* if and only if:

-
$$\int_0^T \int_\Omega \tilde{\varrho}\tilde{\mathbf{u}} \cdot \nabla_x \phi \mathrm{d}x\mathrm{d}t = 0 \tag{4.22}$$
for all $\phi \in C_C^\infty((0,T) \times \Omega)$,
-
$$g\tilde{\eta} = -\frac{D}{\zeta}\frac{\mathrm{d}\tilde{\eta}}{\mathrm{d}x_3} \tag{4.23}$$
$$\frac{\mathrm{d}}{\mathrm{d}x_3}\left[a\tilde{\varrho}^\gamma\right] = -\beta g \tilde{\varrho} \tag{4.24}$$
with the conditions
$$\int_\Omega \tilde{\varrho}\mathrm{d}x = \int_\Omega \varrho_0 \mathrm{d}x$$
$$\int_\Omega \tilde{\eta}\mathrm{d}x = \int_\Omega \eta_0 \mathrm{d}x,$$
-
$$\begin{aligned}&\int_0^T \int_\Omega \tilde{\varrho}\tilde{\mathbf{u}} \cdot \mathbf{w} + \tilde{\varrho}\tilde{\mathbf{u}} \otimes \tilde{\mathbf{u}} : \nabla_x \mathbf{w} \mathrm{d}x\mathrm{d}t \\ &= \int_0^T \int_\Omega \mu \nabla_x \tilde{\mathbf{u}} \nabla_x \mathbf{w} - \left(\beta\varrho^{(2)} + \eta^{(2)}\right) \nabla_x \Phi \cdot \mathbf{w} \mathrm{d}x\mathrm{d}t\end{aligned} \tag{4.25}$$
for all $\mathbf{w} \in C_C^\infty((0,T) \times \Omega; \mathbb{R}^3)$ such that $\mathrm{div}_x \mathbf{w} = 0$.

Much like for the low stratification limit, many of the bounds and convergences used in the analysis arise from the free energies defined as

$$E_F(\varrho, \tilde{\varrho}) := \frac{a}{\gamma - 1}\varrho^\gamma - (\varrho - \tilde{\varrho})\frac{a\gamma}{\gamma - 1}\tilde{\varrho}^{\gamma - 1} - \frac{a}{\gamma - 1}\tilde{\varrho}^\gamma$$

$$E_P(\eta, \tilde{\eta}) := \frac{D}{\zeta}\eta \ln \eta - \frac{D}{\zeta}(\eta - \tilde{\eta})(\ln \tilde{\eta} + 1) - \frac{D}{\zeta}\tilde{\eta}\ln\tilde{\eta},$$

and the resulting inequality formed from these and the energy inequality:

$$\begin{aligned}&\int_\Omega \frac{1}{2}\varrho_\varepsilon |\mathbf{u}_\varepsilon|^2 + \frac{1}{\varepsilon^2}\left[E_F(\varrho_\varepsilon, \tilde{\varrho}) + E_P(\eta_\varepsilon, \tilde{\eta})\right] \mathrm{d}x(T) \\ &\int_0^T \int_\Omega \mathbb{S}(\nabla_x \mathbf{u}_\varepsilon) : \nabla_x \mathbf{u}_\varepsilon \mathrm{d}x\mathrm{d}t + \frac{1}{\varepsilon^3}\int_0^T \int_\Omega \left| \frac{D\nabla_x \eta_\varepsilon}{\sqrt{\zeta\eta_\varepsilon}} + \sqrt{\zeta\eta_\varepsilon}\nabla_x \Phi \right|^2 \mathrm{d}x\mathrm{d}t \\ &\leq \int_\Omega \frac{1}{2}\varrho_0 |\mathbf{u}_0|^2 + \frac{1}{\varepsilon^2}[E_F(\varrho_0, \tilde{\varrho}) + E_P(\eta_0, \tilde{\eta})]\mathrm{d}x.\end{aligned} \tag{4.26}$$

Next, we define the essential and residual sets:

$$\begin{aligned}
&\mathcal{O}_{\text{ess}} := \{(\varrho,\eta)\in\mathbb{R}^2 | \tilde{\varrho}/2 \le \varrho \le 2\tilde{\varrho}, \tilde{\eta}/2 \le \eta \le 2\tilde{\eta}\}\\
&\mathcal{M}^{\varepsilon}_{\text{ess}} := \{(x,t)\in(0,T)\times\Omega | (\varrho_\varepsilon(t,x),\eta_\varepsilon(t,x))\in\mathcal{O}_{\text{ess}}\}\\
&\mathcal{M}^{\varepsilon}_{\text{res}} := ((0,T)\times\Omega) - \mathcal{M}^{\varepsilon}_{\text{ess}}
\end{aligned}$$

Thus, by using (4.26), assuming appropriate bounds on the initial data, we obtain that

$$\begin{aligned}
&\{\sqrt{\varrho_\varepsilon}\mathbf{u}_\varepsilon\}_{\varepsilon>0} \in_b L^\infty(0,T;L^2(\Omega;\mathbb{R}^3))\\
&\|[\varrho_\varepsilon - \tilde{\varrho}]\text{ess}\|_{L^\infty(0,T;L^2(\Omega))} \le \varepsilon^2 c\\
&\|[\eta_\varepsilon - \tilde{\eta}]\text{ess}\|_{L^\infty(0,T;L^2(\Omega))} \le \varepsilon^2 c\\
&\{\mathbf{u}_\varepsilon\}_{\varepsilon>0} \in_b L^2(0,T;W^{1,2}_0(\Omega;\mathbb{R}^3))\\
&\left\|\frac{D\nabla_x\eta_\varepsilon}{\sqrt{\zeta\eta_\varepsilon}} + \sqrt{\zeta\eta_\varepsilon}\nabla_x\Phi\right\|_{L^2(0,T;L^2(\Omega;\mathbb{R}^3))} \le \varepsilon^3 c\\
&\left\{\left[\frac{\varrho_\varepsilon - \tilde{\varrho}}{\varepsilon}\right]_{\text{ess}}\right\}_{\varepsilon>0} \in_b L^\infty(0,T;L^2(\Omega))\\
&\left\{\left[\frac{\eta_\varepsilon - \tilde{\eta}}{\varepsilon}\right]_{\text{ess}}\right\}_{\varepsilon>0} \in_b L^\infty(0,T;L^2(\Omega))
\end{aligned}$$

and since the measure of the residual set goes as ε^2 for each fixed t, we have

$$\begin{aligned}
&\|[\varrho_\varepsilon]\text{res}\|_{L^\infty(0,T;L^\gamma(\Omega))} \le \varepsilon^2 c\\
&\{\varrho_\varepsilon\mathbf{u}_\varepsilon\}_{\varepsilon>0} \in_b L^\infty(0,T;L^{2q/q+1}(\Omega;\mathbb{R}^3)) \cap L^{6q/q+6}(\Omega;\mathbb{R}^3))
\end{aligned}$$

where $q := \min\{2, q\}$. Thus, we have the existence of $\varrho^{(1)}, \eta^{(1)} \in L^\infty(0,T;L^2(\Omega))$ and $\tilde{\mathbf{u}} \in L^2(0,T;W^{1,2}_0(\Omega;\mathbb{R}^3))$ such that up to subsequences

$$\begin{aligned}
&\varrho_\varepsilon \to \tilde{\varrho} \text{ strongly in } L^\infty(0,T;L^q(\Omega))\\
&\eta_\varepsilon \to \tilde{\eta} \text{ strongly in } L^\infty(0,T;L^2(\Omega))\\
&\mathbf{u}_\varepsilon \rightharpoonup \tilde{\mathbf{u}} \text{ weakly in } L^2(0,T;W^{1,2}_0(\Omega;\mathbb{R}^3))\\
&\frac{\varrho_\varepsilon - \tilde{\varrho}}{\varepsilon} \rightharpoonup \varrho^{(1)} \text{ weakly-}* \text{ in } L^\infty(0,T;L^q(\Omega))\\
&\frac{\eta_\varepsilon - \tilde{\eta}}{\varepsilon} \rightharpoonup \varrho^{(1)} \text{ weakly-}* \text{ in } L^\infty(0,T;L^2(\Omega)).
\end{aligned}$$

Now, we are in a position to state the main result of this section:

Theorem 4.3. *Let (Ω,Φ) satisfy the confinement hypothesis and for each $\varepsilon > 0$, assume $\{\varrho_\varepsilon, \boldsymbol{u}_\varepsilon, \eta_\varepsilon\}$ solves the scaled strong stratification system in the sense of Definition 4.1. Assume the initial data can be expressed as follows:*

$$\varrho_\varepsilon(0,\cdot) = \varrho_{\varepsilon,0} = \tilde{\varrho} + \varepsilon\varrho^{(1)}_{\varepsilon,0},\ \boldsymbol{u}_\varepsilon(0,\cdot) = \boldsymbol{u}_{\varepsilon,0},\ \textit{and}\ \eta_\varepsilon(0,\cdot) = \eta_{\varepsilon,0} = \tilde{\eta} + \varepsilon\eta^{(1)}_{\varepsilon,0}.$$

where $\tilde{\varrho}, \tilde{\eta}$ are the densities defined by (4.24)-(4.23). Assume also that as $\varepsilon \to 0$,

$$\varrho^{(1)}_{\varepsilon,0} \rightharpoonup \varrho^{(1)}_0,\ \boldsymbol{u}_{\varepsilon,0} \rightharpoonup \tilde{\boldsymbol{u}}_0,\ \eta^{(1)}_{\varepsilon,0} \rightharpoonup \eta^{(1)}_0$$

weakly-∗ in $L^\infty(\Omega)$ or $L^\infty(\Omega;\mathbb{R}^3)$ as the case may be. Then up to a subsequence and letting $q := \min\{\gamma, 2\}$,

$$\varrho_\varepsilon \to \tilde{\varrho} \ \textit{in}\ C([0,T];L^1(\Omega)) \cap L^\infty(0,T;L^q(\Omega))$$

$$\eta_\varepsilon \to \tilde{\eta} \textit{ in } L^2(0,T;L^2(\Omega))$$
$$\boldsymbol{u}_\varepsilon \to \tilde{\boldsymbol{u}} \textit{ weakly in } L^2(0,T;W^{1,2}(\Omega;\mathbb{R}^3))$$

where $\{\tilde{\varrho}, \tilde{\boldsymbol{u}}, \tilde{\eta}\}$ *solve the target system* (4.22)-(4.25).

Proof. The result follows from the bounds listed above and analysis similar to that done in Section 3 and in [8]. For the details, see [1] □

Acknowledgments. The author wishes to acknowledge the support of Konstantina Trivisa, under whose direction he has been researching. He also acknowledges support in part by the *John Osborn Memorial Fellowship*, the *Ann G. Wylie Dissertation Fellowship*, and the National Science Foundation under the grants DMS 0807815 and DMS 1109397.

REFERENCES

[1] J. Ballew, A strong stratification limit for the Navier-Stokes-Smoluchowski system. In preparation.

[2] J. Ballew and K. Trivisa, Suitable weak solutions and low stratification singular limit for a fluid particle interaction model. *Quart. Appl. Math.* **70**:469-494, 2012.

[3] J. Ballew and K. Trivisa, Weakly dissipative solutions and weak-strong uniqueness for the Navier-Stokes-Smoluchowski system. *Nonlinear Analysis Series A: Theory, Methods & Applications.* **91**:1-19, 2013.

[4] J.A. Carrillo, T. Karper, and K. Trivisa. On the dynamics of a fluid-particle interaction model: The bubbling regime. *Nonlinear Analysis*, **74**:2778-2801, 2011.

[5] B. Desjardins, E. Grenier, P.-L. Lions, and N. Masmoudi. Incompressible limit for solutions of the isentropic Navier-Stokes equations with Dirichlet boundary conditions. *J. Math. Pures Appl.*, **78**:461-471, 1999.

[6] E. Feireisl. Flows of viscous compressible fluids under strong stratification: incompressible limits for long-range potential forces. *Mathematical Models and Methods in Applied Sciences*, **21**:7-27, 2011.

[7] E. Feireisl and A. Novotný. On the low Mach number limit for the full Navier-Stokes-Fourier system. *Arch. Rational Mech. Anal.*, **186**:77-107, 2007.

[8] E. Feireisl and A. Novotný. *Singular limits in thermodynamics of viscous fluids.* Birkhauser, Basel, 2009.

E-mail address: `jballew@math.umd.edu`

METHOD FOR SOLVING A STOCHASTIC CONSERVATION LAW

Caroline Bauzet

Laboratoire de Mathématiques et Applications de Pau
UMR-CNRS 5142
IPRA BP 1155
64013 Pau Cedex, France

Abstract. This paper presents techniques introduced in a joint work with G. Vallet and P. Wittbold for solving the Cauchy problem for a multi-dimensional nonlinear conservation law with stochastic perturbation [2]. We propose here to present main difficulties met in the use of deterministic tools for studying stochastic scalar conservation law and alternative methods.

1. **Introduction.** We are interested in the formal stochastic nonlinear conservation law of type:

$$\mathrm{du} - \mathrm{div}(\vec{\mathbf{f}}(u))\mathrm{dt} = h(u)\mathrm{dw} \quad \text{in } \Omega \times \mathbb{R}^d \times]0,T[, \tag{1}$$

with an initial condition u_0 and $d \geq 1$.

In the sequel we assume that T is a positive number, that $Q =]0,T[\times\mathbb{R}^d$ and that $W = \{w_t, \mathcal{F}_t; 0 \leq t \leq T\}$ denotes a standard adapted one-dimensional continuous Brownian motion, defined on the classical Wiener space $(\Omega, \mathcal{F}, P)$. Note that we consider for convenience a real-valued noise and that the present work could be generalized to a class of multi-dimensional noise. Let us assume that

H$_1$: $\vec{\mathbf{f}} = (f_1, .., f_d) : \mathbb{R} \to \mathbb{R}^d$ is a Lipschitz-continuous function with $\vec{\mathbf{f}}(0) = \vec{0}$.
H$_2$: $h : \mathbb{R} \to \mathbb{R}$ is a Lipschitz-continuous function with $h(0) = 0$.
H$_3$: $u_0 \in L^2(\mathbb{R}^d)$.

We propose to present tools for showing existence and uniqueness of the stochastic entropy solution to the above-mentioned problem. Our aim is to adapt the known methods for deterministic first-order nonlinear P.D.E. to noise perturbed ones.

Note that, even in the deterministic case, a weak solution to a nonlinear scalar conservation law is not unique in general. One needs to introduce the notion of entropy solution in order to select the "physical solution".

Only a few papers are devoted to the study of multiplicative stochastic perturbation of nonlinear first-order hyperbolic problems in the $\mathbb{R}^d$ case. Let us mention, without any claim of being exhaustive, the work of J. Feng and D. Nualart [7] where they

2000 *Mathematics Subject Classification.* Primary: 60H15, 35R60, 35L60.

Key words and phrases. Stochastic PDE; first-order hyperbolic equation; Cauchy problem; multiplicative stochastic perturbation; Young measures; Kruzhkov's entropy.

The author is supported by grant of the C.G. 64 (France).

introduced a notion of strong entropy solution in order to prove the uniqueness of the entropy solution for the Cauchy problem:

$$\mathrm{du} + \mathrm{div}(\vec{\mathbf{f}}(u))\mathrm{dt} = \int_{z\in Z} \sigma(.,u,z)\mathrm{dw}(t,z).$$

Using vanishing viscosity and compensated compactness arguments, the authors established existence of strong entropy solutions only in the $1D$ case.
In the recent paper [3], G.-Q. Chen, Q. Ding and K. H. Karlsen propose to revisit the work of J. Feng and D. Nualart. They prove that the multidimensional stochastic problem is well-posed by using uniform spatial BV-bound. They show the existence of strong stochastic entropy solutions in $L^p \cap BV$ for any finite p and develop a "continuous dependence" theory for stochastic entropy solutions in BV.
Finally, let us mention the paper by A. Debussche and J. Vovelle [5] concerning the d-dimensional problem with multiplicative noise

$$\mathrm{du} + \mathrm{div}\,\vec{\mathbf{f}}(u)\mathrm{dt} = h(u)\mathrm{dw},$$

which is considered on a torus. The authors use the kinetic formulation of the problem and prove existence and uniqueness of a kinetic solution in L^p for any finite p.
The aim of C. Bauzet, G. Vallet and P. Wittbold in [2] is to complete the results of [7] by showing existence and uniqueness of solution in $L^2(\mathbb{R}^d)$ without using the notion of strong entropy solution. The authors propose a method of artificial viscosity to prove the existence of a solution. The compactness properties used are based on the theory of Young measures and on measure-valued solutions. Then, an appropriate adaptation of Kruzhkov's doubling variables technique, and of the way J. Feng and D. Nualart propose to treat the stochastic source term, is presented to prove that any stochastic entropy solution is equal to a solution given by the artificial viscosity method. Thus, the entropy inequalities seem to suffice for the uniqueness *via* Kato-type inequality. This yields the uniqueness of the measure-valued entropy solution, and, by standard arguments, this allows us to deduce the existence and the uniqueness of the stochastic entropy solution.
We propose in this paper to present difficulties (brought by the stochastic perturbation) met by the authors in the use of classical tools from the deterministic setting, and techniques developed to treat the stochastic terms in [2].

First of all, let us introduce some notations and make precise the functional setting.

- Denote by E the integral over Ω with respect to the probability measure P.
- $\mathcal{D}^+([0,T]\times\mathbb{R}^d)$ denotes the subset of nonnegative elements of $\mathcal{D}([0,T]\times\mathbb{R}^d)$.
- For a given separable Banach space X we denote by $N^2_w(0,T,X)$ the space of the predictable X-valued processes (cf. [4]). This space is $L^2(]0,T[\times\Omega,X)$ endowed with the product measure $dt\otimes dP$ and the predictable σ-field $\mathcal{P}_T$(*i.e.* the σ-field generated by the sets $\{0\}\times\mathcal{F}_0$ and the rectangles $]s,t]\times A$ for any $A\in\mathcal{F}_s$).
- $\mathcal{E} = \{\eta \in C^{2,1}(\mathbb{R}),\ \eta \geq 0,\ \text{convex},\ \eta(0)=0,\ \mathrm{supp}\,\eta'' \text{ compact}\}$, the set of smooth entropies.
- $\eta_\delta \in \mathcal{E}$ denotes a uniform approximation of the absolute value function : $\eta'_\delta(r) = 1$ if $r\geq\delta$, $\eta'_\delta(r) = \sin\left(\frac{\pi}{2\delta}r\right)$ if $-\delta<r<\delta$ and $\eta'_\delta(r) = -1$ else.
- $\forall\eta\in\mathcal{E}$, $F^\eta(a,b) = \displaystyle\int_a^b \eta'(\sigma-a)\vec{\mathbf{f}}'(\sigma)\mathrm{d}\sigma$.
- $F(a,b) = \mathrm{Sgn}_0(a-b)[\vec{\mathbf{f}}(a)-\vec{\mathbf{f}}(b)] = \lim\limits_{\delta\to0^+} F^{\eta_\delta}(a,b)$ denotes the entropy flux.

- For any $u \in \mathcal{N}^2_w(0,T,L^2(\mathbb{R}^d))$, $k \in \mathbb{R}$, $\eta \in \mathcal{E}$ and $\varphi \in \mathcal{D}(\mathbb{R}^{d+1})$ denote by

$$\begin{aligned}\mu_{u,\eta,k}(\varphi) \quad &= \quad \int_{\mathbb{R}^d} \eta(u_0-k)\varphi(0)\mathrm{dx} + \int_Q \eta(u-k)\partial_t\varphi - F^\eta(u,k)\nabla\varphi \mathrm{dxdt} \\ &\quad + \int_Q \eta'(u-k)h(u)\varphi\mathrm{dxdw}(t) + \frac{1}{2}\int_Q h^2(u)\eta''(u-k)\varphi\mathrm{dxdt}.\end{aligned}$$

Definition 1.1. A function u of $\mathcal{N}^2_w(0,T,L^2(\mathbb{R}^d))$ is an entropy solution of the stochastic conservation law (1) with the initial condition $u_0 \in L^2(\mathbb{R}^d)$ if $u \in L^\infty(0,T,L^2(\Omega,L^2(\mathbb{R}^d)))$ and, for any $\varphi \in \mathcal{D}^+([0,T]\times\mathbb{R}^d)$, any real k and any $\eta \in \mathcal{E}$

$$0 \leq \mu_{u,\eta,k}(\varphi) \qquad P-\text{a.s.}$$

Remark 1. As $u \in L^\infty(0,T,L^2(\Omega,L^2(\mathbb{R}^d)))$, we can follow [8] p.84 to show that $\operatorname{ess}\lim_{t\to 0^+} E\int_K |u(t,x)-u_0|\mathrm{dx} = 0$ for any compact set $K \subset \mathbb{R}^d$, here the random variable doesn't bring new difficulty.

2. **The parabolic case.** The following existence and uniqueness result is a classic one. One can refer to [4] Section 7.3 and many others authors.

Proposition 1. *For any positive ϵ, there exists a unique $u_\epsilon \in \mathcal{N}^2_w(0,T;H^1(\mathbb{R}^d)) \cap L^\infty(0,T;L^2(\Omega\times\mathbb{R}^d))$, with $\partial_t[u_\epsilon - \int_0^\cdot h(u_\epsilon)\mathrm{dw}]$ and Δu_ϵ in $L^2(\Omega\times Q)$ and such that u_ϵ is a weak solution of the stochastic nonlinear parabolic problem*

$$\mathrm{d}u_\epsilon - [\epsilon\Delta u_\epsilon + \operatorname{div}(\vec{\mathbf{f}}(u_\epsilon))]\mathrm{dt} = h(u_\epsilon)\mathrm{dw} \quad \text{in } \Omega\times\mathbb{R}^d\times]0,T[, \tag{2}$$

for the initial condition $u_0^\epsilon \in \mathcal{D}(\mathbb{R}^d)$. Moreover, there exists a positive constant C such that,

$$\forall\epsilon>0, \quad ||u_\epsilon||^2_{L^\infty(0,T;L^2(\Omega\times\mathbb{R}^d))} + \epsilon||u_\epsilon||^2_{L^2(]0,T[\times\Omega;H^1_0(\mathbb{R}^d))} \leq C.$$

Remark 2. We consider here $(u_0^\epsilon)_\epsilon$ a sequence approximating our initial condition u_0 in $L^2(\mathbb{R}^d)$. The regularity of $\partial_t[u_\epsilon - \int_0^\cdot h(u_\epsilon)\mathrm{dw}]$ and Δu_ϵ in $L^2(\Omega\times Q)$ is not obvious. It comes from the suitable choice of $u_0^\epsilon \in \mathcal{D}(\mathbb{R}^d)$. One refers to the work of G. Vallet [10].

Consider φ in $\mathcal{D}^+([0,T]\times\mathbb{R}^d)$, k a real number and $\eta \in \mathcal{E}$. Since $\eta(u_\epsilon-k)\varphi \in L^2(0,T,H^1(\mathbb{R}^d))$ a.s., it is possible to apply the Itô formula to the operator $\Psi(t,u_\epsilon) := \int_{\mathbb{R}^d}\eta(u_\epsilon-k)\varphi\mathrm{dx}$ and thus, P-a.s.:

$$\begin{aligned}&\int_{\mathbb{R}^d}\eta(u_\epsilon(T)-k)\varphi(T)\mathrm{dx} \\ = \quad &\int_{\mathbb{R}^d}\eta(u_0^\epsilon-k)\varphi(0)\mathrm{dx} + \int_Q \eta(u_\epsilon-k)\partial_t\varphi\mathrm{dxdt} \\ &-\epsilon\int_Q \eta'(u_\epsilon-k)\nabla u_\epsilon\nabla\varphi\mathrm{dxdt} - \epsilon\int_Q \eta''(u_\epsilon-k)\varphi\nabla u_\epsilon\nabla u_\epsilon\mathrm{dxdt} \\ &-\int_Q \eta'(u_\epsilon-k)\vec{\mathbf{f}}(u_\epsilon)\nabla\varphi\mathrm{dxdt} - \int_Q \eta''(u_\epsilon-k)\varphi\vec{\mathbf{f}}(u_\epsilon)\nabla u_\epsilon\mathrm{dxdt} \\ &+\int_0^T\int_{\mathbb{R}^d}\eta'(u_\epsilon-k)h(u_\epsilon)\varphi\mathrm{dxdw}(t) + \frac{1}{2}\int_Q h^2(u_\epsilon)\eta''(u_\epsilon-k)\varphi\mathrm{dxdt}.\end{aligned}$$

Remark 3. Let us mention that in the deterministic setting, the viscous entropy formulation is obtained by testing the parabolic regularization with $\eta(\tilde{u}_\epsilon - k)\varphi$, if $\tilde{u}_\epsilon$ denotes the solution of this regularization. In the stochastic case, testing the stochastic parabolic regularization with $\eta(u_\epsilon - k)\varphi$ corresponds to the application of Itô's derivation formula to $\Psi(t, u_\epsilon) := \int_{\mathbb{R}^d} \eta(u_\epsilon - k)\varphi dx$. Notice that the stochastic perturbation brings two new terms in this derivation formula: one containing an Itô integral, and another one containing the second-order derivative of η.

Since the support of η'' is compact, for any $i = 1, \dots, d$, $\mathbb{R} \ni r \mapsto \eta''(r - k)f_i(r)$ is a bounded continuous function. Then, thanks to the chain-rule for Sobolev functions, we obtain the following viscous entropy formulation for any P-measurable set A

$$\begin{aligned} 0 \quad \leq \quad & E\left[1_A \int_0^T \int_{\mathbb{R}^d} \eta'(u_\epsilon - k)h(u_\epsilon)\varphi \mathrm{dxdw}(t)\right] \\ & -\epsilon E\left[1_A \int_Q \eta'(u_\epsilon - k)\nabla u_\epsilon \nabla \varphi \mathrm{dxdt}\right] + E\left[1_A \int_{\mathbb{R}^d} \eta(u_0^\epsilon - k)\varphi(0)\mathrm{dx}\right] \\ & +E\left[1_A \int_Q \eta(u_\epsilon - k)\partial_t \varphi - F^\eta(u_\epsilon, k)\nabla\varphi + \frac{1}{2}h^2(u_\epsilon)\eta''(u_\epsilon - k)\varphi \mathrm{dxdt}\right] \quad (3) \\ := \quad & E[1_A \mu^\epsilon_{u_\epsilon,\eta,k}(\varphi)]. \end{aligned}$$

3. **Entropy formulation.** We would like to pass to the limit in (3) with respect to ϵ. Because of the random variable, we are not able to use classical results of compactness. But the one given by the concept of Young measures is appropriate here, and the technique is based on the narrow convergence of Young measures (or entropy processes), we refer to E.J. Balder [1] but also to R. Eymard *et al.* [6].
Since u_ϵ is a bounded sequence in $\mathcal{N}^2_w(0, T, L^2(\mathbb{R}^d))$ and thanks to the compact support of φ in $\mathbb{R}^d$, $\mathbf{u}_\epsilon$ converges (up to a subsequence still denoted u_ϵ) in the sense of Young measures to an "entropy process" denoted by $\mathbf{u}$ in $L^\infty(0, T, L^2(\Omega \times \mathbb{R}^d \times]0, 1[))$ with an additional variable $\alpha \in (0, 1)$ (*cf.* [6] §-2.2 or [2] §-5.3.). Precisely, given a Carathéodory function $\Psi : Q \times \Omega \times \mathbb{R} \to \mathbb{R}$ such that $\Psi(., \mathbf{u}_\epsilon)$ is uniformly integrable, one has:

$$E\int_Q \psi(., \mathbf{u}_\epsilon)dxdt \underset{\epsilon \to 0}{\longrightarrow} E\int_Q \int_0^1 \psi(., \mathbf{u}(., \alpha))d\alpha dxdt.$$

By assumptions on η, all the integrands in the third line of (3) are uniformly integrable and passing to the limit is possible in all the integrals. One is also able to pass to the limit in the first term of (3) using the weak continuity of the stochastic integral from $L^2(\Omega \times Q)$ to $L^2(\Omega \times \mathbb{R}^d)$, see [4]. Finally, the *a priori* estimate on ∇u_ϵ yields that the second term of (3) tends to 0 with ϵ.

Therefore at the limit one gets

$$\begin{aligned}0 \quad \leq \quad & E\left[1_A \int_0^T \int_{\mathbb{R}^d} \int_0^1 \eta'(\mathbf{u}(.,\alpha)-k)h(\mathbf{u}(.,\alpha))\varphi \mathrm{d}\alpha \mathrm{dx} \mathrm{dw}(t)\right] \\ & +E\left[1_A \int_{\mathbb{R}^d} \eta(u_0-k)\varphi(0)\mathrm{dx}\right] \\ & +E\left[1_A \int_Q \int_0^1 \left[\eta(\mathbf{u}(.,\alpha)-k)\partial_t\varphi - F^\eta(\mathbf{u}(.,\alpha),k)\nabla\varphi\right] \mathrm{d}\alpha \mathrm{dxdt}\right] \\ & +\frac{1}{2}E\left[1_A \int_Q \int_0^1 h^2(\mathbf{u}(.,\alpha))\eta''(\mathbf{u}(.,\alpha)-k)\varphi \mathrm{d}\alpha \mathrm{dxdt}\right].\end{aligned}$$

Remark 4. Since (u_ϵ) is bounded in the Hilbert space $\mathcal{N}^2_w(0,T,L^2(\mathbb{R}^d))$, by identification, one shows that $u_\epsilon \rightharpoonup \int_0^1 \mathbf{u}(.,\alpha)\mathrm{d}\alpha$ in the same space, and so $\int_0^1 \mathbf{u}(.,\alpha)\mathrm{d}\alpha$ is a predictable process. The interesting point is the measurability of $\mathbf{u}$ with respect to all its variables (t,x,ω,α). Revisiting the work of E. Yu. Panov [9] with the σ-field $\mathcal{P}_T \otimes L(\mathbb{R}^d)$, one shows that $\mathbf{u}$ is measurable for the σ-field $\mathcal{P}_T \otimes L(\mathbb{R}^d \times]0,1[)$.

A separability argument for the norm of $H^1(Q)$ yields the existence of a Young measure solution in the sense of the following definition.

Definition 3.1. $\mathbf{u} \in \mathcal{N}^2_w\left(0,T,L^2(\mathbb{R}^d\times]0,1[)\right) \cap L^\infty\left(0,T,L^2(\Omega\times\mathbb{R}^d\times]0,1[)\right)$ is a measure-valued entropy solution of (1) with the initial data $u_0 \in L^2(\mathbb{R}^d)$ if for any $\eta \in \mathcal{E}$ and any $(k,\varphi) \in \mathbb{R}\times\mathcal{D}^+([0,T]\times\mathbb{R}^d)$,

$$0 \quad \leq \quad \int_0^1 \mu_{\mathbf{u},\eta,k}(\varphi)d\alpha, \quad P-\text{a.s.}$$

4. **Local Kato inequality.** The aim of this section is to discuss about the way to obtain the following interior Kato inequality, which permits to prove that the measure-valued solution is an entropy solution in the sense of Definition 1.1.

Proposition 2. *Let $\mathbf{u}_1$, $\mathbf{u}_2$ be measure-valued entropy solutions to (1) with initial data $u_{1,0}, u_{2,0} \in L^2(\mathbb{R}^d)$, respectively. Then, for any function φ in $\mathcal{D}^+([0,T]\times\mathbb{R}^d)$, one has*

$$\begin{aligned}0 \quad \leq \quad & \int_{\mathbb{R}^d} |u_{1,0}-u_{2,0}|\varphi(0)\mathrm{dx} + E\int_{Q\times]0,1[^2} \Big|\mathbf{u}_1(t,x,\alpha)-\mathbf{u}_2(t,x,\beta)\Big|\partial_t\varphi \mathrm{dxdtd}\alpha\mathrm{d}\beta \\ & -E\int_{Q\times]0,1[^2} F\Big(\mathbf{u}_1(t,x,\alpha),\mathbf{u}_2(t,x,\beta)\Big).\nabla\varphi \mathrm{dxdtd}\alpha\mathrm{d}\beta. \qquad (4)\end{aligned}$$

Proof. We propose here to present stages of the proof introduced in [2], emphasizing on differences with the deterministic setting, and stochastic calculus tools chosen. The main idea is to use Kruzhkov's doubling variables method. Let us apply the usual technique and advice when we meet difficulties. For this, we consider two measure-valued solutions $\mathbf{u_1}$, $\mathbf{u_2}$ and these inequalities P-a.s.:

$$0 \leqslant \int_0^1 \mu_{\mathbf{u_1}(t,x,\alpha),\eta_\delta,k_2}(\psi)d\alpha \qquad ; \qquad 0 \leqslant \int_0^1 \mu_{\mathbf{u_2}(s,y,\beta),\eta_\delta,k_1}(\psi)\mathrm{d}\beta, \qquad (5)$$

where $k_1,k_2 \in \mathbb{R}$ and $\psi \in \mathcal{D}^+([0,T]\times\mathbb{R}^d)$.
Notice that, by comparing with the deterministic case, the stochastic perturbation of our conservation law brings new terms in the entropy inequalities, the ones

containing an Itô integral:

$$\int_0^1 \int_Q \eta'_\delta(\mathbf{u_1}(t,x,\alpha)-k_2)h(\mathbf{u_1})\psi \mathrm{d}x\mathrm{d}w(t)\mathrm{d}\alpha$$

$$\int_0^1 \int_Q \eta'_\delta(\mathbf{u_2}(s,y,\beta)-k_1)h(\mathbf{u_2})\psi \mathrm{d}y\mathrm{d}w(s)\mathrm{d}\beta, \qquad (6)$$

and others containing the second derivative of η_δ:

$$\frac{1}{2}\int_0^1 \int_Q h^2(\mathbf{u_1}(t,x,\alpha))\eta''_\delta(\mathbf{u_1}-k_2)\psi \mathrm{d}t\mathrm{d}x\mathrm{d}\alpha$$

$$\frac{1}{2}\int_0^1 \int_Q h^2(\mathbf{u_2}(s,y,\beta))\eta''_\delta(\mathbf{u_2}-k_1)\psi \mathrm{d}s\mathrm{d}y\mathrm{d}\beta. \qquad (7)$$

Usually, we take in (5) $k_1 = \mathbf{u_1}(t,x,\alpha)$, $k_2 = \mathbf{u_2}(s,y,\beta)$, $\psi(t,x,s,y) = \varphi(s,y)\rho_m(x-y)\rho_n(t-s)$ with $\varphi \in \mathcal{D}^+([0,T]\times\mathbb{R}^d)$, $\operatorname{supp}\varphi(t,.) \subset K$ a compact set of $\mathbb{R}^d$, ρ_n and ρ_m the usual mollifier sequences in $\mathbb{R}$ and $\mathbb{R}^d$ respectively. Then, we integrate the first inequality with respect to (s,y,β) and the second one with respect to (t,x,α), we add these two new inequalities and pass to the limit with respect to δ, n and m.

In our case, there is a problem with this technique when we treat the stochastic integrals (6). Indeed, because of the definition of the Itô integral, we require $\mathbf{u_2}(s,y,\beta)$ to be $\mathcal{F}_t$-measurable and $\mathbf{u_1}(t,x,\alpha)$to be $\mathcal{F}_s$-measurable. This is not satisfied because we ignore if $s>t$ or $s<t$ (recall that $\mathcal{F}_s \subset \mathcal{F}_t$ for $0 \leqslant s \leqslant t$).

For this reason, we keep the variable k in (5), we consider a new mollifier sequence ρ_l and we multiply by $\rho_l(\mathbf{u_1}(t,x,\alpha)-k)$ the inequality coming from $\mu_{\mathbf{u_2},\eta_\delta,k}$ and integrate with respect to (t,x,α). We also multiply by $\rho_l(\mathbf{u_2}(s,y,\beta)-k)$ the inequality coming from $\mu_{\mathbf{u_1},\eta_\delta,k}$ and integrate with respect to (s,y,β). Then we add these two inequalities, integrate over k in $\mathbb{R}$ all the formulation and take the expectation, to get:

$$0 \leqslant E\int_Q \int_{]0,1[^2} \int_{\mathbb{R}} \mu_{\mathbf{u_1},\eta_\delta,k}(\varphi(s,y)\rho_m(x-y)\rho_n(t-s))\rho_l(\mathbf{u_2}(s,y,\beta)-k)\mathrm{d}k\mathrm{d}\alpha\mathrm{d}\beta\mathrm{d}s\mathrm{d}y$$

$$+E\int_Q \int_{]0,1[^2} \int_{\mathbb{R}} \mu_{\mathbf{u_2},\eta_\delta,k}(\varphi(s,y)\rho_m(x-y)\rho_n(t-s))\rho_l(\mathbf{u_1}(t,x,\alpha)-k)\mathrm{d}k\mathrm{d}\beta\mathrm{d}\alpha\mathrm{d}t\mathrm{d}x.$$

With a judicious order for the passage to the limit, we are able to overcome our initial difficulty. Indeed, we first pass to the limit as $n \to \infty$, then we get the same time everywhere (t or s), and the problem of measurability with respect to the σ-field $\mathcal{P}_T$ is forgotten. Then, passing to the limit on l, amounts to replacing $\mathbf{u_1}(t,x,\alpha)$ and $\mathbf{u_2}(t,y,\beta)$ in our formulation, as we wished at the beginning. And we pass to the limit on δ and m.

The second delicate point appears with terms containing the second derivative of η_δ (7) when we want to pass to the limit on δ. Indeed, because of η''_δ, we are not able to identify the limit of those terms, all we can say is that the limit exists (Tanaka formula). The problem is that we need to know the limit to obtain the local Kato inequality. For this reason, we decide to apply the method, not with $\mathbf{u_2}$, but with a viscous regular solution u_ϵ. Thus, the suitable regularity of such a solution allows us to apply the Itô formula. Following the idea of J. Feng and D. Nualart concerning the stochastic terms, the idea remains on combining terms containing η''_δ with others coming from stochastic calculus. Then, passing to the limit as n and l go to infinity on terms containing η''_δ gives:

$$\begin{aligned} & \frac{1}{2}E\int_{\mathbb{R}^d}\int_Q\int_0^1 h^2(\hat{\mathbf{u}})\eta''_\delta(\hat{\mathbf{u}}-u_\epsilon(t,y))\rho_m(x-y)\varphi \mathrm{d}\alpha\mathrm{dtdxdy} \\ & +\frac{1}{2}E\int_{\mathbb{R}^d}\int_Q\int_0^1 h^2(u_\epsilon)\eta''_\delta(u_\epsilon-\hat{\mathbf{u}}(t,x,\alpha))\rho_m(x-y)\varphi \mathrm{d}\alpha\mathrm{dtdxdy} \\ := \quad & A+B. \end{aligned}$$

Moreover, thanks to the martingale property of the Itô integral, the stochastic terms can be written like this

$$\begin{aligned} & E\int_Q\int_{\mathbb{R}}\int_{s-2/n}^s\int_{\mathbb{R}^d}\int_0^1 \eta'_\delta(\hat{\mathbf{u}}-k)h(\hat{\mathbf{u}})\mathrm{d}\alpha\varphi\rho_m(x-y)\rho_n(t-s)\mathrm{dxdw}(t) \\ & \qquad \times\rho_l(u_\epsilon(s,y)-k)\mathrm{dkdyds} \\ = \quad & E\int_Q\int_{\mathbb{R}}\int_{\mathbb{R}^d}\int_{s-2/n}^s\int_0^1 \eta'_\delta(\hat{\mathbf{u}}-k)h(\hat{\mathbf{u}})\mathrm{d}\alpha\rho_n(t-s)\mathrm{dw}(t)\varphi\rho_m(x-y)\mathrm{dx} \\ & \qquad \times[\rho_l(u_\epsilon(s,y)-k)-\rho_l(u_\epsilon(s-2/n,y)-k)]\mathrm{dkdyds} \\ := \quad & C_{n,l}. \end{aligned}$$

Here the choice of u_ϵ instead of $\mathbf{u_2}$ is crucial. Indeed, the regularity of u_ϵ allows us to apply Itô's formula with $\mathrm{d}u_\epsilon = [\epsilon\Delta u_\epsilon + div\vec{\mathbf{f}}(u_\epsilon)]\mathrm{dt} + h(u_\epsilon)\mathrm{dw} = A_\epsilon\mathrm{dt} + h(u_\epsilon)\mathrm{dw}$ and to get:

$$\begin{aligned} & \rho_l(u_\epsilon(s,y)-k)-\rho_l(u_\epsilon(s-2/n,y)-k) \\ = \quad & \int_{s-\frac{2}{n}}^s \rho'_l(u_\epsilon(\sigma,y)-k)A_\epsilon(\sigma,y)\mathrm{d}\sigma + \int_{s-\frac{2}{n}}^s \rho'_l(u_\epsilon(\sigma,y)-k)h(u_\epsilon(\sigma,y))\mathrm{dw}(\sigma) \\ & +\frac{1}{2}\int_{s-\frac{2}{n}}^s \rho''_l(u_\epsilon(\sigma,y)-k)h^2(u_\epsilon(\sigma,y))\mathrm{d}\sigma, \end{aligned}$$

which wasn't possible with a measure-valued solution. Thus, an integration by parts with respect to the variable k yields:

$$\begin{aligned} C_{n,l} \quad = \quad & -E\int_Q\int_{\mathbb{R}}\int_{\mathbb{R}^d}\int_{s-2/n}^s\int_0^1 \eta''_\delta(\hat{\mathbf{u}}-k)h(\hat{\mathbf{u}})\mathrm{d}\alpha\rho_n(t-s)\mathrm{dw}(t)\varphi\rho_m(x-y)\mathrm{dx} \\ & \times\Big[\int_{s-\frac{2}{n}}^s \rho_l(u_\epsilon(\sigma,y)-k)A_\epsilon(\sigma,y)\mathrm{d}\sigma + \int_{s-\frac{2}{n}}^s \rho_l(u_\epsilon(\sigma,y)-k)h(u_\epsilon(\sigma,y))\mathrm{dw}(\sigma) \\ & +\frac{1}{2}\int_{s-\frac{2}{n}}^s \rho'_l(u_\epsilon(\sigma,y)-k)h^2(u_\epsilon(\sigma,y))\mathrm{d}\sigma\Big]\mathrm{dkdyds} \\ \to_{n,l} \quad & -\int_Q\int_{\mathbb{R}^d}\int_0^1 E\Big[\eta''_\delta(\hat{\mathbf{u}}(s,x,\alpha)-u_\epsilon(s,y)))h(\hat{\mathbf{u}})h(u_\epsilon)\Big]\varphi\rho_m(x-y)\mathrm{d}\alpha\mathrm{dydxds} \\ := \quad & C. \end{aligned}$$

Thus,

$$\begin{aligned} A+B+C \quad = \quad & \frac{1}{2}E\int_Q\int_{\mathbb{R}^d}\int_0^1 [h(\hat{\mathbf{u}})-h(u_\epsilon)]^2\eta''_\delta(u_\epsilon-\hat{\mathbf{u}})\rho_m(x-y)\varphi\mathrm{d}\alpha\mathrm{dydxds} \\ \to_\delta \quad & 0. \end{aligned}$$

In summary, this is the plan of the proof. By doing stochastic computations on the Itô integral and passing to the limit (with classical techniques) with respect to n,

l, δ, ϵ, m in this order in

$$0 \leqslant E\int_Q \int_0^1 \int_{\mathbb{R}} \mu_{\hat{\mathbf{u}},\eta_\delta,k}(\varphi(s,y)\rho_m(x-y)\rho_n(t-s))\rho_l(u_\epsilon(s,y)-k)\mathrm{dkd}\alpha\mathrm{dsdy}$$

$$+E\int_Q \int_0^1 \int_{\mathbb{R}} \mu^\epsilon_{u_\epsilon,\eta_\delta,k}(\varphi(s,y)\rho_m(x-y)\rho_n(t-s))\rho_l(\hat{\mathbf{u}}(t,x,\alpha)-k)\mathrm{dkd}\alpha\mathrm{dtdx},$$

we finally obtain the local Kato inequality. □

Proposition 3. *The measure-valued entropy solution is unique. Moreover, it is the unique stochastic entropy solution.*

Proof. As in the deterministic case, set $\kappa = \|f'\|_\infty$, $\hat{u}_0 = u_0$, $\gamma(t) = \frac{(T-t)^+}{T}$, and denote by ψ any nonincreasing regular function with $1_{]-\infty,K]} \leq \psi \leq 1_{]-\infty,K+1]}$, where $K > 0$. Then, considering $K = R + \kappa T$ for any $R > 0$ and $\varphi(t,x) = \psi(|x|+\kappa t)\gamma(t)$ in (4) implies that, $\mathbf{u}(t,x,\beta) = \hat{\mathbf{u}}(t,x,\alpha)$ for almost any $x \in B(0,R)$, $t \in]0,T[$, $\omega \in \Omega$, $\alpha, \beta \in]0,1[$. Thus, on the one hand $\mathbf{u} = \hat{\mathbf{u}}$; on the other hand $\mathbf{u}(t,x,\alpha) = u(t,x)$ is independent of α. Hence, u is the unique stochastic entropy solution in the sense of Definition 1.1.

□

Proposition 4. *Stochastic entropy solutions satisfy a "contraction principle": if u_1, u_2 are stochastic entropy solutions of (1) corresponding to initial data $u_{1,0}, u_{2,0} \in L^2(\mathbb{R}^d)$, respectively, then, for any positive K and time t,*

$$E\int_{B(0,K-\kappa t)} |u_1 - u_2|\mathrm{dx} \leq \int_{B(0,K)} |u_{1,0} - u_{2,0}|\mathrm{dx},$$

where $\kappa = \|f'\|_\infty$.

Proof. This is a consequence of the previous proof when ψ converges to $1_{]-\infty,K]}$. □

REFERENCES

[1] E. J. Balder, *Lectures on Young measure theory and its applications in economics*, Rend. Istit. Mat. Univ. Trieste, **31** (2000), 1–69.

[2] C. Bauzet, G. Vallet and P. Wittbold, *The Cauchy problem for a conservation law with a multiplicative stochastic perturbation*, Journal of Hyperbolic Differential Equations, **9** (2012), 661–709.

[3] G.-Q. Chen, Q. Ding, and K.H. Karlsen, *On nonlinear stochastic balance laws* Arch. Ration. Mech. Anal., **204** (2012), 707–743.

[4] G. Da. Prato and J. Zabczyk, "Stochastic Equations in Infinite Dimensions", Cambridge University Press, Cambridge, 1992.

[5] A. Debussche and J. Vovelle, *Scalar conservation laws with stochastic forcing* Journal of Functional Analysis, **259** (2010), 1014–1042.

[6] R. Eymard, T. Gallouët and R. Herbin, *Existence and uniqueness of the entropy solution to a nonlinear hyperbolic equation* Chin. Ann. Math., **16** (1995), 1–14.

[7] J. Feng and D. Nualart, *Stochastic scalar conservation laws* Journal of Functional Analysis, **255** (2008), 313–373.

[8] J. Malek, J. Necas, M. Rokyta and Ruzicka, "Weak and Measure-valued Solutions to Evolutionary PDE's", Applied mathematics and mathematical computation, 1996.

[9] E. Yu. Panov, *On measure-valued solutions of the Cauchy problem for a first-order quasilinear equation* Investig. Mathematics, **60**(1996), 335–377.

[10] G. Vallet, *Stochastic perturbation of nonlinear degenerate parabolic problems* Differential and integral equation, **21**(2008), 1055–1082.

E-mail address: caroline.bauzet@univ-pau.fr

A MULTI-PHASE MODEL FOR PEDESTRIAN FLOW WITH STRATEGIC AND TACTIC COMPONENT

Stefan Berres

Departamento de Ciencias Matemáticas y Físicas, Facultad de Ingeniería
Universidad Católica de Temuco, Temuco, Chile

Frank Huth, Hartmut Schwandt and Günter Bärwolff

Technische Universität Berlin, Fakultät II, Institut für Mathematik
Straße des 17. Juni 136, D-10623 Berlin, Germany

Abstract. In this contribution, a macroscopic multi-pedestrian flow is modelled by a system of convection-diffusion equations. The convection corresponds to a movement towards a strategic direction, whereas the diffusion corresponds to a tactical movement that avoids jams. Different populations moving in different directions are represented by different phases. Numerical experiments demonstrate the qualitative behaviour of the simulation model.

1. **Introduction.** In the last two decades the modelling and simulation of pedestrian traffic has become an important tool both for scientific purposes and management decisions in the fields of planning and operation of public spaces like airports, shopping malls, sport stadiums, or public events and manifestations in the context of trouble-free operation and security. While single destination problems, in particular evacuation scenarios, have been studied quite intensively, multi destination problems, where distinct streams of pedestrians move from one or more starting points to multiple destinations, still yield a broad field for further research. In particular, intersecting pedestrian streams have not yet been studied in detail. In the modelling of pedestrian behaviour, methods and models from two basic model classes are usually applied [5]. In *microscopic* models, pedestrians are considered as individual objects interacting with each other, while in it macroscopic models, pedestrian behaviour is analyzed in terms of more global properties of a continuous stream. Macroscopic models focus on the balancing relationships of particle density, interpreting pedestrians as particles, flow intensity and flow speed etc. We refer to [6] for a detailed overview of both vehicular and pedestrian traffic and the main modelling and simulation approaches. The modelling of pedestrian flows with macroscopic approaches has received various contributions in the last decade, see [2, 4, 7, 8, 9, 10]. One reason to model pedestrian flows by fluid physical models from gas or fluid dynamics results from the observation that in very dense pedestrian crowds, pedestrians behave quite similarly to gas particles. Most of the

2000 *Mathematics Subject Classification.* Primary: 90B20, 35Q70; Secondary: 35L65, 91D10.

Key words and phrases. Crowd dynamics, multi-phase flow, traffic flow, convection-diffusion equation, pedestrian model.

The first author has been supported by Conicyt (Chile) through Fondecyt project #1120587. The third and forth author have been supported by the Deutsche Forschungsgemeinschaft through projects SCHW548/5-1 and BA1189/4-1.

research in the macroscopic area of pedestrian simulation is, therefore, focussed on the discussion of partial differential equations, based on physical principles like mass, momentum and energy balances. This contribution is a further development of a model, where a convection-diffusion equation with only linear diffusion has been considered [2, 3]; this further development models crowd avoiding behaviour and will be part of the full paper version.

2. **Modelling.** We assume pedestrian flow to be a transport problem principally defined by a mass balance. Let Ω be an open sufficiently smooth bounded domain in $\mathbb{R}^2$ and $(0,T)$ an open interval. For $(x,y)\in\Omega$, $t\in(0,T)$ the mass equation

$$\frac{\partial \varrho_i}{\partial t}+\nabla\cdot(\varrho_i\,\boldsymbol{v}_i)=0,\qquad i=1,\dots,n, \tag{1}$$

describes the mass flow, where t denotes time, $\varrho_i\in[0,1]$, $\boldsymbol{v}_i=\boldsymbol{v}_i(\varrho_1,\dots,\varrho_n), 1\leq i\leq n$, the local densities and speeds, respectively, of the i-th component of $n\in\mathbb{N}$ distinct pedestrian species. The model developed in the sequel has the general form

$$\frac{\partial \varrho_i}{\partial t}+\nabla\cdot\boldsymbol{f}_i(\varrho_1,\dots,\varrho_n;x,y)=\sum_{j=1}^{n}\nabla\cdot(b_{ij}(\varrho_1,\dots,\varrho_n)\nabla\varrho_j),\quad i=1,\dots,n, \tag{2}$$

or written in vector form

$$\frac{\partial \boldsymbol{\varrho}}{\partial t}+\nabla\cdot\boldsymbol{f}(\boldsymbol{\varrho})=\nabla\cdot\Big(\boldsymbol{B}(\boldsymbol{\varrho})\,\nabla\boldsymbol{\varrho}\Big), \tag{3}$$

with the diffusion matrix $\boldsymbol{B}(\boldsymbol{\varrho})=(b_{ij})_{i,j=1}^{n}$, $b_{ij}\equiv b_{ij}(\varrho_1,\dots,\varrho_n)$, $1\leq i,j\leq n$, a flow function $\boldsymbol{f}(\boldsymbol{\varrho})=\boldsymbol{f}_i(\varrho_1,\dots,\varrho_n;x,y))_{i=1,\dots,n}$, and where $\boldsymbol{\varrho}=(\varrho_1,\dots,\varrho_n)^T$, $\nabla\boldsymbol{\varrho}=(\nabla\varrho_1,\dots,\nabla\varrho_n)^T$. For $n=2$, the equation (2) has been considered in [2] with the constant diffusion matrix

$$\boldsymbol{B}(\varrho_1,\varrho_2)=\begin{pmatrix}\varepsilon & \delta\\ \delta & \varepsilon\end{pmatrix}. \tag{4}$$

In the sequel we develop a model with a nonlinear diffusion matrix. We assume that pedestrians, instead of following an uncontrolled uniform diffusion, will try to evade from crowded spaces. Our model will be based on the assumption that pedestrians avoid densely populated areas by modifying their motion into the direction of the negative gradient of the total local density $\varrho=\varrho_1+\dots+\varrho_n$. This local orientation can be metaphorically interpreted by the hypothetical behaviour of blind persons with white canes, who generally stick to their planned direction, but modify it by taking a direction away from congestion when they “detect” the gradient.

Our model has to reflect two aspects of pedestrian behaviour that correspond to strategic and tactic decision making, respectively. On the one hand, a pedestrian has a target, which he tries to reach, and on the other hand, he might be forced to deal with local problems like high densities. To take account of these two aspects, we restart from the mass equation (1) and decompose the velocity as

$$\boldsymbol{v}_i(\varrho_1,\dots,\varrho_n)=\boldsymbol{v}_i^{\mathrm{s}}(\varrho_1,\dots,\varrho_n)+\boldsymbol{v}_i^{\mathrm{t}}(\varrho_1,\dots,\varrho_n),\quad i=1,\dots,n, \tag{5}$$

consisting of the following two components:

- a strategic component $\boldsymbol{v}_i^{\mathrm{s}}$ following a strategic goal of reaching a certain destination on a desired path,
- and a tactical component $\boldsymbol{v}_i^{\mathrm{t}}$, which locally avoids densely populated areas.

Both velocity components are modelled as a product of velocity and direction,

$$\boldsymbol{v}_i^{\mathrm{s}} = a_i V \boldsymbol{d}_i^{\mathrm{s}}, \qquad \boldsymbol{v}_i^{\mathrm{t}} = b_i W \boldsymbol{d}_i^{\mathrm{t}}, \tag{6}$$

where $\boldsymbol{d}_i^{\mathrm{s}}$ and $\boldsymbol{d}_i^{\mathrm{t}}$ denote the direction field giving a desired strategic and an adapted tactical walking direction, respectively. The functions $V = V(\varrho)$ and $W = W(\varrho)$ with $\varrho = \sum_{i=1}^{n} \varrho_i$ are velocity modules denoting the magnitudes of the speeds in the respective direction. The constants a_i and b_i are maximal direction velocities.

Standard strategic directions are for example opposite or perpendicular. For a two-species model, where $n = 2$, the most simple examples are given by $\boldsymbol{d}_1^{\mathrm{s}} = (1,0)^{\mathrm{T}}$, $\boldsymbol{d}_2^{\mathrm{s}} = (-1,0)^{\mathrm{T}}$ and $\boldsymbol{d}_1^{\mathrm{s}} = (1,0)^{\mathrm{T}}$, $\boldsymbol{d}_2^{\mathrm{s}} = (0,1)^{\mathrm{T}}$, respectively.

With these specifications equation (1) can be written as

$$\frac{\partial \varrho_i}{\partial t} + \nabla \cdot \left\{ \varrho_i \left[a_i V \boldsymbol{d}_i^s + b_i W \boldsymbol{d}_i^t \right] \right\} = 0 \tag{7}$$

The tactical direction is modelled as

$$\boldsymbol{d}_i^{\mathrm{t}} = \begin{cases} -\nabla \varrho / |\nabla \varrho| & \text{for} \quad |\nabla \varrho| > 1, \\ -\nabla \varrho & \text{for} \quad |\nabla \varrho| \leq 1, \end{cases} \tag{8}$$

using a partial normalization which avoids unrealistic "escape" velocities. Defining

$$\chi(\varrho) = \begin{cases} 1/|\nabla \varrho| & \text{if} \quad |\nabla \varrho| > 1, \\ 1 & \text{if} \quad |\nabla \varrho| \leq 1, \end{cases} \tag{9}$$

we can interpret (7) as an example of model type (2) with

$$b_{ij}(\varrho_1, \dots, \varrho_n) = -b_i \, \varrho_i \, W \, \chi(\varrho), \quad i,j = 1, \dots, n, \tag{10}$$

which describes a diffusive flux opposite to the gradient of the total density and proportional to ϱ_i. Here, V is weighting the strategic part of the pedestrian way, and W the tactical part. A generic assumption for V is to be decreasing,

$$V(0) = 1, \quad V' \leq 0, \quad V(1) = 0, \tag{11}$$

which describes a throttling effect at higher concentrations. The more persons are in a given region, the more they get stuck on their way towards the strategic direction. The tactical velocity W is assumed to increase,

$$W(0) = 0, \quad W' \geq 0, \quad W(1) = 1, \tag{12}$$

which reflects the model assumption that at higher concentrations the tendency to evade increases. The more persons are blocking the way, the stronger is the tendency to move on along an alternative trajectory. Being aware of the symmetry of (11) and (12), one can impose the constraint

$$V + W = 1 \tag{13}$$

in order to model the realistic assumption that the partitioning of the flux into a strategically and a tactically caused directed part also results in a partitioning of a total velocity, which is normalized to 1. Equality (13) expresses that a pedestrian has an individual level of moving activity, which he partitions on the alternatives of either rather moving to the desired target or rather evading from jams.

Condition (13) does not constrain the qualitative behaviour but reduces the degree of freedom, which is helpful in absence of other modelling guidelines. In this contribution, we choose $1-\varrho$ and opt for the assumption (13). This yields $W=\varrho$, hence (10) becomes

$$b_{ij}(\varrho_1,\dots,\varrho_n) = -b_i\,\varrho_i\,\varrho\,\chi(\varrho), \quad i,j=1,\dots,n, \tag{14}$$

that yields a diffusion matrix of the form

$$\hat{\boldsymbol{B}}(\boldsymbol{\varrho}) = \varrho \begin{pmatrix} b_1\varrho_1 & \dots & b_1\varrho_1 \\ \vdots & & \vdots \\ b_n\varrho_n & \dots & b_n\varrho_n \end{pmatrix} \chi(\varrho). \tag{15}$$

3. **Numerical implementation.** We apply a finite volume discretization, as this approach is well suited for simulating pedestrian dynamics due to its mass preserving properties. For the implementation we use the OpenFOAM package. Our implementation allows for triangulations with possibly irregular grids. In view of the reciprocal complementation of discrete microscopic and PDE based macroscopic models, the further development of our approach will integrate data from a discrete grid-based microscopic method [1]. In particular, we will use the same grids for both approaches. This leads to the requirement of cells of approximately the same size allover the used grids. We use an upwind approximation, obtaining thus a spatial approximation of first order. This upwind strategy provides a necessary numerical stabilization of a convective dominated transport equation and produces an artificial viscosity of at least the order of the meshsize. As a practical consequence we also get a numerical tool for the resolution of jams reflecting real-life behaviour.

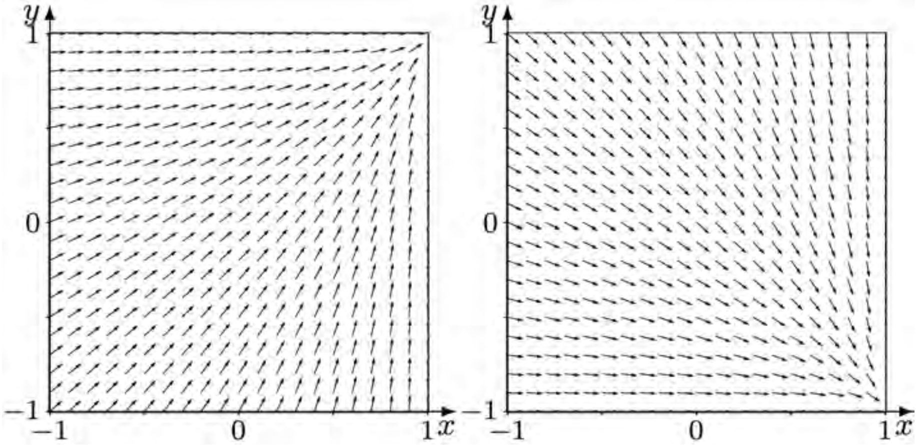


FIGURE 1. Direction fields of stream 1 (left) and 2 (right).

4. **Examples.** The following examples illustrate the different behaviour of model variants (4) with linear diffusion matrix $\mathbf{B}$, which we refer to as the "divergence" or Δ-simulation model and (15) with the nonlinear diffusion matrix $\hat{\mathbf{B}}$, which we refer to as the "gradient" or ∇-simulation model. We choose the domain $\Omega=[-1,1]^2$, which we assume to be empty at the beginning, i.e. we set

$$\varrho_i(0,x,y)=0 \quad \text{for all} \quad (x,y)\in\Omega, \quad i=1,2. \tag{16}$$

The subboundaries

$$\Gamma_{sw} = \Big\{(x,y) : x = -1, y \in [-1,-0,7]\Big\}, \quad \Gamma_{nw} = \Big\{(x,y) : x = -1, y \in [0.7,1]\Big\},$$
$$\Gamma_{ne} = \Big\{(x,y) : x = 1, y \in [0.7,1]\Big\}, \quad \Gamma_{se} = \Big\{(x,y) : x = 1, y \in [-1,-0.7]\Big\},$$

represent dedicated entries ($\Gamma_{sw,nw}$) and exits ($\Gamma_{ne,se}$), while

$$\Gamma_c = \partial\Omega \setminus (\Gamma_{sw} \cup \Gamma_{nw} \cup \Gamma_{ne} \cup \Gamma_{se}) \tag{17}$$

denotes the walls. Stream 1 and 2 enter the domain by the entry Γ_{sw} and Γ_{nw}, respectively, and leave it by exit Γ_{ne} and Γ_{se}, respectively.

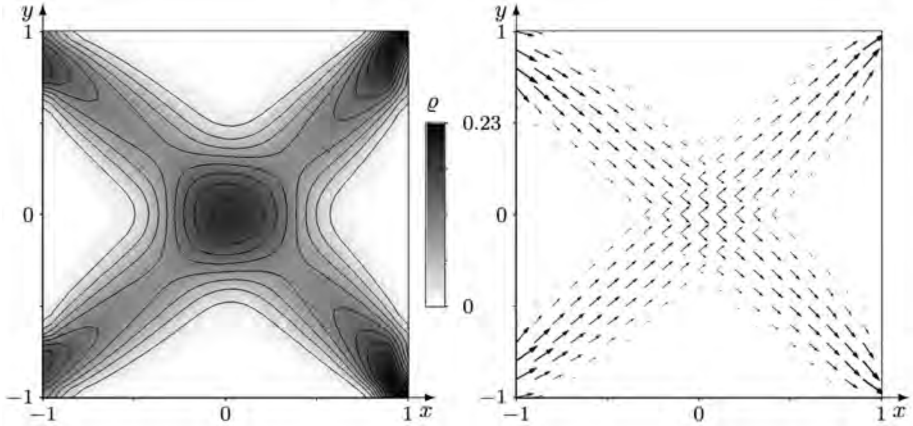


FIGURE 2. Simulation of the the ∇-model. Total density $\varrho = \varrho_1 + \varrho_2$ (left) and intersecting flux fields (right) at time $t = 8.65$ (steady state) for weak influxes $V\varrho_{1,2}|_{\Gamma_{sw,nw}} = 0.0625$.

We consider a situation with two pedestrian streams ($n = 2$) intersecting in an angle of 90°. In the strategic direction fields $\boldsymbol{d}_1^{\mathrm{s}}$ and $\boldsymbol{d}_2^{\mathrm{s}}$ the entrances are adjacent corners. The corresponding exit is located in the remote corners of the square (Figure 1).

The specification of the convective fluxes is completed by setting $a_1 = a_2 = 1$, $V = 1 - \varrho$. In order to illustrate the different qualitative behaviours of the two approaches, we use normalized parameters for the coefficients of the linear diffusion matrix by setting $\varepsilon = 0.01, \delta = 0$ in (4) and $b_1 = b_2 = 1$ for the nonlinear diffusion matrix (15). This choice is motivated by the experimental observation that a larger ε corresponds to a larger diffusion that is spreading pedestrians rather quickly over the available space. This takes place independently of the other phase and the total density. As a consequence, a violation of the condition $\varrho \leq 1$ has been observed for larger values of ε. This effect of undirected diffusion contradicts the assumption that pedestrian streams are oriented towards specific targets. With these specifications the model variants reduce for the linear "Δ-model" to

$$\frac{\partial \varrho_i}{\partial t} + \nabla \cdot \Big(\varrho_i(1-\varrho)\boldsymbol{d}_i^{\mathrm{s}}\Big) = \varepsilon\, \Delta \varrho_i, \quad i = 1,2, \tag{18}$$

and for the nonlinear "∇-model" to

$$\frac{\partial \varrho_i}{\partial t} + \nabla \cdot \Big(\varrho_i(1-\varrho)\boldsymbol{d}_i^{\mathrm{s}}\Big) = \nabla \cdot \Big(\varrho_i\, \varrho\, \chi(\varrho) \nabla \varrho\Big), \quad i = 1,2. \tag{19}$$

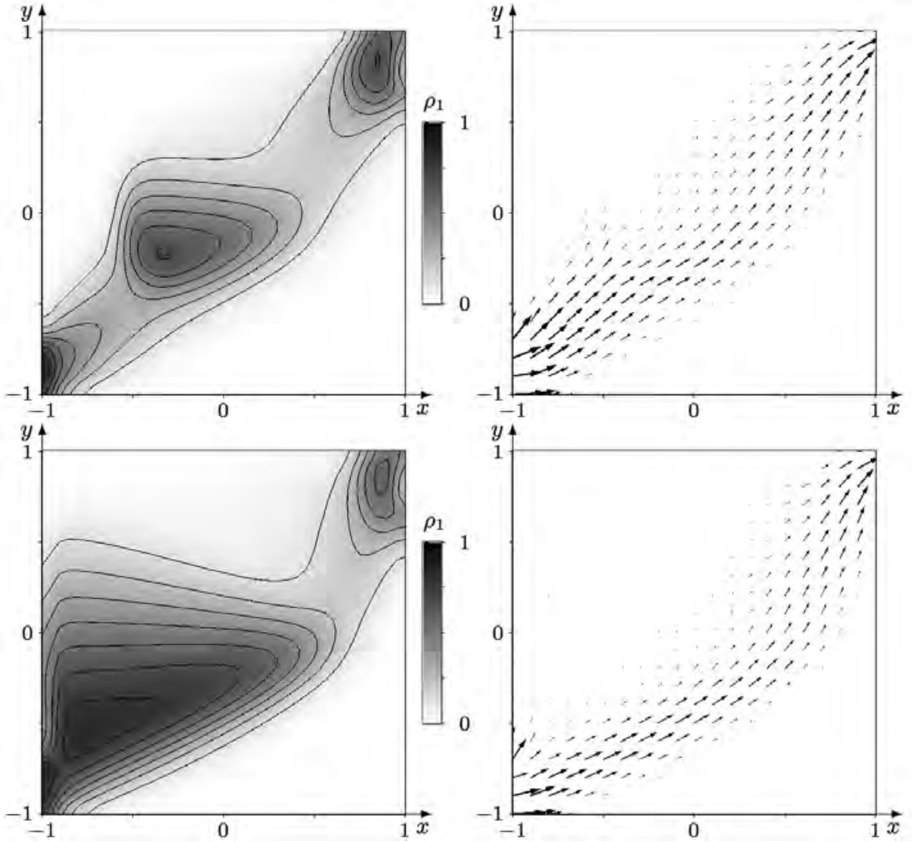


FIGURE 3. Stimulation results for Δ-model. Density ϱ_1 (left) and flux field (right) at times $t = 10$ (top) and $t = 155$ (bottom).

On the walls Γ_c with exterior normal vector ν one would expect zero-flux boundary conditions

$$\begin{aligned} f_\Gamma &= \Big(\varrho_i V \boldsymbol{d}_i^{\mathrm{s}} - \varepsilon\,\nabla\varrho_i\Big)\cdot\nu && \text{for the } \Delta\text{-model (18)},\\ f_\Gamma &= \Big(\varrho_i V \boldsymbol{d}_i^{\mathrm{s}} - \varrho_i\,W\,\chi(\varrho)\nabla\varrho\Big)\cdot\nu && \text{for the } \nabla\text{-model (19)}, \end{aligned} \tag{20}$$

where $V = V(\varrho)$ is a speed function derived form the fundamental diagram and $W = W(V)$ a density-evasion speed weight in Ω. For numerical reasons, these boundary conditions are not sufficient because boundary layer effects would lead to numerical densities greater than one in contradiction to the model. Therefore, we impose the additional conditions

$$\begin{aligned} &\nabla\rho_i\cdot\nu = 0,\ \boldsymbol{d}_i^{\mathrm{s}}\cdot\nu = 0 && \text{for the } \Delta\text{-model (18)},\\ &\rho_i = 0 && \text{for the } \nabla\text{-model (19)}. \end{aligned} \tag{21}$$

The following flux condition at the exits guarantees that all pedestrians entering the region are able to leave it. The flux condition at the entries guarantees a density dependent inflow rate while a constant influx rate would lead to a back flow through the entry in the case of completely occupied regions (or cells in the discrete model)

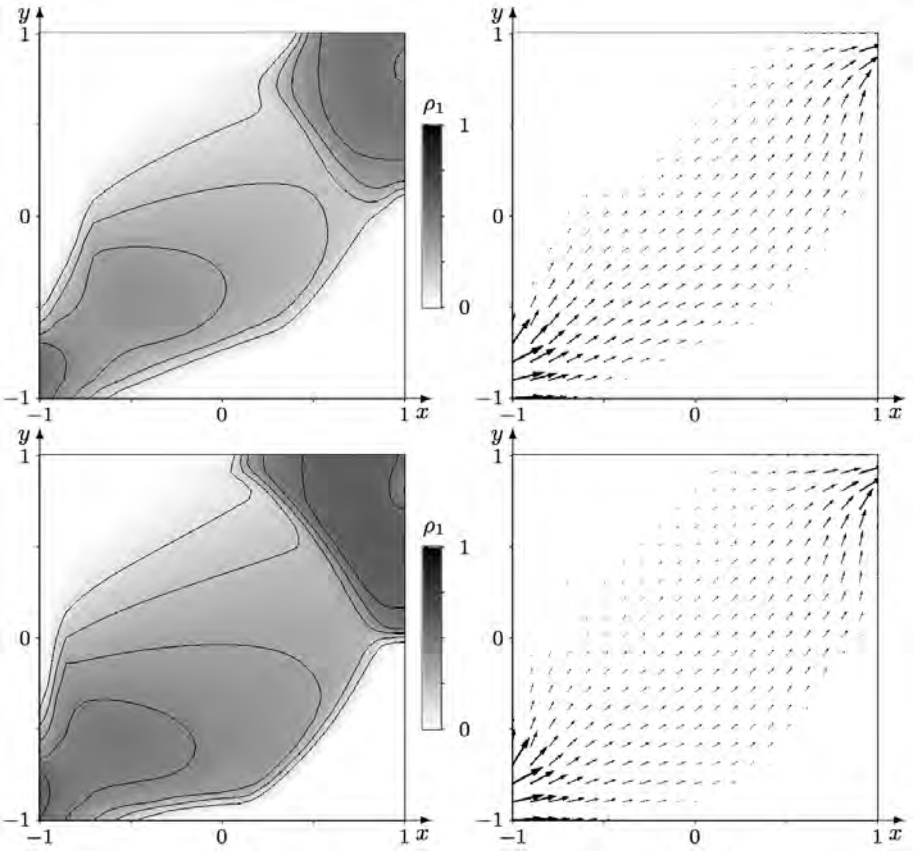


FIGURE 4. Simulation results for ∇-model. Density ϱ_1 (left) and flux field (right) at times $t = 10$ and $t = 395$.

near the entry in contradiction to the model assumptions.

$$
\begin{array}{lll}
\nabla\varrho_i\cdot\nu = 0, \quad \boldsymbol{d}_i^{\mathrm{s}}\cdot\nu = 1, \quad V|_\Gamma = 1 & \text{on} \left\{\begin{array}{ll}\Gamma_{ne} & \text{for } i=1\\ \Gamma_{se} & \text{for } i=2\end{array}\right. & \text{(exits)},\\[2ex]
\begin{array}{l}\varrho_i = \rho_i = \text{const}, \quad \boldsymbol{d}_i^{\mathrm{s}}\cdot\nu = -1,\\ V|_\Gamma = 1 \;\; (\Delta\text{-model}),\\ V|_\Gamma = 1,\, W|_\Gamma = 0 \;\; (\nabla\text{-model}).\end{array} & \text{on} \left\{\begin{array}{ll}\Gamma_{sw} & \text{for } i=1\\ \Gamma_{nw} & \text{for } i=2.\end{array}\right. & \text{(entries)}
\end{array}
\tag{22}
$$

Obviously, numerical stability and conformity with real life situations have been paid for by possibly discontinuous boundary conditions. This is one of the consequences of the fact, that, on the one hand, PDE based macroscopic models are indispensable because of their global properties, but, on the other hand, that for very dense crowds the observed behaviour is only similar to gas dynamics.

The upwind direction is controlled by sign of $\boldsymbol{d}_i^{\mathrm{s}}$ in (18) and the sign of $(1-\varrho)\boldsymbol{d}_i^{\mathrm{s}} - \varrho\,\chi(\varrho)\nabla\varrho$ in (19), respectively. The examples are computed on an 40×40 -grid with regular quadratic cells.

In the case of a 90°-intersection, supercritical flows should behave more symmetric than in a 180°-intersection, where the entry of one phase is the shared exit of the other, such that a complete blocking by one "winning" phase takes place at the entry of the other phase.

For small influxes the intersecting pedestrian streams behave qualitatively quite similar for both models see Figure 2 for a simulation of the ∇-model.

The simulation results for high influxes are compared in Figure 3 and Figure 4 for the Δ-model and the ∇-model, respectively. Since both stream behave symmetrically only stream 1 is shown. In a simulation of the Δ-model, a mutual blocking of both streams occurs in the center of the domain. Also a reciprocal obstruction already near the entries is observed. These blockings are is caused by a missing evasion strategy and gradually growing in time. The locking due to a missing evasion strategy The simulation of the ∇-model in contrast there is no mutual blocking in the center thanks to the avoidance strategy that allows for a better use of the space to pass by each other. The larger tailback at the exits can be naturally explained by the low exit rate. For the ∇-model, we observed a maximum flux of about 0.24375 which is well in accordance with the theoretically expected 0.25. A further increasing of the influx would show the effect of overcrowding with a succeeding breakdown of the flux. This behaviour is better suited to real life situations and probably due to the more realistic densitiy distribution that emerges which includes a tendency to phase separation. As a conclusion, comparing both models, the ∇-model is better adapted at the intended simulation of pedestrian behaviour.

REFERENCES

[1] G. Bärwolff, M.-J. Chen, F. Huth, G. Lämmel, M. Plaue, and H. Schwandt, *Methods for modeling and simulation of multi-destination pedestrian crowds* in "Proceedings of Pedestrian and Evacuation Dynamics", 2012, to appear.

[2] S. Berres, R. Ruiz-Baier, H. Schwandt and E.M. Tory, *A two-dimensional model of pedestrian flow generating pattern formation*, in "Hyperbolic problems – theory, numerics and applications. Volume 1, Ser. Contemp. Appl. Math. CAM, **17**", World Sci. Publishing, Singapore, (2012), 304–311.

[3] S. Berres, R. Ruiz-Baier, H. Schwandt and E.M. Tory, *An adaptive finite-volume method for a model of two-phase pedestrian flow*, Networks and Heterogeneous Media, **6**, (2011), 401–423.

[4] L. Bruno, A. Tosin, P. Tricerri and F. Venuti, *Non-local first-order modelling of crowd dynamics: A multidimensional framework with applications*, Appl. Math. Model., **35**, (2011), 426–445.

[5] W. Daamen, P. H. L. Bovy and S. P. Hoogendoorn, *Modelling pedestrians in transfer stations*, in "Pedestrian and Evacuation Dynamics" (eds. M. Schreckenberg and S. D. Sharma), Springer-Verlag Berlin Heidelberg, 2002, 59–73.

[6] D. Helbing, I. Farkas and T. Vicsek, *Traffic and related self-driven many-particle systems*, Reviews of Modern Physics, **73**, (2001), 1067–1141.

[7] S.P. Hoogendoorn and W. Daamen, *Self-organization in pedestrian flow*, Traff. Granul. Flow, **3**, (2005), 373–382.

[8] Y. Jiang, P. Zhang, S.C. Wong and R. Liu, *A higher-order macroscopic model for pedestrian flows*, Physica A, **389** (2010), 4623–4635.

[9] A. Nakayama, K. Hasebe and Y. Sugiyama, *Instability of pedestrian flow in 2D optimal velocity model with attractive interaction*, Comput. Phys. Comm., **177** (2007), 162–163.

[10] Y. Xia, S.C. Wong, M.P. Zhang, C.-W. Shu and W.H.K. Lam, *An efficient discontinuous Galerkin method on triangular meshes for a pedestrian flow model*, Int. J. Numer. Meth. Engrg. **76**, (2008), 337–350.

E-mail address: sberres@uct.cl
E-mail address: huth@math.tu-berlin.de
E-mail address: schwandt@math.tu-berlin.de
E-mail address: baerwolf@math.tu-berlin.de

SOME RESULTS ON THE TWO-DIMENSIONAL DISSIPATIVE EULER EQUATIONS

Luigi C. Berselli

Dipartimento di Matematica
Università degli Studi di Pisa
Pisa, I-56126, Italy

Abstract. We make a review of some recent results concerning special solutions and behavior at infinity for 2D dissipative Euler equations. In particular, we give a simplified proof –in the space-periodic setting– of the uniform space/time boundedness of the first derivatives of the velocity, under suitable assumptions on the external force and on the dissipation (damping) coefficient. This is used to sketch the proof of existence of almost-periodic solutions.

1. Introduction.

In this paper we summarize some results related with the long-time behavior of the Euler equations for incompressible fluids in two space dimensions. It is well-known that in the 2D case it is possible to prove, for smooth enough data, existence and uniqueness of smooth solution, for all positive times (see also the discussion in the next section for certain less-standard results). It is also clear that without any smoothing or dissipation, one cannot expect to have uniform boundedness of the energy and of other interesting quantities as the enstrophy or higher norms of the velocity. To this end we consider the so-called *dissipative Euler equations*

$$\begin{aligned} \partial_t u + \chi\, u + (u \cdot \nabla)\, u + \nabla p = f \quad & \text{in }]0, +\infty[\times \mathbf{T}, \\ \nabla \cdot u = 0 \quad & \text{in }]0, +\infty[\times \mathbf{T}, \end{aligned} \tag{1}$$

where $u = (u_1, u_2)$ is the velocity of the fluid with the initial condition $u(0) = u_0$, p is the kinematic pressure, $f = f(t, x)$ is the external force field, $\mathbf{T} := (\mathbf{R}/2\pi\mathbf{Z})^2$ is a two dimensional torus and all quantities are 2π-space periodic and with vanishing mean value. The damping term $\chi\, u$ (with $\chi > 0$) models the bottom friction in some 2D oceanic models (when the system is considered in a bounded domain; in that case, the system is called the viscous Charney-Stommel barotropic ocean circulation model of the gulf stream) or the Rayleigh friction in the planetary boundary layer (with space-periodic boundary conditions). The positive constant χ is the Rayleigh friction coefficient (or the Ekman pumping/dissipation constant) or also the sticky viscosity, when the model is used to study motion in presence of rough boundaries, see for instance Gallavotti [10]. Early existence results can be found in Barcilon, Constantin, and Titi [2], while links between the driven and damped 2D Navier-Stokes, attractors, and statistical solutions are proved in Ilyin, Miranville, and Titi [12] and Constantin and Ramos [8]. The model (1) represents (probably) the "weakest" dissipative modification of the Euler equations and results on the

2000 *Mathematics Subject Classification.* Primary: 35Q31; Secondary: 76B03, 35B40, 35L60.

Key words and phrases. Euler equations, continuous and Dini-continuous functions, almost-periodic solutions, transport equation.

long-time behavior of the damped/driven Navier-Stokes do not directly pass to the limit as the "viscosity goes to zero," hence a completely different treatment is required to study the problem without viscosity. This paper is aimed at sketching the fundamental steps needed to show existence of almost-periodic solutions and one key-result is that of showing a sort of asymptotic stability, cf. [15]. In order to use standard tools based on dissipation to construct almost-periodic solutions we need a control on the difference of two solutions. The presence of the nonlinear convection term seems to require an estimate on $\|\nabla u\|_{L^\infty}$. To this end we analyze the equation for the vorticity. Taking the curl of (1) (define $\xi := \operatorname{curl} u := \partial_2 u_1 - \partial_1 u_2$ and $\phi := \operatorname{curl} f$) one obtains

$$\partial_t \xi + \chi\, \xi + (u \cdot \nabla)\, \xi = \phi \quad \text{in }]0, \infty[\times \mathbf{T}, \tag{2}$$

which is a non-local scalar transport equation (with damping), which plays a fundamental role in the sequel.

Moreover, it is well-known that (by the Biot-Savart formula) the velocity can be reconstructed from the vorticity by recalling that $-\Delta u = \nabla^\perp \xi$. Basic Calderon-Zygmund or Schauder estimates for the Poisson equations allow us to state that ∇u and ξ are at the same level of regularity in L^p spaces $(1 < p < \infty)$ or in Hölder spaces $C^{0,\alpha}$. Roughly speaking (full details are given in [5, 6]) the L^p-setting, with $p < +\infty$ is too weak, while the $C^{0,\alpha}$ setting seems too strong in order to obtain uniform estimates. This suggest to use a more precise functional framework and in particular to employ the following well-known potential theoretic result:

$$\exists\, C_0 = C_0(\mathbf{T}) > 0 : \quad \|\nabla u\|_{L^\infty(\overline{\mathbf{T}})} \leq C_0 \|\xi\|_{C_D(\overline{\mathbf{T}})}, \tag{3}$$

to show boundedness of the gradient of u. We recall that the set of Dini-continuous functions $C_D(\overline{\mathbf{T}}) \subset C(\overline{\mathbf{T}})$ is the subset of continuous functions $f : \overline{\mathbf{T}} \to \mathbf{R}$ such that

$$\|f\|_{C_D(\overline{\mathbf{T}})} := \|f\|_{L^\infty(\overline{\mathbf{T}})} + [f]_{C_D} := \|f\|_{L^\infty(\overline{\mathbf{T}})} + \int_0^{\sqrt{2}2\pi} \omega(f, \sigma)\, \frac{d\sigma}{\sigma} < +\infty,$$

where

$$\omega(f, \sigma) := \sup \Big\{ |f(x) - f(y)| : \; x, y \in \overline{\mathbf{T}}, \; 0 < |x - y| < \sigma, \Big\}.$$

The main reason for the use of this functional space to study the vorticity stems in the uniform estimate proved in Proposition 1, which –together with (3)– gives the requested bound. We emphasize that the first use of these spaces for the vorticity of Euler equations dates back to Beirão da Veiga [4] in the context of global well-posedness of the 2D problem. In questions of stability the role of Dini-continuous vorticity has been first recognised by Koch [14], while recent results on global attractors are those proved in [5]. Close relationship between Dini and critical Besov spaces is analyzed in [11]. We consider Stepanov almost-periodic solutions (see [1] for further details), which seems the most natural setting for problems related with the Euler equations. If X is a Banach space we define $L^2_{uloc}(X)$ as the space of *uniformly locally square integrable X-valued functions*

$$L^2_{uloc}(X) := \big\{ v \in L^2_{loc}(\mathbf{R}; X) : \sup_{t \in \mathbf{R}} \int_t^{t+1} \|v(s)\|^2_X \, ds < \infty \big\}.$$

Next, we say that $v \in L^2_{uloc}(X)$ belongs to $\mathcal{S}^2(X)$ or is Stepanov almost-periodic (with values in X) if and only if the set of the time-translates of v is relatively compact with respect to the $L^2_{uloc}(X)$-topology.

The main result of this paper is then the following

Theorem 1.1. *Let be given a divergence-free external force* $f \in \mathcal{S}^2(L^2(\mathbf{T}))$ *with* $\operatorname{curl} f \in L^\infty(\mathbf{R}; C_D(\overline{\mathbf{T}}))$. *There exists* $\chi_0 = \chi_0(f) > 0$ *such that if* $\chi > \chi_0$, *then there exists an almost-periodic solution* $u \in \mathcal{S}^2(L^2(\mathbf{T}))$ *to the dissipative Euler equations* (1).

Remark 1. The condition on χ can be also read as a smallness condition on f. Moreover, by standard results due to Dafermos [9], obtained by compact embedding and interpolation, the solution u will belong also to $\mathcal{S}^2(H^1(\mathbf{T}))$.

Remark 2. Appropriate modifications of the calculations from the next sections can be used to handle also the more general case of a bounded smooth domain $\Omega \subset \mathbf{R}^2$ for the problem endowed with the boundary condition $u \cdot n = 0$ on $\partial\Omega$, see [6] for full details.

The same approach can be also used (with some additional technical steps) to prove, in the case of a time-independent force, the following result concerning the existence of a global attractor, see [5] for full details.

Theorem 1.2. *Let be given* $f \in H^1(\mathbf{T})$ *such that* $\phi = \operatorname{curl} f \in C_D(\overline{\mathbf{T}})$. *There exists* $\chi_0(f) > 0$ *such that if* $\chi > \chi_0$, *then, there exists a global attractor* $\mathcal{A} \subset C(\overline{\mathbf{T}})$, *for the dissipative 2D Euler equations* (1).

Remark 3. Also Thm. 1.2 holds true in a bounded smooth domain $\Omega \subset \mathbf{R}^2$ and moreover the Hausdorff dimension of $\mathcal{A}$ turns out to be finite, cf. [5]

2. **Existence of weak solutions.** In this section we recall some basic results on existence and uniqueness of weak solutions, proved in Bessaih and Flandoli [7], by adapting classical results by Yudovich [16] and Bardos [3]. Let $\mathcal{V}$ be the space of infinitely differentiable, periodic, divergence-free, and with vanishing mean value vector-fields on $\mathbf{T}$. We introduce the usual Hilbert space H defined as the closure of $\mathcal{V}$ with respect to the norm $|\cdot|$ of $L^2(\mathbf{T})^2$, with the inner product of $L^2(\mathbf{T})^2$, denoted in the sequel by $\langle\cdot,\cdot\rangle$. As usual, V is the closure of $\mathcal{V}$ with respect to the norm $\|\cdot\|$ of $H^1(\mathbf{T})^2$. Identifying H with its dual H', and H' with the corresponding natural subspace of V', we have the standard Gelfand triple $V \subset H \subset V'$ with continuous and dense injections. (For simplicity we denote the dual pairing between V and V' by the same symbol as for the inner product of H.)

Definition 2.1. We say that the vector field $u \in C(0,\infty;H) \cap L^\infty_{loc}(0,\infty;V)$, with $\partial_t u \in L^2_{loc}(0,\infty;V')$, is a weak solution to (1) on $[0,\infty[$ if the following properties hold $\forall\, v \in \mathcal{V}$ and all $t \geq t_0 \geq 0$:

$$\|u(t)\|^2 \leq \|u(t_0)\|^2 \mathrm{e}^{-\chi(t-t_0)} + \chi^{-1} \int_{t_0}^t \|f(s)\|^2 \mathrm{e}^{-\chi(t-s)}\, ds,$$

$$|u(t)|^2 + 2\chi \int_{t_0}^t |u(s)|^2\, ds \leq |u(t_0)|^2 + \int_{t_0}^t \langle f(s), u(s)\rangle\, ds,$$

$$\langle u(t) - u(t_0), v\rangle + \chi \int_{t_0}^t \langle u(s), v\rangle\, ds + \int_{t_0}^t \langle (u(s)\cdot\nabla)\, u(s), v\rangle\, ds = \int_{t_0}^t \langle f(s), v\rangle\, ds.$$

We have the following result:

Theorem 2.2. *Let be given* $u_0 \in V$ *and* $f \in L^1_{loc}(0,+\infty;V)$. *Then, there exists at least a weak solution to* (1). *Moreover, if* $\operatorname{curl} u_0 \in L^\infty(\mathbf{T})$ *and* $\operatorname{curl} f \in L^1_{loc}(0,+\infty;L^\infty(\mathbf{T}))$, *such a solution is unique.*

Proof. The proof of this result is classically based on a vanishing-viscosity approximation. The Navier-Stokes equations are considered for $\nu > 0$

$$\begin{aligned} \partial_t u^\nu + \chi\, u^\nu + (u^\nu \cdot \nabla)\, u^\nu - \nu\Delta u^\nu + \nabla p^\nu = f \quad &\text{in }]0,T[\times\mathbf{T}, \\ \nabla\cdot u^\nu = 0 \quad &\text{in }]0,T[\times\mathbf{T}, \end{aligned}$$

for which existence of Leray-Hopf weak solutions in $[0,T]$ for any positive T is well-known. Next, by using the vorticity equation for $\xi^\nu = \operatorname{curl} u^\nu$ it is easy to prove (along Galerkin approximation) that

$$\frac{d}{dt}|\xi^\nu(t)|^2 + \chi|\xi^\nu(t)|^2 + \nu|\nabla\xi^\nu(t)|^2 \leq \frac{1}{\chi}|\phi|^2,$$

which can be used to show an uniform bound for the vorticity in $L^2(\mathbf{T})$. Then, with this it is possible to show that the limit $u := \lim_{\nu\to 0^+} u^\nu$ is a weak solution to the dissipative Euler equations.

The uniqueness in the case of a bounded vorticity for the Euler equations is more delicate, and it is based on the inequality proved in [16].

$$\exists\, C > 0,\ \text{independent of } p: \quad \|u\|_{L^p(\mathbf{T})} \leq C\sqrt{p}\|u\|_{W^{1,2}(\mathbf{T})} \qquad \forall\, p \geq 2.$$

□

Since we have a unique solution of (1) we can prove better regularity on it simply by using representation formulas. It is well-known that if $\xi \in L^\infty(\mathbf{T})$, then this is not enough to have $\nabla u \in L^\infty(\mathbf{T})$ (being the endpoint estimate) hence Lipschitz characteristics. The boundedness of the vorticity implies that the velocity is Lip-Log (called also quasi-Lipschitz) and then that the characteristics are unique and Hölder continuous. In particular, the following result is well-known, see for instance [13].

Lemma 2.3. *Let* $|||\xi||| := \sup_{(s,y)\in[0,T]\times\mathbf{T}} |\xi(s,y)|$, *then there exists a constant* $c > 0$ *such that, for all* $x, x_1 \in \mathbf{T}$ *such that* $|x - x_1| < 1$

$$|u(t,x) - u(t,x_1)| \leq c\,|||\xi|||\,|x - x_1|[1 - \log(|x - x_1|)].$$

If $U(s,t,x)$ *denotes the solution of the Cauchy problem*

$$\begin{cases} \dfrac{dU(t,s,x)}{dt} = u(t,U(t,s,x)), \\ U(s,s,x) = x, \end{cases} \tag{4}$$

then, defining δ *as follows* $\delta := \mathrm{e}^{-c|||\xi|||\,T}$, *it holds*

$$|U(s,t,x) - U(s_1,t_1,x_1)| \leq c|||\xi|||\,|t - t_1| + \mathrm{e}(1 + \mathrm{e}\,c|||\xi|||)(|x - x_1|^\delta + |s - s_1|^\delta).$$

In order to have Lipschitz characteristics, it would be enough, to have bounded gradient of the velocity, which will follow from Dini-continuous vorticity.

Moreover, remaining in the setting of Hölder functions it follows (by direct computation) that the composition of Dini and of an Hölder continuous functions is again a Dini-continuous function.

Lemma 2.4. *Let be given* $f \in C_D(\overline{\mathbf{T}})$ *and* $U \in C^{0,\delta}(\overline{\mathbf{T}})$, *then the following estimate for the Dini's semi-norm holds true:*

$$[f\circ U]_{C_D} \leq \frac{1}{\delta}[f]_{C_D} + \frac{2}{\delta}\log[U]_\delta(\sqrt{2}2\pi)^{\delta-1}.$$

Since the Hölder exponent of the characteristics decreases with time, we first fix an interval $[0,T]$ and the previous lemma allows to control the Dini-norm of the vorticity, by using the representation formula obtained by following the characteristics in the equation for the vorticity

$$\xi(t,x)=\xi_0(U(0,t,x))\,\mathrm{e}^{-\chi t}+\int_0^t \phi(s,U(s,t,x))\,\mathrm{e}^{-\chi(t-s)}\,ds, \qquad t\in[0,T]. \tag{5}$$

By using Lemma 2.4, formula (5), and by reasoning as in [4, 14] one can easily show that if $\xi_0\in C_D(\overline{\mathbf{T}})$ and $\phi\in L^1_{\mathrm{loc}}(0,+\infty;C_D(\overline{\mathbf{T}}))$, then $\xi\in L^\infty(0,T;C_D(\overline{\mathbf{T}}))$, for all positive T.

Remark 4. By using the Schauder's fixed point theorem (employed in two slightly different manners in Ref. [4, 14]) it is possible also to show that $\xi\in C([0,T];C_D(\overline{\mathbf{T}}))$, for all $T>0$, but this is not needed here.

For our purposes the continuity is not so important, but what will be relevant is the following result.

Proposition 1. *Let $u_0\in V$ such that $\xi_0\in C_D(\overline{\mathbf{T}})$ and $\phi\in L^\infty(0,+\infty;C_D(\overline{\mathbf{T}}))$. Then, for large enough $\chi>0$, the Dini-norm of ξ is uniformly bounded over $[0,+\infty[$.*

Proof. We are assuming that we have a unique solution $\xi\in L^\infty(0,T;C_D(\overline{\mathbf{T}}))$ of the transport equation (2), for any given $T>0$. Then for a.e. $t\in[0,T]$ it follows $\nabla u(t,\cdot)\in L^\infty$ and U is Lipschitz continuous (especially in the space variable) and the Lip-norm depends on the Dini-norm of ξ. More precisely, we have the estimate

$$|\nabla U(s,t,x)|\le \mathrm{e}^{\int_s^t \sup_{y\in\mathbf{T}}|\nabla u(\tau,y)|\,d\tau} \qquad \text{for } (s,t,x)\in[0,T]^2\times\mathbf{T}, \tag{6}$$

but, since the bound on $\|\nabla u\|_{L^\infty}$ depends on $\|\xi(t)\|_{C_D}$, it may depend on $T>0$. To show an uniform bound we first observe that the L^∞ bound for the vorticity (shown also in [7]) follows directly from (5) and it is independent of T:

$$\|\xi(t)\|_{L^\infty}\le\|\xi_0\|_{L^\infty}\mathrm{e}^{-\chi t}+\sup_{t\ge0}\|\phi(t)\|_{L^\infty}\frac{1-\mathrm{e}^{-\chi t}}{\chi}.$$

We estimate the Dini-continuity of $\eta=\xi\,\mathrm{e}^{\chi t}$ on $[0,T]$. Observe that, for η we have the representation formula $\eta(t,x)=\xi_0(U(0,t,x))+\int_0^t\phi(s,U(s,t,x))\,\mathrm{e}^{\chi s}\,ds$, and

$$\|\eta(t)\|_{L^\infty}\le\|\xi_0\|_{L^\infty}+\sup_{t\ge0}\|\phi(t)\|_{L^\infty}\frac{\mathrm{e}^{\chi t}-1}{\chi}.$$

Moreover, we observe that $[\eta(t)]_{C_D}=[\xi(t)]_{C_D}\mathrm{e}^{\chi t}$, and we split it as follows:

$$\begin{aligned}
[\eta(t)]_{C_D}&:=\int_0^1\sup_{|x-y|\le\rho}|\eta(t,x)-\eta(t,y)|\frac{d\rho}{\rho}\\
&\le\int_0^1\sup_{|x-y|\le\rho}|\xi_0(U(0,t,x))-\xi_0(U(0,t,y))|\frac{d\rho}{\rho}\\
&\quad+\int_0^t\int_0^1\sup_{|x-y|\le\rho}|\phi(s,U(s,t,x))-\phi(s,U(s,t,y))|\,\mathrm{e}^{\chi s}\frac{d\rho}{\rho}ds\\
&=:B_1+B_2.
\end{aligned} \tag{7}$$

By making a change of variable by means of the unitary diffeomorphism $U(0,t,x)$ we have that

$$\begin{aligned}
B_1 &\leq \int_0^1 \sup_{|x-y|\leq\rho\|\nabla U(0,t,\cdot)\|_{L^\infty}} |\xi_0(x)-\xi_0(y)|\frac{d\rho}{\rho}\\
&\leq \int_0^1 \sup_{|x-y|\leq\rho}|\xi_0(x)-\xi_0(y)|\frac{d\rho}{\rho} + 2\|\xi_0\|_{L^\infty}\int_1^{\|\nabla U(0,t,\cdot)\|_{L^\infty}}\frac{d\rho}{\rho}\\
&\leq [\xi_0]_{C_D} + 2\|\xi_0\|_{L^\infty}\log\|\nabla U(0,t,\cdot)\|_{L^\infty},
\end{aligned}$$

and, by appealing to (6), we get

$$B_1 \leq [\xi_0]_{C_D} + 2\|\xi_0\|_{L^\infty}\int_0^t \|\nabla u(s)\|_{L^\infty}ds \quad \leq [\xi_0]_{C_D} + 2C_0\|\xi_0\|_{L^\infty}\int_0^t \|\eta(s)\|_{C_D}\,ds.$$

Concerning B_2, by making the change of variables by means of $U(s,t,x)$, we have

$$\begin{aligned}
B_2 &\leq \int_0^t\int_0^1 \sup_{|x-y|\leq\rho\|\nabla U(s,t,\cdot)\|_{L^\infty}}|\phi(s,x)-\phi(s,y)|\frac{d\rho}{\rho}\mathrm{e}^{\chi s}\,ds\\
&\leq \int_0^t[\phi(s)]_{C_D}\mathrm{e}^{\chi s}ds + 2\|\phi(s)\|_{L^\infty}\int_0^t\int_1^{\|\nabla U(s,t,\cdot)\|_{L^\infty}}\frac{d\rho}{\rho}\mathrm{e}^{\chi s}ds\\
&\leq \sup_{t\geq0}[\phi(t)]_{C_D}\int_0^t \mathrm{e}^{\chi s}ds + 2\sup_{t\geq0}\|\phi(t)\|_{L^\infty}\int_0^t\log\|\nabla U(s,t,\cdot)\|_{L^\infty}\mathrm{e}^{\chi s}ds\\
&\leq \sup_{t\geq0}[\phi(t)]_{C_D}\int_0^t \mathrm{e}^{\chi s}ds + 2\sup_{t\geq0}\|\phi(t)\|_{L^\infty}\int_0^t\log\|\nabla U(s,t,\cdot)\|_{L^\infty}\mathrm{e}^{\chi s}ds\\
&\leq \sup_{t\geq0}[\phi(t)]_{C_D}\int_0^t \mathrm{e}^{\chi s}ds + 2\sup_{t\geq0}\|\phi(t)\|_{L^\infty}\int_0^t\int_s^t\|\nabla u(\tau)\|_{L^\infty})\mathrm{e}^{\chi s}d\tau ds.
\end{aligned}$$

Changing the order of integration in the last integral we have

$$\begin{aligned}
B_2 &\leq \sup_{t\geq0}[\phi(t)]_{C_D}\int_0^t \mathrm{e}^{\chi s}ds + 2\sup_{t\geq0}\|\phi(t)\|_{L^\infty}\int_0^t\int_0^\tau\|\nabla u(\tau)\|_{L^\infty}\mathrm{e}^{\chi s}ds d\tau\\
&\leq \sup_{t\geq0}[\phi(t)]_{C_D}\frac{\mathrm{e}^{\chi t}}{\chi} + \frac{2C_0}{\chi}\sup_{t\geq0}\|\phi(t)\|_{L^\infty}\int_0^t\|\eta(\tau)\|_{C_D}d\tau.
\end{aligned}$$

Collecting all the estimates and by defining $\Phi := \sup_{t\geq0}\|\phi(t)\|_{C_D}$ we arrive at

$$\|\eta(t)\|_{C_D} \leq \|\xi_0\|_{C_D} + \frac{2\Phi}{\chi}\mathrm{e}^{\chi t} + 2C_0\Big[\|\xi_0\|_{C_D} + \frac{\Phi}{\chi}\Big]\int_0^t\|\eta(s)\|_{C_D}\,ds.$$

By using Gronwall lemma and by coming back to the variable ξ we get

$$\begin{aligned}
\|\xi(t)\|_{C_D} \leq \Bigg[\|\xi_0\|_{C_D} + \frac{2\Phi}{\chi} - \frac{2\Phi\chi}{\chi^2-2C_0(\Phi+\|\xi_0\|_{C_D}\chi)}\Bigg]\ \mathrm{e}^{t\Big[\frac{2C_0(\Phi+\|\xi_0\|_{C_D}\chi)}{\chi}-\chi\Big]}\\
+\frac{2\Phi\chi}{\chi^2-2C_0(\Phi+\|\xi_0\|_{C_D}\chi)}
\end{aligned}$$

which is uniformly bounded on $[0+\infty[$ if $2C_0\Phi + 2C_0\|\xi_0\|_{C_D}\chi - \chi^2 < 0$, that is if $\chi > \chi_0 := C_0\|\xi_0\|_{C_D} + \sqrt{C_0^2\|\xi_0\|_{C_D}^2 + 2C_0\Phi}$, ending the proof. □

We are now ready to proceed to the proof of the main result. The first step consists in proving existence of weak solution defined for all $t \in \mathbf{R}$. This is classically obtained by constructing solutions of the following problems

$$\begin{cases} \partial_t u_n + \chi\, u_n + (u_n \cdot \nabla)\, u_n + \nabla p_n = f & \text{in }]-n, +\infty[\times \mathbf{T}, \\ \nabla \cdot u_n = 0 & \text{in }]-n, +\infty[\times \mathbf{T}, \\ u_n(-n) = 0 & \text{in } \mathbf{T}. \end{cases} \tag{8}$$

The results of the previous section show the following result.

Proposition 2. *Under the hypotheses of Thm. 1.1, for $\chi > \sqrt{2C_0\Phi}$ the unique weak solution of (8) satisfies $u_n \in L^\infty(-n, +\infty; V)$ and $\operatorname{curl} u_n \in L^\infty(-n, +\infty; C_D(\overline{\mathbf{T}}))$.*

By extending u_n to zero for $t < n$ and by standard compactness tools it follows that $u_n \overset{*}{\rightharpoonup} u$ in $L^\infty(\mathbf{R}; V)$ where u is a weak solution to the dissipative Euler equations on the whole line. The uniform bounds on $\|\nabla u_n\|_{L^\infty}$ imply also that, for χ large enough

$$\exists\, C_2 = C_2(f, \chi): \qquad \sup_{t\in\mathbf{R}} \|\nabla u(t)\|_{L^\infty(\mathbf{T})} \leq C_2 < +\infty, \tag{9}$$

With the above estimate at hand we can give an outline of an existence result for almost-periodic solutions.

Sketch of the Proof of Thm. 1.1. The condition that f is $S^2(H)$-almost-periodic reads: for any sequence $\{r_m\}$ there exists a sub-sequence $\{r_{m_k}\}$ and a function $\widetilde{f}(t, x)$ such that

$$\sup_{t\in\mathbf{R}} \int_t^{t+1} |f(\tau + r_{m_k}) - \widetilde{f}(\tau)|^2 d\tau \to 0.$$

As in [15, §4], we proceed by contradiction. Therefore, there is a weak solution u to (1) and a sequence $\{h_m\}$ such that

$$\sup_{t\in\mathbf{R}} \int_t^{t+1} |f(\tau + h_m) - \widetilde{f}(\tau)|^2 d\tau \to 0,$$

and there exist three sequences $\{t_k\}$, $\{h_{m_k}\}$, $\{h_{n_k}\}$ and a positive constant $\delta_0 > 0$ such that

$$\int_{t_k}^{t_k+1} |u(s + h_{m_k}) - u(s + h_{n_k})|^2\, ds \geq \delta_0, \qquad \forall k \in \mathbf{N}. \tag{10}$$

Since f is $S^2(H)$-almost-periodic, there exist $f^*(x, t)$ such that

$$\sup_{t\in\mathbf{R}} \int_t^{t+1} |f(\tau + t_k + h_{m_k}) - f^*(\tau)|^2 d\tau \to 0,$$

$$\sup_{t\in\mathbf{R}} \int_t^{t+1} |f(\tau + t_k + h_{n_k}) - f^*(\tau)|^2 d\tau \to 0.$$

By defining the maps $u_1^k(s) := u(s + t_k + h_{m_k})$ and $u_2^k(s) := u(s + t_k + h_{n_k})$, inequality (10) can be rewritten as follows

$$\delta_0 \leq \int_{t_k}^{t_k+1} |u_1^k(s - t_k) - u_2^k(s - t_k)|^2 ds = \int_0^1 |u_1^k(s) - u_2^k(s)|^2\, ds. \tag{11}$$

Using the a priori bounds on u we can extract a sub-sequence $\{u_i^{k_l}\}$ of $\{u_i^k\}$, $i = 1, 2$, strongly convergent to u_i in $L^2_{loc}(\mathbf{R}; H)$, for $i = 1, 2$, respectively. Hence, we can pass to the limit in (11) to get

$$\delta_0 \leq C \int_0^1 |u_1(s) - u_2(s)|^2 ds. \tag{12}$$

By studying the difference $u_1 - u_2$ on the interval $[t_0, 1]$, with $t_0 < 0$, one can show (by using (9)) that $\int_0^1 |u_1(s) - u_2(s)|^2 \, ds$ can be made smaller than any positive constant, by taking t_0 sufficiently small. This shows a contradiction and ends the proof. □

REFERENCES

[1] L. Amerio and G. Prouse. "Almost-periodic functions and functional equations," Van Nostrand Reinhold Co., New York, 1971.

[2] V. Barcilon, P. Constantin, and E.S. Titi. Existence of solutions to the Stommel-Charney model of the Gulf Stream. SIAM J. Math. Anal., **19** (1988), 1355–1364.

[3] C. Bardos. Existence et unicité de la solution de l'équation d'Euler en dimension deux. J. Math. Anal. Appl., **40** (1972), 769–790.

[4] H. Beirão da Veiga. On the solutions in the large of the two-dimensional flow of a nonviscous incompressible fluid. J. Differential Equations, **54** (1984), 373–389.

[5] L.C. Berselli. On the long-time behavior of 2D dissipative Euler equations: The role of (Dini)continuous vorticity. In preparation.

[6] L.C. Berselli and L. Bisconti. Almost-periodic solutions to a 2D dissipative Euler equation. Submitted.

[7] H. Bessaih and F. Flandoli. Weak attractor for a dissipative Euler equation. J. Dynam. Differential Equations, **12** (2000), 713–732.

[8] P. Constantin and F. Ramos. Inviscid limit for damped and driven incompressible Navier-Stokes equations in $\mathbb{R}^2$. Comm. Math. Phys., **275** (2007), 529–551.

[9] C.M. Dafermos. *Almost periodic processes and almost periodic solutions of evolution equations*, In "Dynamical systems" (Proc. Internat. Sympos., Gainesville, Fl., 1976), Acad. Press, (1977) N.Y., 43–57.

[10] G. Gallavotti. "Foundations of fluid dynamics," Texts and Monographs in Physics. Springer-Verlag, Berlin, 2002.

[11] T. Hmidi. Private communication.

[12] A.A. Ilyin, A. Miranville, and E. S. Titi. Small viscosity sharp estimates for the global attractor of the 2-D damped-driven Navier-Stokes equations. Commun. Math. Sci., **2** (2004), 403–426.

[13] T. Kato. On classical solutions of the two-dimensional nonstationary Euler equation. Arch. Rational Mech. Anal., **25** (1967), 188–200.

[14] H. Koch. Transport and instability for perfect fluids. Math. Ann., **323** (2002), 491–523.

[15] P. Marcati and A. Valli. Almost-periodic solutions to the Navier-Stokes equations for compressible fluids. Boll. Un. Mat. Ital. B (6), **4** (1985), 969–986.

[16] V.I. Yudovich. Non stationary flow of an ideal incompressible liquid. Comput. Math. Math. Phys., **3** (1963), 1407–1456. (Russian.)

E-mail address: `berselli@dma.unipi.it`

ADER-SCHEMES ON NETWORKS OF SCALAR CONSERVATION LAWS

Raul Borsche and Jochen Kall

Fachbereich Mathematik
Technische Universität Kaiserslautern, Germany

Abstract. In this paper we extend the ADER-approach [12] onto networks of scalar hyperbolic conservation laws. Within a single edge of the network the classical scheme can be used. At the nodes we use the shrinking stencils of [11] to compute the outflow. The inflow conditions are computed by applying the inverse Lax-Wendroff method to the coupling conditions. Numerical examples confirm the aimed rates convergence and the stability for discontinuous solutions.

1. **Introduction.** Networks of hyperbolic conservation laws have a wide field of applications. Classical examples are the flow of water or gas in networks of pipelines [1, 3], as well as the human circulatory system [6]. Networks of scalar conservation laws can be used to model traffic flow in a network of roads [2, 9], the productivity of supply chains [8] or telecommunication networks [5]. All these applications need robust and accurate numerical methods in order to resolve the involved flow phenomena.

One possible choice are ADER-schemes [12]. This class of schemes combines high order accuracy in regions of smooth data and robustness in the presence of discontinuous solutions. They especially provide good numerical results for waves traveling over long times [12], which is important for networks of large spatial extent. One main difficulty lies in capturing accurately the solution at the nodes. Since the structure at the coupling points can be treated similarly to classical boundary value problems [1], the present approach adapts the ideas of [11], in order to maintain the high order of accuracy across the nodes .

In the following section the notation of the network problem is shortly recalled. In section 3 we present the numerical method to compute the solutions at the nodes. First the shrinking stencil approach is recalled for the outflow edges at a node. For the inflow edges the temporal derivatives are determined by a successive use of the coupling conditions and its derivatives. The spatial information is restored via the inverse Lax-Wendroff procedure. In section 4 some numerical examples for networks of advection equations and the Burgers equations are presented.

2. **Network Notation.** We consider a network $\mathcal{N} = (E, N)$ consisting of a set of edges E and a set of nodes N. Along each edge $e \in E$ the quantity u_e evolves in time according to a scalar hyperbolic conservation law

$$(u_e(t,x))_t + f_e(u_e(t,x))_x = 0 \qquad\qquad x \in [0, L_e] \ . \tag{1}$$

2000 *Mathematics Subject Classification.* Primary: 58F15, 58F17; Secondary: 53C35.

Key words and phrases. ADER, coupling, hyperbolic conservation laws, network.

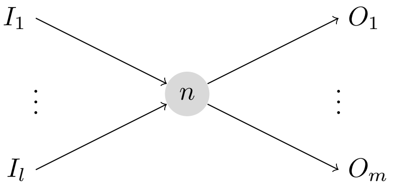


FIGURE 1. A single node n with l incoming and m outgoing edges.

Here $t \in \mathbb{R}^+$ denotes the time, x the position on the edge, L_e the length of the edge and f_e is the flux function.

At each end of the edge suitable boundary or coupling conditions have to be prescribed. For classical boundary conditions we refer to [4] and only recall the notation at a node. The coupling conditions at a single node $n \in N$ connecting the set of edges $E(n) = \{e_1, ..., e_m\}$ are given by

$$\Phi\left(u_{e_1}(t,0), ..., u_{e_m}(t,0)\right) = 0 \ . \tag{2}$$

For simplicity we assume w.l.o.g. that all edges connected to node n are numbered from 1 to m and that in each edge x starts at node n. The vector valued function Φ usually represents a system of nonlinear equations. The well-posedness of such coupling conditions is studied e.g. in [3] and the references therein. The main assumptions are that no sonic points occur and in the scalar case the following condition holds:

$$\det\left[\frac{\partial}{\partial u_{e_1}}\Phi \dots \frac{\partial}{\partial u_{e_k}}\Phi\right] \neq 0 \ , \tag{3}$$

i.e. to each outgoing edge, $e_1, \dots, e_k$, a value can be assigned.

A common example for coupling conditions at a node n, with incoming edges $I(n)$ and outgoing ones $O(n)$ (Figure 1), is the following set of equations

$$\begin{aligned} 0 &= \sum_{e \in I(n)} f_e(u_e) - \sum_{e \in O(n)} f_e(u_e) \ , \\ f_l(u_l) &= \alpha_l \sum_{e \in I(n)} f_e(u_e) \quad \forall l \in O(n) \ . \end{aligned} \tag{4}$$

The first equation assures the conservation of mass at the node. The second one distributes the incoming mass onto the outgoing edges according to the distribution parameters $0 \leq \alpha_l \leq 1$, $l \in O(n)$ with $\sum_{l \in O(n)} \alpha_l = 1$. Conditions of this type are used in e.g. [7, 9].

3. **Numerical Scheme.** The conservation law (1) along the edges is solved using a generic ADER scheme [12] with the classical WENO reconstruction [10]. For further details we refer to the given references and focus in the following on the behavior at the nodes.

3.1. **Outflow ends.** At a node, as well as at free outflow ends, no information from outside the given edge is needed. We therefore use a one sided WENO extrapolation at the boundary, a so called shrinking stencil as in [11].

The idea is, instead of mixing shifted stencils in a non-oscillatory way, to combine a series of shortening stencils, as depicted in Figure 2 for e_1. This resulting reconstruction polynomial is used to extrapolate the data within the computational domain onto the ghost cells across the boundary. The actual update can then be performed by the scheme used inside the edge.

The shrinking stencil only needs some mild modifications compared to the classical WENO approach. For example the smoothness indicators, detecting possible discontinuities, can be chosen as usual, except for the smallest stencil of size one. The computations in [11] show, that for smooth data the magnitude of β_0 is of order $\mathcal{O}(\Delta_x^2)$, which motivates the following choices

$$\beta_0 = (\Delta_x)^2 \ , \qquad\qquad \beta_r = \sum_{l=1}^{k-1} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \Delta_x^{2l-1} \left(\frac{\partial^l p_r(x)}{\partial^l x} \right)^2 dx \ .$$

Here p_r denotes the reconstruction polynomial of the sub-stencil of length $r \leq k_{\max}$. According to these choices the prototype weights now change depending to the stencil length:

$$d_k = \begin{cases} \Delta_x^{k_{\max}-1-k} & 0 \leq i < k_{\max}-1 \\ 1 - \sum_{l=0}^{k_{\max}-2} \Delta_x^{k_{\max}-1-l} & k = k_{\max} \end{cases} \ .$$

The actual weights for the convex combination of the different reconstruction polynomials can be computed as in the classical WENO procedure

$$\alpha_r = \frac{d_r}{(\epsilon + \beta_r)^2} \ , \qquad\qquad \omega_r = \frac{\alpha_r}{\sum \alpha_i} \ .$$

Finally the resulting non-oscillatory reconstruction polynomial

$$p_{k_{\max}}^{\text{shrink}}(x) = \sum_{r=1}^{k_{\max}} \omega_r p_r(x)$$

can be used to compute the cell averages of the ghost cells. This approach maintains the full order of accuracy in case of smooth data, but as well provides sharp resolutions in case of strong shock waves.

3.2. **Inflow ends.** For the inflow ends of the edges, we distinguish between classical boundaries and inflow from other edges at a node. For the inflow ends of the edges, we distinguish between classical boundaries and inflow from other edges at a node. The classical boundaries can be treated exactly as described in [11] or considered as a node without connection to other edges. The following procedure is used to fill the missing information in the ghost cells of the edges, but small modifications might allow a direct incorporation into the ADER approach.

Solving the coupling conditions (2) only restores information of order zero at the node [9, 2]. The temporal derivatives are governed by the derivatives of the coupling conditions. In order to achieve a higher order of accuracy we therefore transfer the available spatial information from the incoming edges into temporal derivatives at the node using the Lax-Wendroff or Cauchy-Kowalewski (CK) procedure. We translate these values into temporal derivatives of the inflow edges by successive application of the temporal derivatives of the coupling conditions (CC^l). By the help of the inverse Lax-Wendroff or Kowalewski-Cauchy (KC) procedure, these data can be converted into spatial derivatives and thus a full reconstruction polynomial is obtained, as sketched in Figure 2.

At a given junction we can use the reconstruction polynomials of the edges flowing into the node, obtained from the shrinking stencils from section 3.1, to collect all the ingoing information. These spatial reconstruction polynomials can be transformed

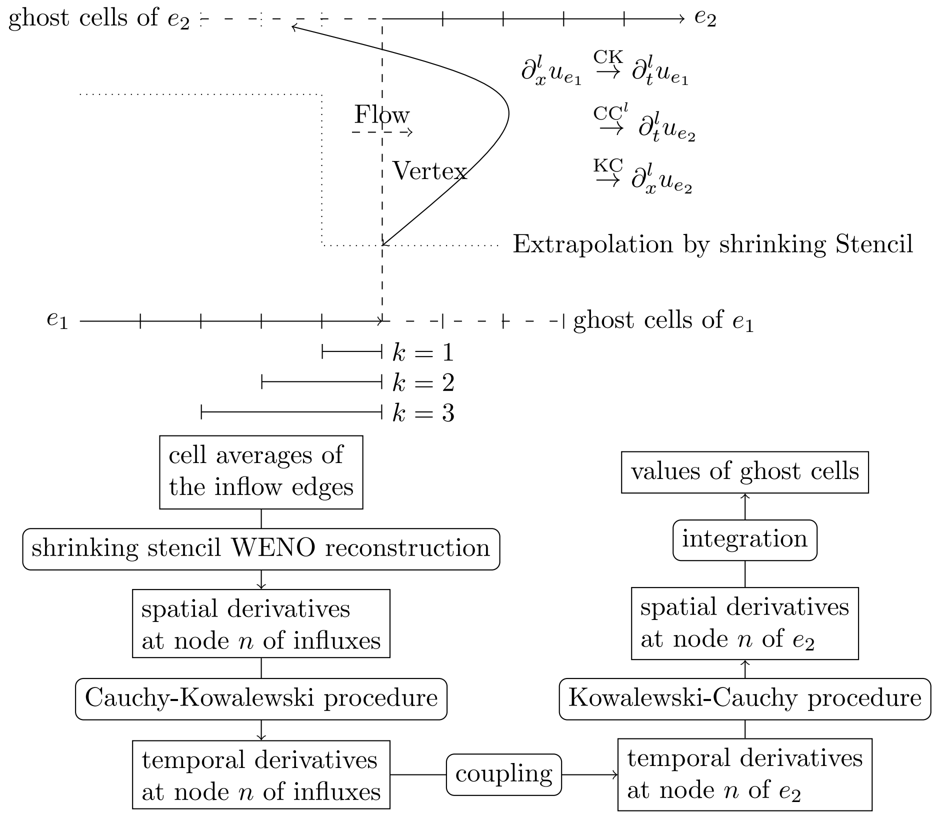

ghost cells of e_2
e_2
$\partial_x^l u_{e_1} \overset{\text{CK}}{\to} \partial_t^l u_{e_1}$
$\overset{\text{CC}^l}{\to} \partial_t^l u_{e_2}$
$\overset{\text{KC}}{\to} \partial_x^l u_{e_2}$
Flow
Vertex
Extrapolation by shrinking Stencil
e_1
ghost cells of e_1
$k = 1$
$k = 2$
$k = 3$
cell averages of the inflow edges
shrinking stencil WENO reconstruction
spatial derivatives at node n of influxes
Cauchy-Kowalewski procedure
temporal derivatives at node n of influxes
coupling
temporal derivatives at node n of e_2
Kowalewski-Cauchy procedure
spatial derivatives at node n of e_2
integration
values of ghost cells

FIGURE 2. A schematic representation of the coupling process.

into polynomials in time at the boundary by the Cauchy-Kavalevsky procedure. As this is done in the same way as for the ADER approach, no additional computational effort is required.

The transfer of the information from the outflow ends onto the inflow ends is managed by the coupling conditions. The leading terms of the temporal polynomials of the edges leaving the node is computed by directly solving the coupling conditions (2). For the temporal derivatives we consider the derivatives of the coupling conditions w.r.t. time

$$\frac{d^l}{dt^l}\Phi(u_{e_1}(t,0),...,u_{e_m}(t,0)) = 0 \qquad\qquad \forall\ 1 \le l \le k_{\max -1}\ . \tag{5}$$

These resulting linear systems can all be solved easily, since the highest order temporal derivatives in u are only accompanied by the first order derivative of Φ w.r.t. u. Due to the well-posedness of the coupling conditions (3) this linear system always has a unique solution. Finally the Kowalewski-Cauchy procedure can be applied, since no sonic points occur, i.e. $f'_{e_i}(u_{e_i}) \neq 0$.

4. **Numerical examples.** As numerical examples we present a simple network of advection or Burgers equations. In both cases we used a network of three edges and two nodes, as depicted in Figure 3.

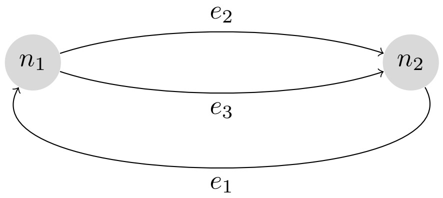


FIGURE 3. A network with one splitting and one merging node.

4.1. **Advection equation.** Consider for each edge of the network the conservation law (1) with the flux function $f(u) = u$. At the nodes we choose the coupling conditions (4), with $\alpha_2 = \frac{2}{3}$ and $\alpha_3 = \frac{1}{3}$ at n_1. For n_2 no parameter has to be defined, since the one equation needed is given by the conservation of mass.
As initial data we take the following functions on the edges of the same length $L_{e_1} = L_{e_2} = L_{e_3} = 1$:

$$u_{e_1}(0,x) = \sin(2\pi x) + 3 \ , \qquad u_{e_2}(0,x) = \frac{2}{3}u^0_{e_1}(x) \ , \qquad u_{e_3}(0,x) = \frac{1}{3}u^0_{e_1}(x) \ .$$

They are chosen such that the coupling conditions (2) are matched, as well as all its temporal derivatives (5) at the initial time. Since all edges transport with the same speed and the frequencies coincide, the data remains smooth over time, even across the coupling at the nodes. It is important to notice that we intentionally used a CFL number less than one, since for a CFL number equal to one the scheme yields the exact solution.

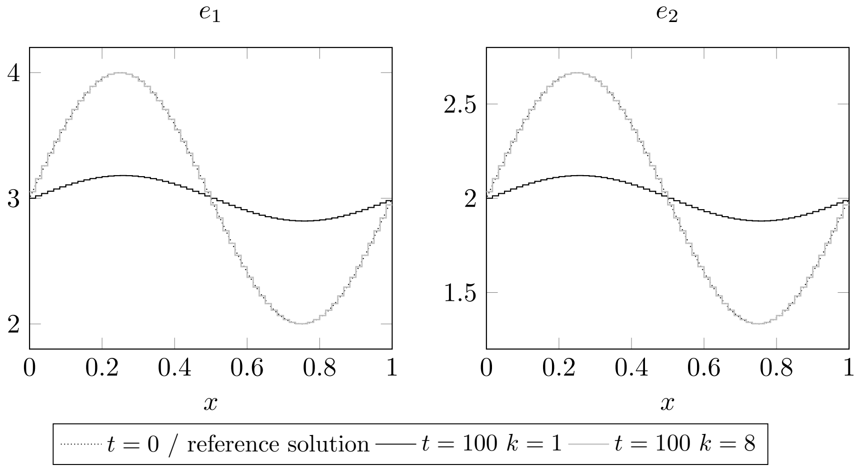


FIGURE 4. The solutions in e_1 and e_2 of an advection network at $t = 100$, computed with 60 cells per edge.

In Figure 4 we show the solution at time $t = 100$ on the edges e_1 and e_2. At this time the sine wave has passed 50 times the network. The solution of the scheme of order $k = 8$ still matches perfectly the exact one, while the first order version suffers remarkable diffusion. In Table 1 we show the errors and the corresponding rates of convergence. We omit the data for finer grids and higher order methods, since in both cases the errors reach the bounds of the machine imprecision.

	$k=4$				$k=5$			
N	L_∞	order	L_1	order	L_∞	order	L_1	order
60	2.34e-07		1.51e-07		5.29e-09		3.37e-09	
120	1.46e-08	4.00	9.33e-09	4.02	1.65e-10	5.00	1.05e-10	5.00
240	9.16e-10	4.00	5.83e-10	4.00	5.22e-12	4.99	3.30e-12	4.99

TABLE 1. Errors and rates of convergence on the advection network.

4.2. **Burgers equation with smooth data.** In this example we present the behavior of the scheme for a network of Burgers equations, i.e. we choose $f(u) = u^2$ for (1). In (4) we choose $\alpha_2 = \frac{16}{25}$ and $\alpha_3 = \frac{9}{25}$. The initial data is set according to the following functions

$$u_{e_1}(0,x) = \frac{1}{10}\sin(2\pi x) + \frac{1}{2}\ , \qquad u_{e_2}(0,x) = \frac{2}{25}\sin\left(\frac{5}{4}2\pi x\right) + \frac{2}{5}\ ,$$

$$u_{e_3}(0,x) = \frac{3}{50}\sin\left(\frac{5}{3}2\pi x\right) + \frac{3}{10}\ .$$

Furthermore, the lengths of the edges are adjusted such that each contains exactly one period of the initial values, i.e. $L_{e_1} = 1$, $L_{e_2} = 4/5$, $L_{e_3} = 3/5$. All these values are chosen such that the characteristics are always positive and the coupling conditions, as well as the equations for all their derivatives (5), are met at the time $t = 0$.

In Figure 5 the initial data and the solution in e_1 and e_2 at time $t = 0.2$ are shown.

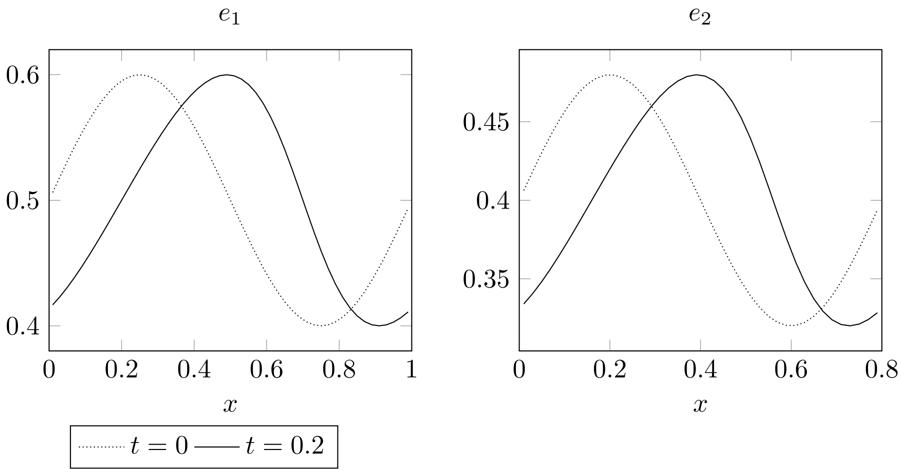


FIGURE 5. Burgers equation on e_1 and e_2, initial data and solution at $t = 0.2$

At this time the waves only have moved a short distance, but the top already starts forming into a shock wave. In Table 2 we show the errors and the rates of convergence for the schemes of order $k = 5$ and $k = 7$. It can be clearly seen that the desired order of convergence is present as long the error is not yet in the magnitude of the machine imprecision. Especially methods of even higher order converge so fast, that the imprecision of floating point operations prevents proper measurement of the rates of convergence.

	$k=5$				$k=7$			
N	L_∞	order	L_1	order	L_∞	order	L_1	order
50	3.19e-07		6.64e-08		3.58e-08		2.22e-09	
100	9.74e-09	5.03	1.91e-09	5.12	2.88e-10	6.96	1.89e-11	6.87
200	2.30e-10	5.40	5.52e-11	5.11	2.22e-12	7.02	1.37e-13	7.11
400	7.31e-12	4.98	1.66e-12	5.06	2.49e-14	6.48	5.15e-15	4.74
800	2.26e-13	5.02	5.06e-14	5.03	3.81e-14	-0.61	5.53e-15	-0.10

TABLE 2. Errors and convergence rates for a network of Burgers equation.

4.3. **Burgers equation with discontinuous data.** As final example we consider the network of Burgers equations with discontinuous data. As initial data a single block is placed in e_1, while the values in e_2 and e_3 are constant

$$u_{e_1}(0,x) = \begin{cases} 4 & x < 0.3 \\ 5 & 0.3 \leq x \leq 0.8 \\ 4 & 0.8 < x \end{cases}, \qquad u_{e_2}(0,x) = \frac{16}{5}, \qquad u_{e_3}(0,x) = \frac{12}{5}.$$

All edges are of the same length $L_e = 1,\ e = e_1, e_2, e_3$ and the coupling conditions are as in the example above. From the initial time $t = 0$ the block in e_1 moves to the right. While the front shock remains sharp, a rarefaction wave develops at the first jump. At about $t = 0.22$ the discontinuity reaches the node n_1 and splits up into two blocks. Both waves pass the junction and travel further in the edges e_2 and e_3. In Figure 6 we show the solution of a scheme of order $k = 5$ on a grid of 50 cells together with a reference solution. The numerical values are displayed as step function to emphasize the difference to the reference solution. We see that the shock has passed the node without oscillations and is resolved sharply in both edges. The not shown order of convergence is near to one due to the discontinuity in the data.

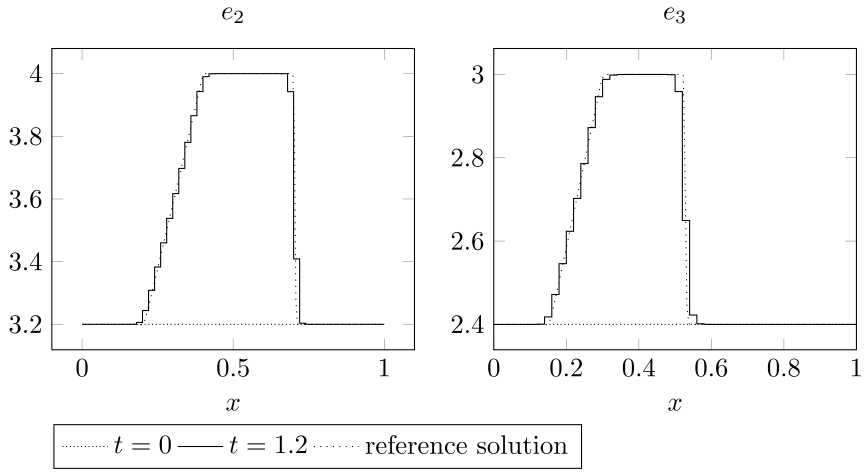


FIGURE 6. The solution at $t = 1.2$ in e_2 and e_3, of a scheme of order $k = 5$ with 50 cells.

5. Conclusion.

Based on the ideas of [11] we developed a ADER method for networks of scalar conservation laws. The numerical examples show a good agreement with the expected accuracy and robustness in case of discontinuous data. The additional computational costs at the nodes are small compared to the total effort of the scheme.

REFERENCES

[1] Raul Borsche, Rinaldo M. Colombo, and Mauro Garavello. On the coupling of systems of hyperbolic conservation laws with ordinary differential equations. *Nonlinearity*, 23(11):2749–2770, 2010.

[2] G. Bretti, R. Natalini, and B. Piccoli. Numerical approximations of a traffic flow model on networks. *Netw. Heterog. Media*, 1(1):57–84, 2006.

[3] R. M. Colombo, G. Guerra, M. Herty, and V. Schleper. Optimal control in networks of pipes and canals. *SIAM J. Control Optim.*, 48(3):2032–2050, 2009.

[4] Rinaldo M. Colombo and Graziano Guerra. On general balance laws with boundary. *J. Differential Equations*, 248(5):1017–1043, 2010.

[5] Ciro D'apice, Rosanna Manzo, and Benedetto Piccoli. Packet flow on telecommunication networks. *SIAM J. Math. Anal.*, 38(3):717–740, 2006.

[6] Miguel Ángel Fernández, Vuk Milišić, and Alfio Quarteroni. Analysis of a geometrical multiscale blood flow model based on the coupling of ODEs and hyperbolic PDEs. *Multiscale Model. Simul.*, 4(1):215–236 (electronic), 2005.

[7] Mauro Garavello and Benedetto Piccoli. Conservation laws on complex networks. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 26(5):1925–1951, 2009.

[8] S. Göttlich, M. Herty, and A. Klar. Network models for supply chains. *Commun. Math. Sci.*, 3(4):545–559, 2005.

[9] M. Herty, C. Kirchner, and A. Klar. Instantaneous control for traffic flow. *Math. Methods Appl. Sci.*, 30(2):153–169, 2007.

[10] Guang-Shan Jiang and Chi-Wang Shu. Efficient implementation of weighted ENO schemes. *J. Comput. Phys.*, 126(1):202–228, 1996.

[11] Sirui Tan and Chi-Wang Shu. Inverse Lax-Wendroff procedure for numerical boundary conditions of conservation laws. *J. Comput. Phys.*, 229(21):8144–8166, 2010.

[12] Eleuterio F. Toro. *Riemann solvers and numerical methods for fluid dynamics*. Springer-Verlag, Berlin, third edition, 2009. A practical introduction.

E-mail address: `borsche@mathematik.uni-kl.de`
E-mail address: `kall@mathematik.uni-kl.de`

ASYMPTOTIC BEHAVIOR OF A DIFFUSIVE SCHEME SOLVING THE INVISCID ONE-DIMENSIONAL PRESSURELESS GASES SYSTEM

Laurent Boudin

UPMC Univ Paris 06, UMR 7598 LJLL
F-75005 Paris, France
and
Inria Paris-Rocquencourt, F-78150 Le Chesnay Cedex, France

Julien Mathiaud

CEA, DAM, Cesta
BP2, F-33114 Le Barp, France
and
CMLA, ENS Cachan, CNRS, UniverSud, 61 Avenue du Président Wilson
F-94230 Cachan, France

Abstract. In this work, we discuss some numerical properties of the viscous numerical scheme introduced in [8] to solve the one-dimensional pressureless gases system, and study in particular, from a computational viewpoint, its asymptotic behavior when the viscosity parameter $\varepsilon > 0$ used in the scheme becomes smaller.

1. **Introduction.** In this work, we focus on the one-dimensional system describing a pressureless gas, which writes, for any $T > 0$,

$$\partial_t \rho + \partial_x(\rho u) = 0, \tag{1}$$

$$\partial_t q + \partial_x(q u) = 0, \tag{2}$$

in $(0,T] \times \mathbb{R}$, where $\rho(t,x) \geq 0$ is the gas density and $q(t,x) \in \mathbb{R}$ is the momentum at time $t \in [0,T]$ and location $x \in \mathbb{R}$. The gas velocity $u(t,x) \in \mathbb{R}$ must be somehow defined as a quotient of q by ρ, but there is trouble when ρ is 0. That is why one needs the notion of duality solutions [5], previously introduced for conservation laws [4]. Clearly, (1) and (2) can be seen as conservation laws, for mass and momentum respectively. System (1)–(2) is supplemented with initial conditions

$$\rho(0,\cdot) = \rho^{\text{in}}, \qquad q(0,\cdot) = q^{\text{in}}. \tag{3}$$

This system arises from very various physical situations (cold plasmas [3], astrophysics [17, 9], traffic models [1, 13]...) and has been mathematically studied in numerous articles, such as [14, 10, 16, 5].

We here choose a periodic framework: we focus on $[0,1]$ and impose that mass density, velocity and momentum have the same values at both $x = 0$ and $x = 1$, so that the solutions are 1-periodic.

2000 *Mathematics Subject Classification.* Primary: 65M12, 58F17; Secondary: 53C35.

Key words and phrases. Pressureless gases, vanishing viscosity, diffusive scheme, regularity loss, asymptotic behavior.

When ρ and u are smooth, and if ρ remains non zero, (2) can be modified into the standard Burgers equation thank to (1):

$$\partial_t u + \partial_x \left(\frac{u^2}{2} \right) = \partial_t u + u \partial_x u = 0. \tag{4}$$

The mass density ρ then solves a basic transport equation, where u is given: it does not depend on ρ because of (4). But it is common knowledge that, in finite time, a mass concentration phenomenon can happen, for instance, when u does not increase. Consequently, the regularity properties of u and ρ are lost, and u does not solve (4) anymore.

From the numerical viewpoint, one can think about several ways to discretize (1)–(3). Kinetic schemes [3, 6] or particle methods [12] allow to use the kinetic framework underlying the pressureless gas dynamics. It is also natural to try numerical schemes related to hyperbolic conservation laws [15] or relaxation schemes [2].

In [8], we prove that upwind schemes were not an option, since they failed to ensure the discrete one-sided Lipschitz (OSL) condition introduced [11] for convex scalar conservation laws. Then, following the strategy of [7] at the continuous level, we add an artificial viscosity. The new diffusive scheme we obtain is proven to be L^∞-stable and consistent, hence converging towards the solution of the viscous pressureless gases system. In particular, it satisfies the discrete OSL condition. We here investigate, at the numerical level, the asymptotic behavior of the same numerical scheme when the artificial viscosity vanishes.

2. Diffusive numerical scheme.

Let us recall the diffusive scheme we presented in [8]. Consider Δt, $\Delta x > 0$ such that $N = T/\Delta t \in \mathbb{N}$ and $I = 1/\Delta x \in \mathbb{N}$, and set $\lambda = \Delta t/\Delta x$. Denote ρ_i^n, q_i^n and u_i^n the approximate values of ρ, q and u at time $n\Delta t \in [0,T]$ and location $(i+1/2)\Delta x \in [0,1]$, for $0 \le n \le N$ and $0 \le i < I$. Of course, thanks to the periodicity property, ρ_i^n, q_i^n and u_i^n can be extended for any $i \in \mathbb{Z}$.

For the sake of readability, in the previous notations, we may drop the time iteration index n and replace $n+1$ by a prime symbol “ $'$ ”.

Note that the discrete OSL condition can be written as $n\lambda(u_{i+1}^n - u_i^n) \le 1$, for any i and $n > 0$.

Let us describe step by step the strategy for our scheme.

2.1. Periodic initial data.

We choose arbitrary 1-periodic initial data $\rho^{\text{in}} \ge 0$, $u^{\text{in}} \ge 0$. Indeed, the viscous problem in [7] deals with mass density and velocity, and not momentum.

2.2. Regularizing initial data.

We take a fixed $\varepsilon > 0$, small enough. We may have to regularize both u^{in} and ρ^{in} so that u^{in}, $\rho^{\text{in}} \in C^1(\mathbb{R}; \mathbb{R}_+^*)$ satisfy the assumptions of Theorem 2 in [8], i.e.

$$\rho^{\text{in}}(x) \ge C\varepsilon^{1/4}, \qquad u^{\text{in}}(x) \le C, \qquad (u^{\text{in}})'(x) \le \frac{C}{\sqrt{\varepsilon}}, \qquad \forall x \in [0,1], \tag{5}$$

where $C \ge 0$ is a constant not depending on ε. Note that ρ^{in} must lie in $\mathbb{R}_+^*$, since the continuous viscous model involves a division by ρ.

The following quantities

$$U = \max_{[0,1]} u^{\text{in}} > 0, \qquad V = \min_{[0,1]} u^{\text{in}} > 0,$$
$$A = \max(0, \max_{[0,1]} (u^{\text{in}})') \geq 0, \qquad R = \min_{[0,1]} \rho^{\text{in}} > 0,$$

may depend on ε. More precisely, they must satisfy properties inherited from (5), i.e.

$$R \geq C\varepsilon^{1/4}, \qquad V \leq U \leq C, \qquad A \leq \frac{C}{\sqrt{\varepsilon}}, \tag{6}$$

where $C \geq 0$ does not depend on ε.

2.3. **Choosing the time and space steps.** The steps Δt and $\Delta x > 0$ are then chosen such that

$$0 < \quad \Delta x \quad \leq \frac{2V}{1+A}, \tag{7}$$
$$0 < \quad \Delta t \quad \leq \min\left(\frac{1}{4A+1}, \frac{1}{4U}\Delta x, \frac{R}{4\varepsilon(1+AT)}\Delta x^2\right), \tag{8}$$

where we set $\lambda = \Delta t/\Delta x$ and $\sigma = \Delta t/\Delta x^2$.

2.4. **Writing the scheme.** We eventually write the discretization of the viscous pressureless gases system as

$$u_i' = u_i - \lambda\left(\frac{{u_i}^2}{2} - \frac{{u_{i-1}}^2}{2}\right) + \frac{\varepsilon\sigma}{\rho_i}(u_{i-1} + u_{i+1} - 2u_i), \tag{9}$$
$$\rho_i' = (1 - \lambda u_i')\rho_i + \lambda u_{i-1}'\rho_{i-1}. \tag{10}$$

2.5. **Properties of the scheme.** In [8], we prove the following

Theorem 2.1. *We assume that (7)–(8) hold. Then we have, for any i and $n \geq 0$,*

$$V \leq u_i^n \leq U, \qquad u_i^n - u_{i-1}^n \leq \frac{A\Delta x}{1 + An\Delta t}, \qquad \rho_i^n \geq \frac{R}{1 + An\Delta t} \geq \frac{R}{1+AT} > 0.$$

Moreover, the discrete total mass is conserved. Finally, when $\varepsilon > 0$ is fixed, scheme (9)–(10) is first-order consistent (in both t and x) with the viscous pressureless gases system, and is monotonic.

In other words, both discrete unknown functions satisfy maximum principles, and the discrete velocity satisfies the OSL condition. Assumptions (8) on Δt are crucial to ensure the stability of the scheme. They are related to the standard assumptions to get stability for explicit schemes on transport or diffusion equations.

3. **Numerical study.** The viscosity parameter $\varepsilon > 0$ is chosen in the first place. We use several values of ε in the following computations: 10^{-4}, 10^{-5}, 10^{-6}, 5×10^{-7}, 10^{-7} and the smallest value $\varepsilon_0 = 10^{-6}/14$, which will be considered as the reference situation. We take $T = 1$ s as the final time. Moreover, we focus on one case with (almost, see the discussion below) smooth initial data. It allows to point out all the problems arising the pressureless gases system, in particular the regularity loss phenomenon, already explained in Section 1.

We set, for any $x \in [0,1]$,

$$\rho^{\text{in}}(x) = 1 + \frac{1}{2}\cos(4\pi x), \qquad u^{\text{in}}(x) = 2x(1-x) + \frac{1}{2}.$$

We clearly have $U = 1$, $V = 1/2$, $A = 2$ and $R = 1/2$. When ε belongs to the set of values above, (6) is clearly satisfied. Nevertheless, u^{in} is not $C^1(\mathbb{R};\mathbb{R}_+^*)$. Indeed, there is an issue at the bounds of $[0,1]$. Therefore, we should need to regularize u^{in}. Fortunately, it is not necessary: the regularization near 0 and 1 can happen on intervals whose length may be chosen smaller than $2\Delta x$. We can choose it so that $(u^{\text{in}})'$ remains in $[-2,2]$, and consequently, A still equals 2.

Let us now choose $\Delta x = 20\varepsilon$. Again, if ε belongs to our same set of values, (7) obviously holds. Then Δt can be taken as the optimal possible value given by (8), i.e. $\Delta t = \varepsilon \min(5, 50/(1+2T)) = 5\varepsilon$.

We want to study the behavior of our numerical solution with respect to ε. In the following, since ϱ has a measure meaning when ε goes to zero, it is really more convenient to focus on the cumulative mass M, defined by

$$M(t,x) = \int_0^x \varrho(t,y)\,\mathrm{d}y.$$

We first draw the evolution of M and u with respect to x, at two different times, before and after approximately $t = 0.5$. Indeed, until $t = 0.5$, at the continuous level, u solves the Burgers equation, and there is no regularity loss. But, near time $t = 0.5$, a regularity loss happens because of mass concentration and subsequent vacuum appearance.

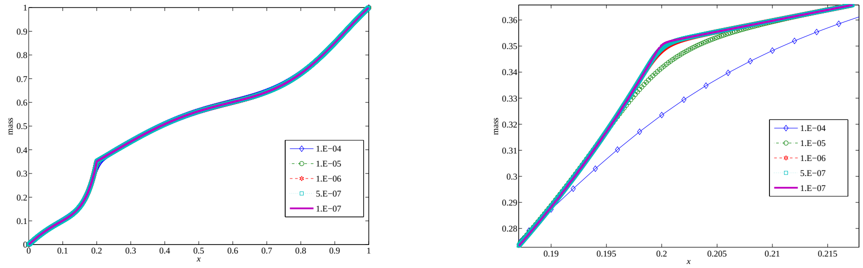


FIGURE 1. Behavior, at time 0.4, of M (a) on $[0,1]$, (b) focused around $x = 0.2$

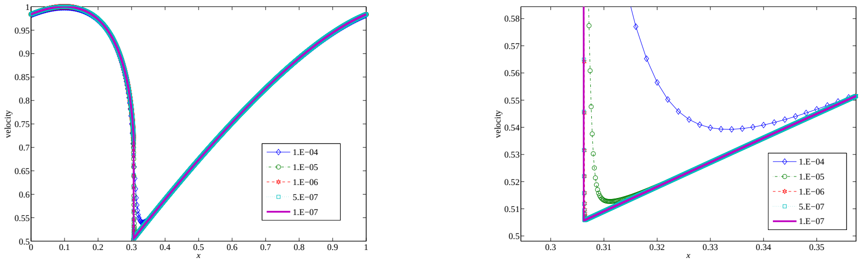


FIGURE 2. Behavior, at time 0.4, of u (a) on $[0,1]$, (b) focused around $x = 0.2$

The solutions are then displayed for various values of ε. More precisely, Fig. 1–2 respectively give the behavior of M and u at time $t = 0.4$, i.e. before the regularity loss. Of course, the curves are sharpened with smaller values of ε, but there is still no jump at all for either ρ or u.

Fig. 3–4 then present the typical profiles of both M and u after the regularity loss, here given at final time $T = 1$. The jump is now visible at $x = 0.59$ on both figures, and u of course still decreases at the jump abscissa.

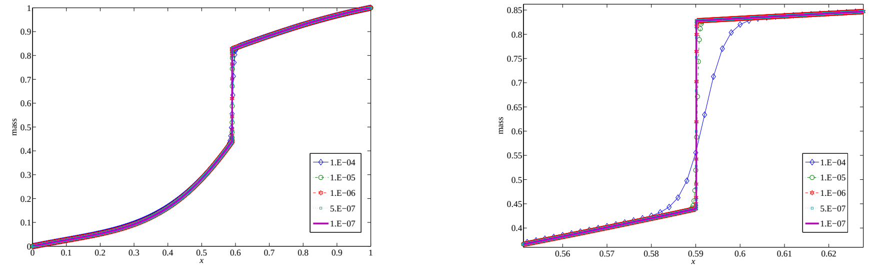


Figure 3. Behavior, at final time, of M (a) on $[0, 1]$, (b) focused around $x = 0.59$

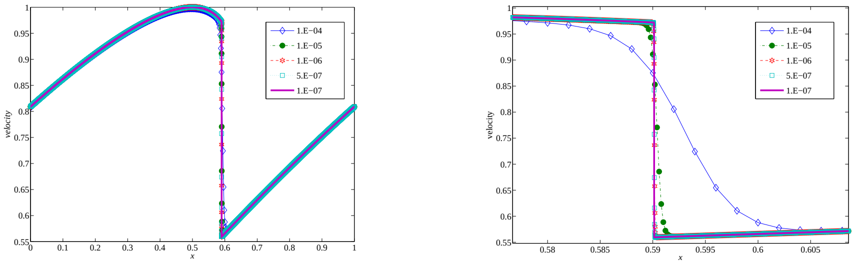


Figure 4. Behavior, at final time, of u (a) on $[0, 1]$, (b) focused around $x = 0.59$

Let us now study the evolution of the total momentum

$$Q(t) = \int_0^1 \rho(t, x)\, u(t, x)\, \mathrm{d}x$$

with respect to time.

As the reader can see on Fig. 5, till the regularity loss, the numerical conservation of Q is quite satisfactorily ensured. The property is not recovered after the loss. This drawback of our scheme (9)–(10) was already pointed out in [8]: we had to choose between the total momentum conservation (if we write a scheme for ρ and q) and the OSL condition (for which we wrote a scheme for ρ and u). The latter is in fact crucial to ensure that we select the right solution to the pressureless gases system. Nevertheless, we can at least state that the order of magnitude of Q is conserved.

Eventually, we focus on the behavior with respect to ε. We focus on the following quantities

$$\mathcal{E}_M(t) = \max_{x\in[0,1]} |M(t, x) - M_0(t, x)|, \tag{11}$$

$$\mathcal{E}_u(t) = \max_{x\in[0,1]} |u(t, x) - u_0(t, x)|, \tag{12}$$

for the values ε under study. Quantities indexed by 0 are of course the ones related to the reference value of ε, ε_0. Quantities on coarser grids are projected on the

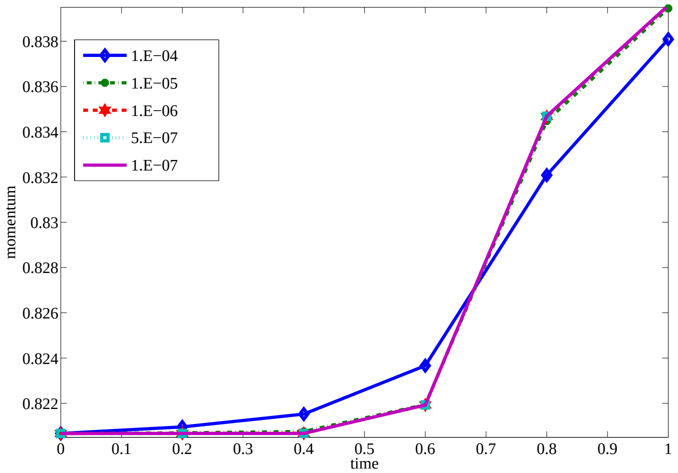


FIGURE 5. Checking the momentum for several values of ε

(fine) grid of the reference value (ε_0 corresponds to 700,000 space cells) in order to compute the discretized versions of (11)–(12).

Fig. 6–7 respectively show the behavior of $\mathcal{E}_M$ and $\mathcal{E}_u$ with respect to time.

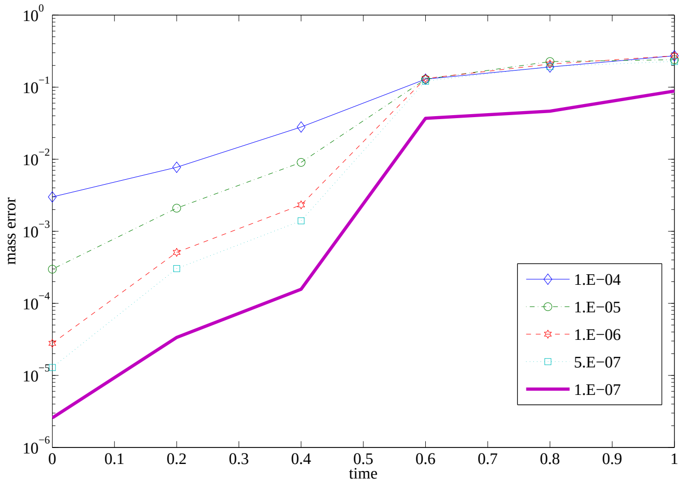


FIGURE 6. Behavior of $\mathcal{E}_M$ w.r.t. t for several values of ε

These results give us hints of what should be the asymptotic behavior of our solutions, in particular when we loose regularity. The convergence seems very slow in ε but it was expected since the smallest power of ε in the main hypotheses of the scheme is equal to one fourth, see Eqn. 6. That partially explains that the errors are quite large, even for quite small values of ε. In fact, we were not able to understand why $\varepsilon = 10^{-7}$ gives such a different behavior (in order of magnitude) compared to not so larger values of ε, whereas, from 10^{-4} to 10^{-6}, the results are close after the regularity loss.

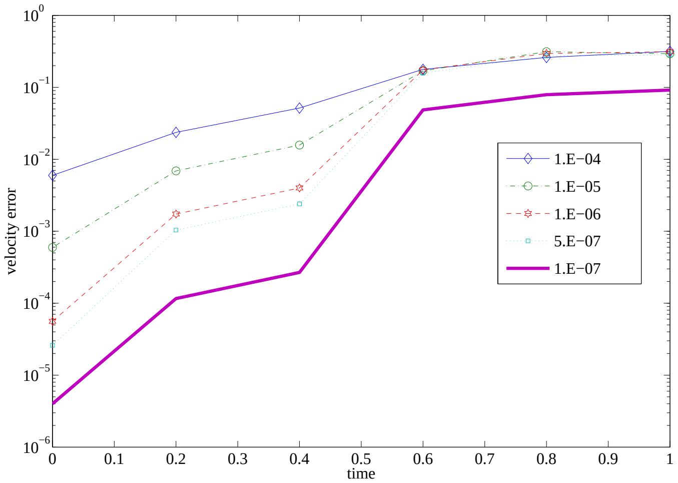


FIGURE 7. Behavior of $\mathcal{E}_u$ w.r.t. t for several values of ε

4. **Perspectives.**

As expected, the crucial point in the asymptotic behavior of our scheme is strongly linked to the regularity loss phenomenon which occurs for the pressureless gases. We should be able to prove reasonable convergence estimates with respect to ε as long as the solutions remain smooth. Beyond the regularity loss, the situation remains quite unclear.

Acknowledgments. The authors want to thank Frédéric Lagoutière for the (ongoing) discussions they share about this topic.

REFERENCES

[1] F. Berthelin, P. Degond, M. Delitala, and M. Rascle. A model for the formation and evolution of traffic jams. *Arch. Ration. Mech. Anal.*, 187(2):185–220, 2008.

[2] C. Berthon, M. Breuss, and M.-O. Titeux. A relaxation scheme for the approximation of the pressureless Euler equations. *Numer. Methods Partial Differential Equations*, 22(2):484–505, 2006.

[3] F. Bouchut. On zero pressure gas dynamics. In *Advances in kinetic theory and computing*, volume 22 of *Ser. Adv. Math. Appl. Sci.*, pages 171–190. World Sci. Publ., River Edge, NJ, 1994.

[4] F. Bouchut and F. James. One-dimensional transport equations with discontinuous coefficients. *Nonlinear Anal.*, 32(7):891–933, 1998.

[5] F. Bouchut and F. James. Duality solutions for pressureless gases, monotone scalar conservation laws, and uniqueness. *Comm. Partial Differential Equations*, 24(11-12):2173–2189, 1999.

[6] F. Bouchut, S. Jin, and X. Li. Numerical approximations of pressureless and isothermal gas dynamics. *SIAM J. Numer. Anal.*, 41(1):135–158 (electronic), 2003.

[7] L. Boudin. A solution with bounded expansion rate to the model of viscous pressureless gases. *SIAM J. Math. Anal.*, 32(1):172–193 (electronic), 2000.

[8] L. Boudin and J. Mathiaud. A numerical scheme for the one-dimensional pressureless gases system. *Numer. Methods Partial Differential Equations*, 28(6):1729–1746, 2012.

[9] Y. Brenier. A modified least action principle allowing mass concentrations for the early universe reconstruction problem. *Confluentes Math.*, 3(3):361–385, 2011.

[10] Y. Brenier and E. Grenier. Sticky particles and scalar conservation laws. *SIAM J. Numer. Anal.*, 35(6):2317–2328 (electronic), 1998.

[11] Y. Brenier and S. Osher. The discrete one-sided Lipschitz condition for convex scalar conservation laws. *SIAM J. Numer. Anal.*, 25(1):8–23, 1988.

[12] A. Chertock, A. Kurganov, and Yu. Rykov. A new sticky particle method for pressureless gas dynamics. *SIAM J. Numer. Anal.*, 45(6):2408—2441 (electronic), 2007.

[13] F. De Vuyst, V. Ricci, and F. Salvarani. Nonlocal second order vehicular traffic flow models and Lagrange-remap finite volumes. In *Finite volumes for complex applications. VI. Problems & perspectives. Volume 1, 2*, volume 4 of *Springer Proc. Math.*, pages 781–789. Springer, Heidelberg, 2011.

[14] W. E, Yu. G. Rykov, and Ya. G. Sinai. Generalized variational principles, global weak solutions and behavior with random initial data for systems of conservation laws arising in adhesion particle dynamics. *Comm. Math. Phys.*, 177(2):349–380, 1996.

[15] L. Gosse and F. James. Numerical approximations of one-dimensional linear conservation equations with discontinuous coefficients. *Math. Comp.*, 69(231):987–1015, 2000.

[16] F. Poupaud and M. Rascle. Measure solutions to the linear multi-dimensional transport equation with non-smooth coefficients. *Comm. Partial Differential Equations*, 22(1-2):337–358, 1997.

[17] Ya.B. Zel'dovich. Gravitational instability: An approximate theory for large density perturbations. *Astron. and Astrophys.*, 5:84–89, 1970.

E-mail address: `laurent.boudin@upmc.fr`
E-mail address: `julien.mathiaud@cea.fr`

COUPLING TECHNIQUES FOR NONLINEAR HYPERBOLIC EQUATIONS

Benjamin Boutin
IRMAR
Université de Rennes 1
35042 Rennes, France

Frédéric Coquel
Centre de Mathématiques Appliquées
& Centre National de la Recherche Scientifique
École Polytechnique
91128 Palaiseau, France

Philippe G. LeFloch
Laboratoire Jacques-Louis Lions
& Centre National de la Recherche Scientifique
Université Pierre et Marie Curie, Paris 6
75252 Paris, France

Abstract. We analyze the coupling between different nonlinear hyperbolic equations across possibly resonant interfaces. The proposed reformulation of the problem involves a nonconservative product that is understood through a self-similar viscous approximation. We obtain the existence of a coupled solution to the Riemann problem in this thin interface regime, and underline the persisting multiplicity of solutions for some Riemann data, even in simple situations. Another regularization strategy is then studied, that corresponds somehow to a thick interface regime. This other selection criterion leads to a well-posed problem and we thus consider a finite volume scheme to approximate its solution.

1. **Introduction.**

1.1. **Context of this work.**

The main aim of this work consists in the improvement of the numerical simulation of multi-scale phenomena in the nuclear technology. Namely, in such a context one has typically to couple different existing codes for thermalhydraulic flows, each one of them computing a different and specific part of the flow in the powerplant installation. There is however a strong practical constraint in play: the coupling procedure has to be as non-invasive as possible, using for example the data for the physical variables in only one cell on each side of the inferface, or making the usefull information passing through a buffer zone. The natural question is then: how to design such a coupling procedure adapted to nonlinear hyperbolic problems ? More importantly: how much is this coupling procedure mathematically and physically satisfactory?

2000 *Mathematics Subject Classification.* 35L65, 76L05, 76N.

Key words and phrases. Hyperbolic conservation law, coupling problem, thick interface model, Riemann problem, resonant effect.

These questions have made the object of a large litterature in the past decade and connects strongly with the problematic of solving discontinuous conservation laws. We refer the reader for example to the works of Audusse and Perthame [2], Seguin and Vovelle [16], Andreianov, Karlsen and Risebro[1], or Panov [15]. However it is important to notice that our point of view is here slightly different. We are interested in the nonconservative coupling. This framework also applies to consider active control devices in which mass may be injected at a well chosen location to prevent from slugging effects in the flow, therefore with a singular (localized) nonconservative term.

1.2. **The state-coupling problem.** Let consider the coupling between different nonlinear hyperbolic equations across a fixed interface, say at $x = 0$ considering for convenience only the one space variable problem, with unknown $w(x,t) \in \mathbb{R}^d$:

$$\partial_t w + \partial_x f^{\pm}(w) = 0, \quad t > 0, \ \pm x > 0. \tag{1}$$

The flux functions f^+ and f^- are supposed to be different and a supplemented coupling condition, modeling the transient exchange of informations at the interface, reads as the continuity of the unknown w or of a nonlinear transformation of it $u(x,t) = \theta_{\pm}(w(x,t))$, $\pm x > 0$, say

$$u(0^-,t) = u(0^+,t), \quad t > 0. \tag{2}$$

In the hyperbolic framework, it is well-known that such a trace condition has to be formulated in a weak form, for example following the theory for boundary conditions in hyperbolic problems developed by Dubois and LeFloch [12] and used more recently in the context of the coupling by Godlewski and Raviart [13]. However, multiple solutions to a given Riemann problem may then occur. We underline that due to the expected continuity condition (2), the Rankine-Hugoniot relation at $x = 0$ only provides another global formulation of (1) with a measure source term:

$$\partial_t w + \partial_x f(x,w) = (f^+(w(0^+,t)) - f^-(w(0^-,t)))\delta_{x=0}, \quad t > 0, \ x \in \mathbb{R}, \tag{3}$$

where $f(x,w) := f^{\pm}(w)$ for $\pm x > 0$.

Using the set of admissible traces in half-Riemann problems, Chalons, Raviart and Seguin [10] have studied the coupling of two Euler systems with different equation of states. They obtain some uniqueness results under condition and prove in certain cases a large non-uniqueness. Another similar result is obtained in the case of general scalar flux functions in [9]. The difficulty arising with the thin interface coupling problem lies in the fact that the initial value problem, even with apparently well-defined interface conditions, is often ill-posed, so that this model does not fully determine the dynamic of the relevant solution. A supplemented selection criterion has to be added to recover uniqueness. As an example for such a selection criterion, one can think at this level of the theory of L^1-dissipative admissibility germs developed by Andreianov, Karlsen and Risebro in [1].

1.3. **Formulation of the extended system.** The keystone of the present work consists in considering the coupling interface as being a standing wave for an augmented system of partial differential equations

$$\partial_t u + A(u,v)\partial_x u = 0, \quad \partial_t v = 0. \tag{4}$$

For the sake of simplicity, we restrict here the situation to the situation where $\theta_{\pm} = \mathrm{Id}$. The new variable $v \in \mathbb{R}$ plays the role of a color function and one sets for example $A(u,v) = v\nabla f^+(u) + (1-v)\nabla f^-(u)$, with $v(x) = 0$ for $x \le 0$ and $v(x) = 1$ for $x > 0$.

Obviously, when ∇f^+ and ∇f^- have eigenvalues with different signs, then the whole system (4) in (u,v) may fall only non-strictly hyperbolic: 0 is an eigenvalue with multiplicity more than two. Think typically to the scalar situation where one has $A(u,v) = v\lambda_+(u) + (1-v)\lambda_-(u)$ which may vanish for some $v \in (0,1)$ as soon as $\lambda_+(u)\lambda_-(u) < 0$.

The system (4) inherits therefore the well-posedness difficulty of the original coupling problem, through its non-strict hyperbolicity and its nonconservativeness.

1.4. **Outline of the paper.** The present paper organizes as follows. In a first part we consider the so-said thin interface regime. Considering vanishing viscosity self-similar solutions, we get an existence result and illustrate the remaining non-uniqueness for some simple scalar examples. Then, in a second part, we explore a thick interface regime derived from the previous extended PDE system and prospect the above mentionned non-uniqueness through some numerical experiments.

2. The thin interface regime.

2.1. **The self-similar viscous approximation.** The definition of the weak solutions for the nonconservative system (4) in the resonant regime is tackled turning back to the self-similar vanishing viscosity analysis of Dafermos [11]. This methodology was successfully introduced by Tzavaras [17] and by Joseph and LeFloch [14] to get existence results in hyperbolic systems in nonconservative form. The first results of existence for such solutions in the coupling framework have been obtained in the scalar case by Boutin, Coquel and Godlewski [4]. In [5], we extend this analysis to the coupling of hyperbolic systems. We consider the solution u^ϵ of

$$(-\xi \mathrm{Id} + A(u^\epsilon, v^\epsilon))u^\epsilon_\xi = \epsilon u^\epsilon_{\xi\xi}, \quad -\xi v^\epsilon_\xi = \epsilon^2 v^\epsilon_{\xi\xi}, \tag{5}$$

together with the boundary conditions:

$$\lim_{\xi\to-\infty} u^\epsilon = u_\ell, \quad \lim_{\xi\to+\infty} u^\epsilon = u_r. \tag{6}$$

2.2. **An existence result.** Under fairly general assumptions we obtain the existence of self-similar weak solutions to the Riemann problem for (4), as limit of (a subsequence of) u^ϵ as ϵ goes to zero.

Theorem 2.1 ([5]). *There exists a solution u^ϵ of (5)-(6) that converges pointwise to $u \in BV$ satisfying in the sense of distributions*

$$-\xi u_\xi + (f_\pm(u))_\xi = 0, \quad \pm\xi > 0, \tag{7}$$

and the entropy inequalities (with $\eta_\pm$ two convex functions and $q_\pm$ the associated entropy fluxes)

$$-\xi(\eta_\pm(u))_\xi + (q_\pm(u))_\xi \le 0, \quad \pm\xi > 0. \tag{8}$$

The proof of this result is based on a fixed point argument and requires some smallness estimate on the interaction coefficients involving the different waves of the hyperbolic system (including the stationnary wave emanating from the supplemented variable v). These smallness estimates follow directly from the strictly hyperbolic character of the whole system in the variable u and from the closeness of the fluxes f^+ and f^- and of the Riemann data u_ℓ and u_r. Moreover the interaction coefficient involving the interfacial wave and the possibly resonant wave is under control. In the easier scalar setting, the same result extends without any closeness assumption on the Riemann data, nor on the flux functions.

2.3. **The analysis of the internal structure of the layer.** In order to understand if the diffusive approach achieves a selection of solutions in the resonant situations, we investigate the internal structure of the coupling interface [6]. To that aim, we consider a usual change of variable in the analysis of shock profiles and set $\mathcal{U}^\epsilon(y) = u^\epsilon(\epsilon y)$ and $\mathcal{V}^\epsilon(y) = v^\epsilon(\epsilon y)$ for $y \in \mathbb{R}$.

Hereafter follow some property of the limiting profile in the scalar setting:

- $\mathcal{U}$ is a monotone bounded smooth function ($\mathcal{U} \in \mathcal{C}^2(\mathbb{R})$), with the same monotonicity as u (given by the sign of $u_r - u_\ell$),
- the limit of $\mathcal{U}$ as y tends to ∞ (denoted $\mathcal{U}_\infty$), and as y tends to $-\infty$ (denoted $\mathcal{U}_{-\infty}$) satisfy to the following conditions: $f^-(u(0^-)) = f^-(\mathcal{U}_{-\infty})$, $q_-(u(0^-)) \geq q_-(\mathcal{U}_{-\infty})$ and $f^+(\mathcal{U}_{+\infty}) = f^+(u(0^+))$, $q_+(\mathcal{U}_{+\infty}) \geq q_+(u(0^+))$. In other words, there may be an interfacial layer corresponding to an entropy 0-shock for the left problem, or for the right problem, or for both,
- $\mathcal{U}$ solves the ODE: $A(\mathcal{U}, \mathcal{V})\mathcal{U}_y = \mathcal{U}_{yy}$, on $y \in \mathbb{R}$.

In Figures 1 and 2, we draw up a map of possible diffusive self-similar Riemann solutions for each given Riemann data (u_ℓ, u_r). To that purpose, we rule out the solutions for which all known necessar conditions are not fullfiled. The Figure 1 concerns the fluxes $f^-(u) = u^2/2$ and $f^+(u) = (u-c)^2/2$ with $c > 0$. In the Figure 2, we choose $c < 0$, reverting the natural order between the sonic points of these quadratic fluxes. As a consequence the resonance phenomenon reveals a large variety of non-uniqueness situations. The vanishing viscosity process however has selected at most a finite number of solutions for each Riemann data, as expected (a Laplace stability analysis provides a new necessar condition for solutions involving two waves and an intermediate constant state, see [6]).

3. The thick interface regime.

3.1. **A well-posed balance law.** In [7] and [8], we consider another regularization strategy based on thick interfaces. First of all, we reformulate the augmented system (4) to handle the case of a smoothed fixed color function v so that it reads as a conservation law with lipschitz source term (it should be understood as a regularization of the singular source term in (3)):

$$\partial_t w + \partial_x f(w,v) = \ell(w,v)\partial_x v, \quad \text{with } \ell(w,v) = \partial_v f(w,v). \tag{9}$$

The choice for the interface profile function v is thought here somehow arbitrary but it suppresses any difficulty due to the above mentionned resonant behavior. It is yet expected that different choices of v lead to different solutions w.

On the other side, we define a thick coupling condition to fully characterize the dynamic of solutions through the thick interface. The expected coupling condition (2), i.e. $\theta_-(w(0^-,t)) = \theta_+(w(0^+,t))$, is now ensured by requiring the PDE (and later the scheme) to preserve the u-constant states w where we set $w = C_0(u,v)$. Here $C_0(u,v)$ is a function connecting smoothly $\theta_-^{-1}(u)$ to $\theta_+^{-1}(u)$ as v goes from 0 to 1. It is supposed to realize the following monotonicity assumptions that $\partial_u C_0(u,v) > 0$ for any $u \in \mathbb{R}$ and any $v \in [0,1]$.

Theorem 3.1 (A Kružkov theorem). *As soon as $v \in W^{2,\infty}(\mathbb{R})$, and for any initial Cauchy data $w_0 \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$, there exists a unique $w \in L^\infty(\mathbb{R}_+, L^1(\mathbb{R}) \cap$*

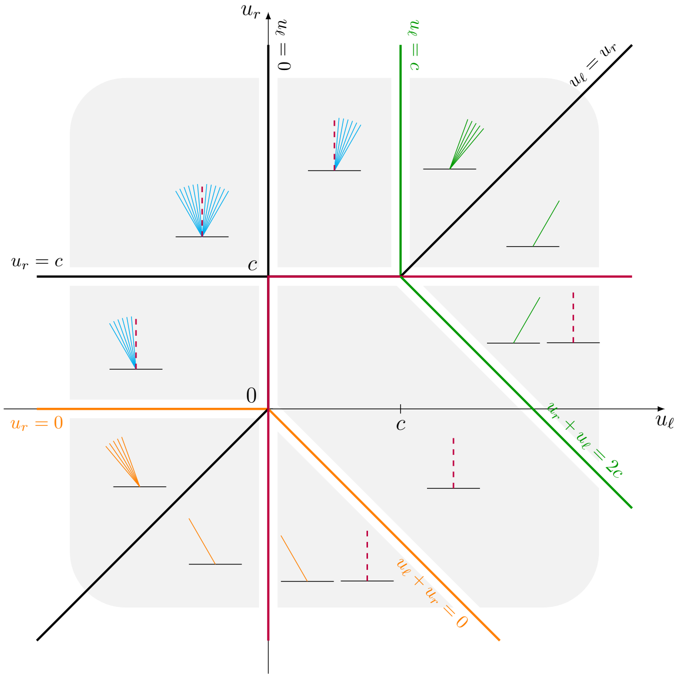


Figure 1. Diffusive self-similar Riemann solutions for the coupling of two quadratic fluxes ($c > 0$)

$L^\infty(\mathbb{R}))$, *solution of* (9) *satisfying the entropy inequalities*

$$\partial_t \mathfrak{U}(w) + \partial_x \mathfrak{F}(w, v) - \mathfrak{L}(w, v)\partial_x v \leq 0. \tag{10}$$

(See [7] *for the details of the notation)*

3.2. **Finite volume approximation and prospecting results.** A new well-balanced finite volume scheme is proposed in [7]. It approximates the entropy solution of (9) and preserves exactly the equilibria satisfying the thick coupling condition. This scheme is based on a non co-localized finite volume method with two distinct grids. An adapted reconstruction step ensures the required well-balanced property (see [3] for a review). The family of approximate solutions is proved to be uniformly bounded in sup-norm under a naturel CFL condition and the convergence of the numerical solution to an entropy measure-valued solution is obtained via infiniety many entropy inequalities and the use of DiPerna's theory. The numerical solution is shown to converge to the unique solution of Kružkov theorem.

Some numerical experiments are conducted in order to understand the sensitiveness of the selected solution with relation to the structure of the interface. We parametrize the choice of the interface profile with two parameters $\eta > 0$ and $\zeta \in \mathbb{R}$ and set $v(x) = \frac{1}{2}(\text{erf}(x/\eta + \zeta) + 1)$. The fluxes are the one corresponding to Figure 2 (with $c < 0$) where the resonant effects occur in the thin interface regime. In Figure 3 we represent three different selected interface profiles with various values for the parameter ζ and the same fixed small value for the parameter η. In Figures 4 and 5, the corresponding solutions w computed with 1000 points for the coupling problem with $\theta_\pm = \text{Id}$. In the first case, the solution consists in a double rarefaction fan with an intermediate constant state. We observe that the obtained intermediate

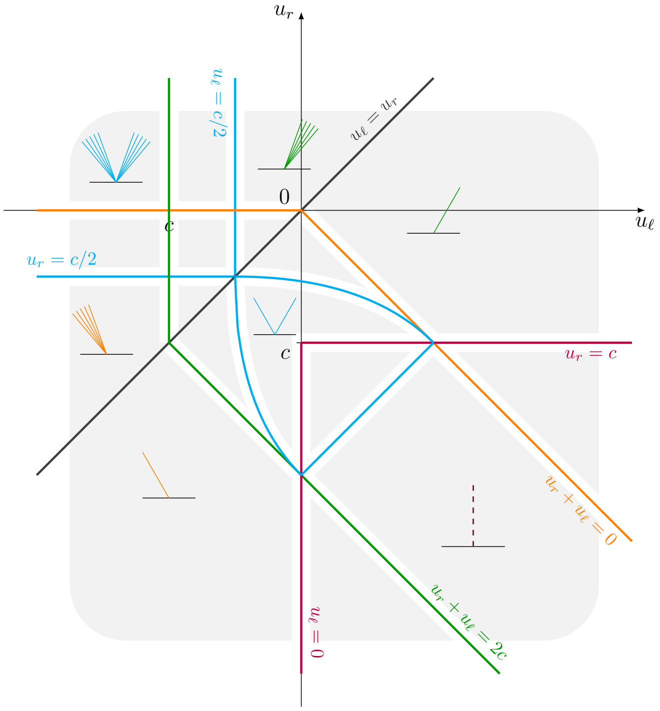


FIGURE 2. Diffusive self-similar Riemann solutions for the coupling of two quadratic fluxes ($c < 0$)

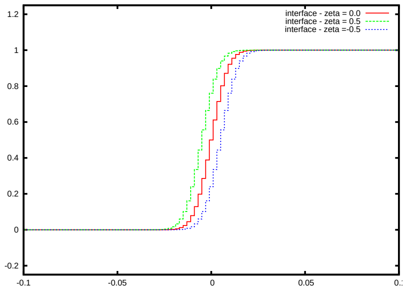


FIGURE 3. Three interface profiles

state depends on the interface profile. In the second case, the solution consist in a unique shock, either for the left problem, or for the right problem, or a standing discontinuity at $x = 0$. These numerical results illustrate the unstability of the solution with respect to the interface profile function in the resonant situations.

REFERENCES

[1] B. Andreianov, K.H. Karlsen, and N.H. Risebro, *A theory of L1-dissipative solvers for scalar conservation laws with discontinuous flux*, Arch. Ration. Mech. Anal. **201** (2011), 27-86.

[2] E. Audusse and B. Perthame, *Uniqueness for scalar conservation laws with discontinuous flux via adapted entropies*, Proc. Roy. Soc. Edinburgh Sect. A **135** (2005), 253-265.

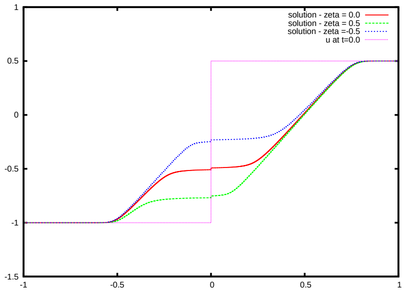


FIGURE 4. Corresponding solutions with $u_\ell = -1$, $u_r = 0.5$: three different double rarefaction fans

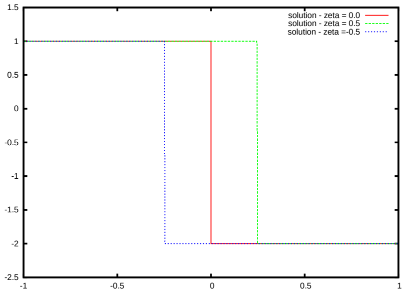


FIGURE 5. Corresponding solutions with $u_\ell = 1$, $u_r = -2$: three different shock solutions

[3] F. Bouchut, “Nonlinear Stability of Finite Volume Methods for Hyperbolic Conservation Laws and Well-balanced Schemes for Sources”, Frontiers in Mathematics, Birkhäuser Verlag, Basel, 2004.

[4] B. Boutin, F. Coquel, and E. Godlewski, *Dafermos' regularization for interface coupling of conservation laws*, in “Hyperbolic problems: Theory, Numerics, Applications” (eds. S. Benzoni-Gavage, Sylvie and D. Serre), Springer, (2008), 567–575.

[5] B. Boutin, F. Coquel, and P.G. LeFloch, *Coupling techniques for nonlinear hyperbolic equations. I. Self-similar diffusion for thin interfaces*, Proc. Roy. Soc. Edinburgh Sect. A, **141** (2011), 921–956.

[6] B. Boutin, F. Coquel, and P.G. LeFloch, *Coupling techniques for nonlinear hyperbolic equations. II. Resonant interfaces with internal structure*, (In preparation).

[7] B. Boutin, F. Coquel, and P.G. LeFloch, *Coupling techniques for nonlinear hyperbolic equations. III. The well-balanced approximation of thick interfaces*, SIAM J. Numer. Anal., **51** (2013), 1108–1133.

[8] B. Boutin, F. Coquel, and P.G. LeFloch, *Coupling techniques for nonlinear hyperbolic equations. IV. A multidimensional finite volume framework*, preprint, arXiv:1206.0248.

[9] B. Boutin, C. Chalons, and P.-A. Raviart, *Existence result for the coupling problem of two scalar conservation laws with Riemann initial data*, Math. Models Methods Appl. Sci. **20** (2010), 1859-1898.

[10] C. Chalons, P.-A. Raviart, and N. Seguin, *The interface coupling of the gas dynamics equations*, Quart. Appl. Math., **66** (2008), 659–705.

[11] C.M. Dafermos, *Solution of the Riemann problem for a class of hyperbolic systems of conservation laws by the viscosity method*, Arch. Rational Mech. Anal., **52** (1973), 1–9.

[12] F. Dubois, and P.G. LeFloch, *Boundary conditions for nonlinear hyperbolic systems of conservation laws*, J. Differential Equations, **71** (1988), 93–122.

[13] E. Godleswki, and P.-A. Raviart, *The numerical interface coupling of nonlinear hyperbolic systems of conservation laws. I. The scalar case*, Numer. Math., **97** (2004), 81–130.

[14] K.T. Joseph, and P.G. LeFloch, *Boundary layers in weak solutions of hyperbolic conservation laws. II. Self-similar vanishing diffusion limits*, Comm. Pure Appl. Anal. **1** (2002), 51-76.

[15] E. Y. Panov, *On existence and uniqueness of entropy solutions to the Cauchy problem for a conservation law with discontinuous flux*, J. Hyper. Differential Equations, **6** (2009), 525–548.

[16] N. Seguin, and J. Vovelle, *Analysis and approximation of a scalar conservation law with a flux function with discontinuous coefficients*, Math. Models Methods Appl. Sci. **13** (2003) 221–257.

[17] A.E. Tzavaras, *Wave interactions and variation estimates for self-similar zero-viscosity limits in systems of conservation laws*, Arch. Rational Mech. Anal., **135** (1996), 1–60.

E-mail address: `benjamin.boutin@univ-rennes1.fr`
E-mail address: `coquel@cmap.polytechnique.fr`
E-mail address: `contact@philippelefloch.org`

GLOBALLY OPTIMAL AND NASH EQUILIBRIUM SOLUTIONS FOR TRAFFIC FLOW ON NETWORKS

Alberto Bressan

Department of Mathematics, Penn State University
University Park, Pa. 16802, USA

Abstract. We consider a conservation law model of traffic flow on a network of roads, where drivers choose their departure times in order to minimize the sum of a departure cost and an arrival cost. Drivers can have different origins and destinations, and different cost functions. Under natural assumptions, two main results have been established: (i) the existence of a globally optimal solution, minimizing the sum of the costs to all drivers, and (ii) the existence of a Nash equilibrium solution, where no driver can lower his own cost by changing his departure time or the route taken to reach destination. In the special case of one single road, the global optimum and the Nash equilibrium are uniquely determined.

1. **Introduction.** Starting with the classical papers [14, 15], conservation law models have been widely used in the analysis of traffic flow [5, 6, 7, 10, 11, 12, 13]. Most of these studies were concerned with modeling, prediction, and control of traffic flow, on a single road or on a network of roads.

We adopt here a different perspective, looking at vehicular traffic in connection with decision problems [1, 2, 3, 4, 8, 9]. Traffic patterns are determined by the choices of a large number of individual drivers; each one choosing his departure time and the route to reach destination in an optimal way, for a given cost criterion.

To begin with a simple example, consider a group of drivers starting from a location A (a residential neighborhood), who wish to reach a destination B (a working place) at a given time T, all driving on the same road. There is a cost for starting early and a cost for arriving late. These costs can also account for the discomfort of waking up early in the morning or spending a long time stuck in traffic. Denoting by τ^d and τ^a respectively the departure and the arrival time, the total cost to each driver can be described as

$$\Psi \doteq \varphi(\tau^d) + \psi(\tau^a) \,. \tag{1}$$

For example, one could choose the penalty functions

$$\varphi(s) = -s \,, \qquad\qquad \psi(s) = ae^{b(s-T)} \tag{2}$$

for suitable constants $a, b > 0$. If L is the length of the road connecting A with B, and v is the (constant) speed of cars, then $\tau^a = \tau^d + \frac{L}{v}$ and the optimal departure

2000 *Mathematics Subject Classification.* Primary: 35L65, 49K20; Secondary: 90B20, 91A10.

Key words and phrases. Traffic flow, road network, scalar conservation law, global optima, user equilibria.

This work was partially supported by NSF through grant EFRI-1024707, A theory of complex transportation network design.

time for each driver is

$$\tau^d_{\text{opt}} \;=\; \underset{s}{\text{argmin}} \left\{ \varphi(s) + \psi\Big(s + \frac{L}{v}\Big) \right\}. \tag{3}$$

However, if everyone adopts this same strategy and departs exactly at the same time, in a real life situation a big traffic jam is created and this strategy is not optimal anymore. Clearly, the simple-minded solution (3) does not take into account the impact of traffic density on the velocity of cars.

Calling $\rho = \rho(t,x)$ the density of cars at time t at the point x along the road, we thus consider the conservation law

$$\rho_t + [\rho\, v(\rho)]_x \;=\; 0\,. \tag{4}$$

Here the decreasing function $v = v(\rho)$ describes the velocity of cars depending on the density. Let $\kappa > 0$ be the total number of drivers. In connection with the above model, a natural problem can be considered.

(I) - Global Optimization Problem. *Find a departure rate $\bar u(\cdot)$ which minimizes the combined total cost to all drivers.*

Let $x \in [0, L]$ be the space variable, denoting points along the road, and let

$$u(t,0) \;\doteq\; \rho(t,0)\, v(\rho(t,0)) \;=\; \bar u(t) \tag{5}$$

be the departure rate at time t, measuring how many drivers enter the highway per unit time. We regard $t \mapsto \bar u(t)$ as a control function, that can be assigned at will, subject to the obvious constraints

$$\bar u(t) \;\geq\; 0\,, \qquad\qquad \int_{-\infty}^{\infty} \bar u(t)\, dt \;=\; \kappa\,. \tag{6}$$

Let $\rho = \rho(t,x)$ be the solution of conservation law (4), defined for $(t,x) \in I\!\!R \times [0,L]$, with boundary data (5) assigned at $x = 0$, and let

$$u(t,x) \;\doteq\; \rho(t,x)\, v(\rho(t,x)) \qquad\qquad t \geq 0\,,\;\; x \in [0,L]$$

be the corresponding flux. The optimization problem can thus be stated as

$$\text{minimize:} \qquad J(u) \;=\; \int \varphi(t)\, u(t,0)\, dt + \int \psi(t)\, u(t,L)\, dt\,. \tag{7}$$

The above problem is relevant if there exists a central planner who can decide the departure time of all vehicles. In a more realistic situation each driver makes an individual choice, minimizing his own cost function given the traffic pattern determined by the decisions of all the other drivers. This leads to a different mathematical problem, namely:

(II) - Equilibrium problem. *Find a departure rate $\bar u(\cdot)$ which yields a* Nash equilibrium solution*, i.e., a solution where no driver can reduce his own cost by choosing a different departure time.*

Clearly, this implies that in an equilibrium solution all drivers share the same total cost (departure cost + arrival cost).

Recent results concerning the existence, uniqueness, and characterization of globally optimal and of Nash equilibrium solutions are described in Section 2. As mentioned in Section 3, the existence results remain valid also in the case of several groups of drivers, with different origins and destinations, different departure and arrival costs, traveling on a general networks of roads. The difficult issue of stability of Nash equilibria is discussed in Section 4.

2. **A single road.**

As usual, we assume that the flux function $\rho \mapsto \rho\, v(\rho)$ is strictly concave down and attains a positive maximum M for some density $\rho^* > 0$. Here M is the maximum flux of cars that can transit on the highway. Since in (6) we are not requiring that $\bar u(t) \leq M$, we need to specify what happens if the flux of cars arriving at the entrance of the highway is strictly larger that this maximum flux. As in [1], we shall simply assume that a queue is formed at the entrance of the highway. The length $q(t)$ of this queue varies in time according to

$$\dot q(t) \;=\; \begin{cases} \bar u(t) - M & \text{if } q(t) > 0\,, \\ 0 & \text{if } q(t) = 0\,. \end{cases}$$

As remarked in [1], it is convenient to switch the usual role of the variables t, x, and write (4) in the form of a conservation law for the flux $u = \rho v(\rho)$:

$$u_x + f(u)_t \;=\; 0\,, \qquad\qquad u(t,0) \;=\; \bar u(t)\,. \tag{8}$$

The function $u \mapsto f(u) \;=\; \rho$ is here defined as a partial inverse of the function $\rho \mapsto \rho\, v(\rho) = u$, mapping $[0, M]$ onto $[0, \rho^*]$. The advantage of this new formulation is that, instead of a boundary value problem, we now have a Cauchy problem, whose solution can be determined by the Lax formula. More precisely, let

$$U(t,x) \;\doteq\; \int_{-\infty}^{t} u(\tau, x)\, d\tau$$

be the total number of drivers that have crossed the point x along the highway before time t. Then the function U provides a viscosity solution to the Hamilton-Jacobi equation

$$U_x + f(U_t) \;=\; 0\,, \qquad\qquad U(t,0) \;=\; \overline U(t)\,. \tag{9}$$

Here $\overline U(t) = \int_{-\infty}^{t} \bar u(s)\, ds$ denotes the total number of drivers that have started their journey before time t (joining the queue at the entrance of the highway, if there is any). Interpreting $U = U(t,x)$ as the value function for an auxiliary optimization problem, and calling f^* the Legendre transform of f, for every $x > 0$ the solution of (9) is provided by

$$U(t,x) \;\doteq\; \inf_{\tau} \left\{ x\, f^*\Big(\frac{t-\tau}{x}\Big) + \overline U(\tau) \right\}. \tag{10}$$

Globally optimal solutions and Nash equilibrium solutions will be studied under the following natural set of assumptions.

(A1) The flux function $f : [0, M] \mapsto I\!R$ is continuous, increasing, and strictly convex. It is twice continuously differentiable on the open interval $]0, M[$ and satisfies

$$f(0) = 0\,, \qquad\qquad f''(u) \geq b > 0 \quad \text{for } 0 < u < M\,. \tag{11}$$

(A2) The cost functions φ, ψ are locally Lipschitz continuous and satisfy

$$\varphi' \;<\; 0\,, \qquad \psi, \psi' \;>\; 0\,, \qquad\qquad \lim_{x\to-\infty} \varphi(x) \;=\; \lim_{x\to+\infty} \Big(\varphi(x) + \psi(x)\Big) = +\infty\,. \tag{12}$$

The following results were proved in [1].

Theorem 1 (globally optimal solutions). *Under the assumptions (A1)-(A2), for every $\kappa > 0$ there exists a unique solution $u = u(t,x)$ of (8), with initial values $u(0,x) = \bar u$ satisfying the constraint (6), which minimizes the total cost (7).*

(i) The optimal solution is continuous, i.e., it contains no shocks. Moreover, it does not produce any queue at the entrance of the highway.

(ii) On the support of u, the sum of the departure and arrival costs along all characteristics is constant. More precisely, if $x \mapsto t(x) = t_0 + xf'(\bar{u}(t_0))$ is any characteristic line where $u > 0$, then

$$\varphi(t_0) + \psi(t(L)) \;=\; C\,, \tag{13}$$

for some constant C independent of t_0.

Theorem 2 (Nash equilibrium solutions). *Under the assumptions (A1)-(A2), for every $\kappa > 0$ there exists a unique Nash equilibrium solution $u = u(t,x)$ of (8), with initial values $u(0,x) = \bar{u}$ satisfying the constraint (6).*

Calling $\tau^a(t)$ the arrival time of a driver departing at time t, one has

$$\varphi(t) + \psi(\tau^a(t)) \;=\; C \tag{14}$$

for every t in the support of $\bar{u}$. Here C is a constant independent of t.

Remark 1. In general, the Nash equilibrium solution produces a queue at the entrance of the highway, and contains shocks. Notice the further difference between (13) and (14):

- For a globally optimal solution, the sum $\varphi(t(0)) + \psi(t(L))$ is constant along characteristics. These are straight lines, defined by $dx/dt = v(\rho) + \rho v'(\rho)$.
- For a Nash equilibrium solution, the sum $\varphi(\tau(0)) + \psi(\tau(L))$ is constant along car trajectories. These are curves defined by $dx/dt = v(\rho)$.

Remark 2. The Nash equilibrium satisfies a minimax property: among all admissible departure rates $\bar{u}$ which satisfy (6), it minimizes the maximum total cost to each driver. With the same notation as in (14), it was proved in [4] that the Nash equilibrium provides a solution to the problem

$$\text{minimize: } \sup \Big\{ \varphi(t) + \psi(\tau^a(t))\,;\ t \in Supp(\bar{u}) \Big\}.$$

For further analysis and examples of these solutions we refer to [1]. Continuous dependence on data was studied in [4].

3. **A network of roads.**

In this section we consider a more general model of traffic flow where drivers travel on a network of roads. Let $A_1, \ldots, A_m$ be the nodes of the network. Along the arc γ_{ij} connecting A_i with A_j, we assume that the flow of traffic is described by the conservation law

$$\rho_t + [\rho\, v_{ij}(\rho)]_x \;=\; 0\,. \tag{15}$$

Here t is time and $x \in [0, L_{ij}]$ is the space variable along γ_{ij}. We assume that the velocity v_{ij} is a continuous, nonincreasing function of the car density ρ. If $v_{ij}(0) > 0$ we say that the arc γ_{ij} is *viable*. If the two nodes i, j are not directly linked by a road, we simply take $v_{ij} \equiv 0$, so that the arc is not viable. We consider n groups of drivers traveling on the network, distinguished by the locations of departure and arrival, or by their cost functions. More precisely:

- All drivers of the k-th group depart from a node $A_{d(k)}$ and arrive at a node $A_{a(k)}$, but can choose different paths to reach destination.
- Any driver of the k-th group, departing at time τ^d and arriving at destination at time τ^a, will incur in the total cost $\varphi_k(\tau^d) + \psi_k(\tau^a)$.

For $k \in \{1,\ldots,n\}$, let G_k be the total number of drivers in the k-th group. Of course, we assume that there exists at least one chain of viable arcs

$$\Gamma \doteq \Big(\gamma_{i(0),i(1)}\,,\ \gamma_{i(1),i(2)}\,,\ \ldots\ ,\gamma_{i(\nu-1),i(\nu)}\Big) \tag{16}$$

with $i(0) = d(k)$ and $i(\nu) = a(k)$, connecting the departure node $A_{d(k)}$ with the arrival node $A_{a(k)}$. We denote by

$$\mathcal{V} \doteq \Big\{\Gamma_1,\, \Gamma_2,\ \ldots\ ,\Gamma_N\Big\}$$

the set of all viable paths (i.e. concatenations of viable arcs) which do not contain any closed loop. For a given $k \in \{1,\ldots,n\}$, we call $\mathcal{V}_k \subset \mathcal{V}$ the set of all viable paths for the k-drivers, connecting $A_{d(k)}$ with $A_{a(k)}$.

Let $G_{k,p}$ be the total number of k-drivers who travel along the path Γ_p. The assumption that every driver eventually reaches destination means that

$$\sum_{\Gamma_p\,\in\,\mathcal{V}_k} G_{k,p} = G_k \qquad\qquad \text{for every } k. \tag{17}$$

We shall use the Lagrangian variable $\beta \in [0, G_{k,p}]$ to label a particular driver in this subgroup. The departure and arrival time of this driver will be denoted by $\tau^d_{k,p}(\beta)$ and $\tau^a_{k,p}(\beta)$, respectively. With this notation, the definition of globally optimal and of Nash equilibrium solution can be more precisely formulated.

Definition 1. Given population sizes $G_1,\ldots,G_n$, a family of departure timings $\tau^d_{k,p} : [0, G_{k,p}] \mapsto I\!R$ is a **globally optimal solution** if it provides a global minimum to the functional

$$J \doteq \sum_{k,p} \int_0^{G_{k,p}} \Big(\varphi_k(\tau^d_{k,p}(\beta)) + \psi_k(\tau^a_{k,p}(\beta))\Big)d\beta\,, \tag{18}$$

subject to the constraints (17).

Definition 2. A family of departure timings $\{\tau^d_{k,p}\}$ is a **Nash equilibrium solution** if no driver can lower his total cost by changing departure time or the route taken to reach destination. This is the case if and only if there exist constants $c_1,\ldots,c_n$ such that:

(i) For almost every $\beta \in [0, G_{k,p}]$ one has

$$\varphi_k(\tau^d_{k,p}(\beta)) + \psi_k(\tau^a_{k,p}(\beta)) = c_k\,. \tag{19}$$

(ii) For all $\tau \in I\!R$, there holds

$$\varphi_k(\tau) + \psi_k(A_{k,p}(\tau)) \geq c_k\,. \tag{20}$$

Here $A_{k,p}(\tau)$ is the arrival time of a driver that starts at time τ from the node $A_{d(k)}$ and reaches the node $A_{a(k)}$ traveling along the path Γ_p.

Notice that (i) means that all k-drivers bear the same cost c_k, regardless of the path Γ_p that they take to reach destination. Moreover, (ii) means that no k-driver can achieve a cost $< c_k$ by choosing any other starting time τ.

We observe that, given the departure times $\tau^d_{k,p}(\beta)$, the corresponding arrival times $\tau^a_{k,p}(\beta)$ depend on the overall traffic pattern on the entire network. This is obtained by solving the various conservation laws (15) on every arc, with suitable conditions at junctions, specifying the priorities assigned to drivers that wish to enter the same road. A simple condition governing junctions was considered in [3], assuming that a separate queue can form at the entrance of each road. Drivers arriving at the node A_i from all incoming roads $\gamma_{\ell i}$, and who want to travel along the arc γ_{ij}, join a queue at the entrance of this outgoing arc. Their place in the queue is determined by the time at which they arrive at A_i, first in first out. With these modeling assumptions, the following results were proved in [3].

Theorem 3 (existence of globally optimal solutions on networks). *Assume that, for every viable arc γ_{ij} the corresponding flux function $f = f_{ij}$ satisfies (A1). Moreover, assume that for every $k = 1, \ldots, n$ the cost functions φ_k, ψ_k satisfy (A2). Then, for any n-tuple $(G_1, \ldots, G_n)$ of nonnegative numbers, there exists departure timings $\tau^d_{k,p} : [0, G_{k,p}] \mapsto I\!R$ satisfying (17) which yield a globally optimal solution of the traffic flow problem.*

Theorem 4 (existence of a Nash equilibrium solutions on networks). *Under the same assumptions on Theorem 3, for any n-tuple $(G_1, \ldots, G_n)$ of nonnegative numbers there exists departure timings $\tau^d_{k,p} : [0, G_{k,p}] \mapsto I\!R$ satisfying (17), which yield a Nash solution of the traffic flow problem.*

Remark 3. In the case of one group of drivers traveling on a single road, the uniqueness of the globally optimal solution stated in Theorem 1 is an easy consequence of the characterization (14). The uniqueness of the Nash equilibrium, stated in Theorem 2, is derived from a monotonicity argument. Indeed, the departure distribution $\overline{U}(t) = \int_{-\infty}^{t} \bar{u}(t)\, dt$ for a Nash equilibrium can be characterized as the pointwise supremum of a family of admissible distributions, satisfying an additional constraint of the form (14). In the case of several groups of drivers on a network of roads, the existence of a Nash equilibrium stated in Theorem 4 is proved by a fixed point argument. By its nature, this topological technique does not yield information about uniqueness or continuous dependence of the Nash equilibrium.

Remark 4. It would be interesting to see if the above theorems remain valid for more general models of road intersections. Instead of putting a buffer at the beginning of each outgoing road, the Riemann solvers studied in [6, 10, 11] allow for the back-propagation of queues along incoming roads. These models are more realistic. However, it is not clear if the corresponding solutions are sufficiently well behaved, in order to to apply the same arguments used in [3].

4. **Stability of the Nash equilibrium.**

For simplicity, consider one group of drivers on a single road. Assume that on a given day the departure rate is $\bar{u}(\cdot)$. If this is not a Nash equilibrium, on the next day some drivers may decide to depart at a different time, hopefully achieving a lower individual cost. An interesting question is whether, iterating this process, after several days the traffic pattern will approach an equilibrium solution.

To set the ideas, given a departure distribution $\bar{u}$, call

$$\Phi^{\bar{u}}(t) \;=\; \varphi(t) + \psi(\tau^a(t))$$

the total cost to a driver who departs at time t. Notice that the arrival time τ^a depends on the departure time t but also on the overall traffic pattern, i.e. on $\bar{u}$. Two models have been proposed in [4], describing how drivers can change their behavior day after day. To simplify the mathematical analysis, it is convenient to replace the discrete variable recording the day on the calendar by a continuous time variable s.

Model 1. Drivers who initially depart at time t continuously modify their departure time, depending on the gradient $\Phi^{\bar u}_t$ of the cost. The evolution of the departure rate $\bar u$ is then described by

$$\frac{d}{ds}\bar u \;=\; (\Phi^{\bar u}_t\,\bar u)_t\,. \tag{21}$$

Model 2. Drivers who depart at time t may decide to jump to a different departure time $\tau \in I\!R$, with probability proportional to the difference in cost. This leads to the integro-differential evolution equation

$$\frac{d}{ds}\,\bar u(t) \;=\; \int \bar u(\tau)\Big[\Phi^{\bar u}(\tau) - \Phi^{\bar u}(t)\Big]_+ d\tau - \int \bar u(t)\Big[\Phi^{\bar u}(t) - \Phi^{\bar u}(\tau)\Big]_+ d\tau\,, \tag{22}$$

where $[a]_+ \doteq \max\{a, 0\}$.

In both cases the key issue is whether, as $s \to \infty$, the departure rate $\bar u(\cdot)$ converges to the unique Nash equilibrium. At the present date, this problem is completely open. Quite surprisingly, numerical simulations reported in [4] indicate that the equilibrium solution may be unstable, while solutions of (21) or (22) approach a chaotic attractor. Even for initial data close to equilibrium, linearized stability has not been rigorously investigated.

REFERENCES

[1] A. Bressan and K. Han, *Optima and equilibria for a model of traffic flow.* SIAM J. Math. Anal. **43** (2011), 2384–2417.

[2] A. Bressan and K. Han, *Nash equilibria for a model of traffic flow with several groups of drivers.* ESAIM, Control, Optim. Calc. Var., **18** (2012), 969–986.

[3] A. Bressan and K. Han, *Existence of optima and equilibria for traffic flow on networks.* Netw. Heter. Media, to appear.

[4] A. Bressan, C. J. Liu, W. Shen, and F. Yu, *Variational analysis of Nash equilibria for a model of traffic flow.* Quarterly Appl. Math. **70** (2012), 495–515.

[5] Y. Chitour and B. Piccoli. *Traffic circles and timing of traffic lights for cars flow.* Discr. Cont. Dyn. Syst. Series B, **5** (2005), 599–630.

[6] G. M. Coclite, M. Garavello, and B. Piccoli, *Traffic flow on a road network.* SIAM J. Math. Anal. **36** (2005), 1862–1886.

[7] C. Daganzo, *"Fundamentals of Transportation and Traffic Operations"* Pergamon-Elsevier, Oxford, 1997.

[8] T. L. Friesz, *"Dynamic Optimization and Differential Games"*, Springer, New York, 2010.

[9] T. L. Friesz, D. Bernstein, T. E. Smith, R. L. Tobin, and B. W. Wie, *A variational inequality formulation of the dynamic network user equilibrium problem.* Oper. Res. **41** (1993), 179–191.

[10] M. Garavello and B. Piccoli, *"Traffic Flow on Networks. Conservation Laws Models".* AIMS Series on Applied Mathematics, Springfield, Mo., 2006.

[11] M. Garavello and B. Piccoli, *Traffic flow on complex networks.* Ann. Inst. H. Poincaré, Anal. Nonlin. **26** (2009), 1925–1951.

[12] M. Gugat, M. Herty, A. Klar, and G. Leugering. *Optimal control for traffic flow networks.* J. Optim. Theory Appl. **126** (2005), 589–616.

[13] D. Helbing, A. Hennecke, and V. Shvetsov, *Micro- and macro-simulation of freeway traffic.* Math. Computer Modeling **35** (2002), 517–547.

[14] M. Lighthill and G. Whitham, *On kinematic waves. II. A theory of traffic flow on long crowded roads.* Proceedings of the Royal Society of London: Series A, **229** (1955), 317–345.

[15] P. I. Richards, *Shock waves on the highway.* Oper. Res. **4** (1956), 42–51.

E-mail address: bressan@math.psu.edu

HOW TO IMPROVE THE DECAY OF THE NUMERICAL ERROR FOR LARGE TIMES: THE CASE OF DISSIPATIVE BGK SYSTEMS

Maya Briani

Istituto per le Applicazioni del Calcolo "Mauro Picone"
Consiglio Nazionale delle Ricerche, Italy

Denise Aregba-Driollet and Roberto Natalini

IMB, UMR CNRS 5251, Université de Bordeaux, France
and
Istituto per le Applicazioni del Calcolo "Mauro Picone"
Consiglio Nazionale delle Ricerche, Italy

Abstract. We introduce new finite differences schemes to approximate one dimensional dissipative semilinear hyperbolic systems with a BGK structure. Using accurate analytical time-decay properties of the local truncation error, it is possible to design schemes based on standard upwinding schemes, which are increasingly accurate for large times when computing small perturbations of constants asymptotic states.

1. Introduction.

Consider the following class of one dimensional BGK systems:

$$\partial_t f^i + \lambda_i \partial_x f^i = M_i(u) - f^i, \quad i = 1, ..., m. \tag{1}$$

Here $f^i \in \mathbb{R}^k$, $u := \sum_{i=1}^m f^i$, $x \in \mathbb{R}$ and $t > 0$, and the functions $M_i = M_i(u) \in \mathbb{R}^k$ are smooth functions of u such that: $\sum_{i=1}^m M_i(u) = u$.

To obtain the time decay rates of these solutions, we need to rewrite the problem in more suitable coordinates. Following [3], we rewrite the BGK system in its *conservative-dissipative* form for the new unknowns

$$Z = (u, \tilde{Z})^T.$$

It is proved in [3] that, under some dissipativity conditions and for initial data which are small and smooth in some suitable norms, the time decay of the global solutions, for large times and in the L^∞-norm, is given by

$$\partial_x^l u = O(t^{-1/2-l/2}), \quad \partial_x^l \tilde{Z} = O(t^{-1-l/2}),$$

and similar estimates are available for their time derivatives. Notice that the improved estimate for $\tilde{Z}$ can only be obtained in these new coordinates.

The aim of this paper is to give a brief overview of the way it is possible to take advantage of these precise decay estimates to build up more accurate numerical schemes. Actually, we can see that, for standard numerical schemes, like for instance

2000 *Mathematics Subject Classification.* Primary: 65M12; Secondary: 35L65.

Key words and phrases. Finite differences methods, dissipative hyperbolic problems, BGK systems, asymptotic behavior, asymptotic high order schemes.

the upwind scheme with the source term approximated pointwise by the standard Euler scheme, the truncation error has the following decay as $t \to +\infty$:

$$\begin{aligned} \mathcal{T}_u(x,t) &= O(\Delta x\ t^{-3/2}) + O(\Delta t\ t^{-3/2}), \\ \mathcal{T}_{\tilde{Z}}(x,t) &= O(\Delta x\ t^{-3/2}) + O(\Delta t\ t^{-3/2}). \end{aligned} \tag{2}$$

It can be seen numerically that the corresponding absolute error, for a fixed space step, decays as

$$e_u(t) = O(t^{-1/2}), e_z(t) = O(t^{-1}),$$

which implies that the relative error is essentially constant in time.

Here, our main goal is to improve the decay estimates on the truncation order to achieve an effective decay in time of the relative error, both in u and $\tilde{Z}$. To obtain this result, we perform a detailed analysis of the behavior of the truncation error for a general class of schemes, called "Time Asymptotically High Order" (TAHO) schemes, which generalize those introduced in [2]. Thanks to this analysis, we are able to select some schemes such that the truncation order behaves as

$$\mathcal{T}_u(x,t) = O(\Delta x\ t^{-2}),\ \mathcal{T}_{\tilde{Z}}(x,t)) = O(\Delta x\ t^{-2}), \tag{3}$$

for a fixed CFL ratio and such that the numerical error observed in the practical tests improves of $t^{-1/2}$ on other schemes.

The plan of the paper is the following. In Section 2, we introduce our analytical framework. The main schemes are derived in Section 3, where we show how to improve the time decay of their local truncation error. Section 4 presents some numerical tests which show the nice behavior of our new schemes in two test cases.

2. **The analytical framework.**

Following [3], we rewrite system 1 in its *conservative-dissipative* form. This means that we assume that there exists an invertible matrix

$$D = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix}, \tag{4}$$

such that, setting $m_1 = k$, $m_2 = k(m-1)$, the new unknown

$$Z = Df = (u, \tilde{Z})^T \in \mathbb{R}^k \times \mathbb{R}^{m_2}, \tag{5}$$

solves the system

$$\begin{cases} \partial_t u + A_{11}\partial_x u + A_{12}\partial_x \tilde{Z} &= 0, \\ \partial_t \tilde{Z} + A_{21}\partial_x u + A_{22}\partial_x \tilde{Z} &= \tilde{Q}(u) - \tilde{Z}, \end{cases} \tag{6}$$

where A is symmetric and $\tilde{Q}(u)$ is quadratic in u, i.e.: $\tilde{Q}(0) = 0$ and $\tilde{Q}'(0) = 0$. Observe that, after this transformation, which a priori is not unique, the source term is zero in the first component and the second one is the sum of a quadratic term and of the dissipative term $-\tilde{Z}$.

Moreover, when transforming system 1 in system 6, we can always assume that blocks D_{11} and D_{12} have the special form

$$D_{11} = I_k, \quad D_{12} = (I_k I_k \cdots I_k) \in \mathbb{R}^{k \times m_2},$$

and, setting $\Lambda = diag(\lambda_1 I_k, ..., \lambda_m I_k)$, we have that

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = D\Lambda D^{-1}.$$

Therefore, we can rewrite our system in a more compact form:

$$\partial_t Z + A\partial_x Z = -Z + DM(u). \tag{7}$$

To guarantee the existence of the matrix D in 4, we can assume that our system is strictly entropy dissipative in the sense of [5] and verifies the Shizuta-Kawashima condition [6, 5, 3].

For instance, using Bouchut's Entropy dissipation condition [4], it is possible to prove the existence of a matrix D in 4, with all the above properties.

2.1. **The simplest example: the Jin-Xin 2×2 relaxation system.**

Consider the following system

$$\begin{cases} \partial_t u + \partial_x v = 0, \\ \partial_t v + \lambda^2 \partial_x u = F(u) - v, \end{cases} \tag{8}$$

with $\lambda > 0$. The unknowns u and v are scalar and the function $F = F(u)$ is smooth, with $F(0) = 0$. This case is obtained from 1 for $k = 1$, $m = 2$, and $\lambda_2 = -\lambda_1 = \lambda$, by setting

$$u = f^1 + f^2,\ v = \lambda(f^2 - f^1),\ F(u) = \lambda(M_2 - M_1).$$

Under the condition

$$\lambda > |F'(0)|, \tag{9}$$

the problem is dissipative, at least in a small neighborhood of the origin, in the sense of [5] and the Shizuta-Kawashima condition is verified.

In this case the conservative-dissipative form is obtained by using

$$D = \begin{pmatrix} 1 & 1 \\ -\mu a_+ & \mu a_- \end{pmatrix},$$

where $a = F'(0)$, $\mu = (\lambda^2 - a^2)^{-1/2}$ is real and positive, $a_\pm = \lambda \pm a > 0$, from assumption 9.

2.2. **A 3×3 BGK example.**

Let us now compute the conservative-dissipative form for the following 3×3 BGK model,

$$\begin{cases} \partial_t f_1 - \lambda \partial_x f_1 = M_1(u) - f_1, \\ \partial_t f_2 = M_2(u) - f_2, \\ \partial_t f_3 + \lambda \partial_x f_3 = M_3(u) - f_3. \end{cases}$$

Let $F = F(u)$ be a smooth scalar function such that $F(0) = 0$ and let γ be such that $\gamma'(u) = |F'(u)|$, with $\gamma(0) = 0$. We choose our three maxwellian functions as follows, for $\beta \in]0, 1[$ and $\lambda > 0$

$$M_1(u) = \frac{1}{2}\left(\frac{\gamma(u) - F(u)}{\lambda} + \beta u\right), M_3(u) = \frac{1}{2}\left(\frac{\gamma(u) + F(u)}{\lambda} + \beta u\right),$$

$$M_2(u) = u - M_1(u) - M_3(u) = (1 - \beta)u - \frac{\gamma(u)}{\lambda}.$$

The functions M_i, $i = 1, 2, 3$, are strictly increasing if for any u under consideration

$$\lambda > \frac{|F'(u)|}{1 - \beta},$$

and so the entropy dissipation condition [4] is verified. Let $a = F'(0)$ and $\alpha = |a| + \beta\lambda$, following the results in [4, 5, 3], the matrix D for the transformation in the conservative–dissipative form 5 is given by

$$D = \begin{pmatrix} 1 & 1 & 1 \\ \frac{\alpha+a}{\alpha-a}\sqrt{\frac{\lambda(\alpha-a)}{\alpha(\alpha+a)}} & 0 & -\sqrt{\frac{\lambda(\alpha-a)}{\alpha(\alpha+a)}} \\ -\sqrt{\frac{\lambda-\alpha}{\alpha}} & -\sqrt{\frac{\lambda-\alpha}{\alpha}}+\frac{\lambda}{\sqrt{\alpha(\lambda-\alpha)}} & -\sqrt{\frac{\lambda-\alpha}{\alpha}} \end{pmatrix}.$$

3. **The numerical approximation.** In this section we first introduce general finite difference approximations for system 1. Then, we compute the local truncation error of these schemes and we discuss its decays properties. The main result is given in Theorem 3.1, where a class of Time Asymptotically High Order (TAHO) schemes is fully characterized. First, we approximate the differential part following the direction of the characteristic velocities, so we study the methods for the system in diagonal form 1.

We denote by $f = (f^1, ..., f^m)$ the exact solution. Let Δx the uniform mesh-length and $x_j = j\,\Delta x$ the spatial grid points for all $j \in \mathbb{Z}$. The time levels t_n, with $t_0 = 0$, are also spaced uniformly with mesh-length $\Delta t = t_{n+1} - t_n$ for $n \in \mathbb{N}$. We denote by ρ the CFL ratio $\rho = \Delta t/\Delta x$, which is taken constant through all the paper.

We consider the Cauchy problem for system 1 possibly subjected to some stability conditions. The initial data f^0 is supposed to be smooth and approximated by its node values. The approximate solution $(f^1_{j,n}, ..., f^m_{j,n})^T$, $f^i_{j,n} \in \mathbb{R}^k$, $i = 1, ..., m$, for $j \in \mathbb{Z}$ and $n \in \mathbb{N}$, is given by

$$\begin{aligned} &\frac{f^i_{j,n+1} - f^i_{j,n}}{\Delta t} + \frac{\lambda_i}{2\Delta x}\left(f^i_{j+1,n} - f^i_{j-1,n}\right) - \frac{q_i}{2\Delta x}\delta^2_x f^i_{j,n} \\ &= \sum_{l=-1,0,1} \left(\mathcal{B}^i_l(u_{j+l,n}) - \beta^i_l f^i_{j+l,n}\right), \end{aligned} \tag{10}$$

with $f^i_{j,0} = f^i_0(x_j)$ and $\delta^2_x f_{j,n} = (f_{j+1,n} - 2f_{j,n} + f_{j+1,n})$, for all $i = 1, ..., m$. The artificial diffusion terms q_i are diagonal matrices in $\mathbb{R}^{k\times k}_+$. The source term approximation is defined, for $l = -1, 0, 1$, by the diagonal matrices $\beta^i_l \in \mathbb{R}^{k\times k}$ and by the vectors of functions $\mathcal{B}^i_l(\cdot) \in \mathbb{R}^k$.

We assume the scheme 10 is consistent with system 1, i.e, for all $i = 1, ..., m$

$$\beta^i_{-1} + \beta^i_0 + \beta^i_1 = I_k + \Delta x C^i,$$
$$\mathcal{B}^i_{-1}(u) + \mathcal{B}^i_0(u) + \mathcal{B}^i_1(u) = M_i(u) + \Delta x \mathcal{C}_i(u),$$

where $C^i = diag(c^i_1, ..., c^i_k) \in \mathbb{R}^{k\times k}$ and $\mathcal{C}_i(u)$ are k functions to be defined.

3.1. **Decay properties of the local truncation error.** In this section we focus on the local truncation error for the general scheme 10. By applying the time decay properties given in [3], we will show how it is possible to build up numerical schemes which are more accurate for large times.

Set, for $i = 1, ..., m$,

$$C = diag(C^i), \quad \bar{C} = DCD^{-1}, \quad \mathcal{C}(u) = (\mathcal{C}_i(u))^T, \quad \gamma^i = (\beta^i_1 - \beta^i_{-1}).$$

Scheme 10 is clearly consistent. Now, using the time decay estimates in [3] and similar estimates for their time derivatives, we obtain for a general approximation the following estimates for the local truncation error, as $t \to +\infty$:

$$\mathcal{T}_u(x,t) = O(\Delta x\ t^{-3/2}) + O(\Delta t\ t^{-3/2}), \mathcal{T}_z(x,t) = O(\Delta x\ t^{-3/2}) + O(\Delta t\ t^{-3/2}).$$

We would like to improve the decay property of this local truncation error to build up more accurate numerical schemes. The main idea is to chose the free parameters of the scheme to delete the terms that decay more slowly in the Taylor expansion of the local truncation error (see [1] for details).

Let $g_i = diag(\gamma_{(i-1)k+1}, ..., \gamma_{ik})$ for $i = 1, ..., m$ and $G = diag(g_1, ..., g_m)$.

Theorem 3.1 (Local Truncation Error). *Let $\Delta t/\Delta x = \rho$ be fixed and let $H = diag(h_1, \dots, h_m)$ be the block diagonal matrix given by $H = D^T D$ and set $P = \sum_{i=1}^m \lambda_i^2 h_i^{-1}$. Assume $A_{11} \neq 0$ and that the following condition holds:*

the matrix $(\lambda_i I_k - A_{11})$ is invertible for $i = 1, ..., m$.

If we make the following choice for the coefficients of the scheme 10,

$$C = -\frac{\rho}{2} I_{km}, \quad \mathcal{C} = CM(u) = -\frac{\rho}{2} M(u), \tag{11}$$

$$g_i = -\left(\frac{1}{2} q_i h_i^{-1} - \frac{\rho}{2} h_i^{-1}\left(P - (\lambda_i I_k - A_{11})^2\right)\right)(\lambda_i I_k - A_{11})^{-1} h_i \tag{12}$$

and

$$\Gamma_i'(u) = g_i M_i'(u) + \frac{\rho}{2}\left((h_i^{-1} - M_i'(u))A_{11} + \lambda_i M_i'(u) - h_i^{-1}\sum_{j=1}^m \lambda_j M_j'(u)\right), \tag{13}$$

both for $i = 1, ..., m$, then the scheme 10 is TAHO and the local truncation error reads

$$\mathcal{T}_u(x,t) = O\left(\Delta x\ t^{-2}\right) + O\left(\Delta x^2\ t^{-3/2}\right), \mathcal{T}_z(x,t) = O\left(\Delta x\ t^{-2}\right) + O\left(\Delta x^2\ t^{-3/2}\right).$$

For the proof and further considerations in case $A_{11} = 0$ we refer to [1].

4. **Numerical tests.**

In this Section we show how, for large time simulations, TAHO schemes give better numerical results than standard approximations for both examples considered in Sections 2.1 and 2.2.

Specifically, we shall compare our TAHO scheme with two numerical approximations: i) a source pointwise approximation, denoted by STD and defined by 10 with $\Gamma_i = 0$, $\gamma^i = 0$, $C^i = 0$ and $\mathcal{C}_i = 0$, for $i = 1 =, ..., m$; ii) a source upwinding approximation, denoted by ROE and defined by

$$\begin{aligned}
&\frac{f^i_{j,n+1} - f^i_{j,n}}{\Delta t} + \frac{\lambda_i}{2\Delta x}\left(f^i_{j+1,n} - f^i_{j-1,n}\right) - \frac{|\lambda_i|}{2\Delta x}\delta_x^2 f^i_{j,n} \\
=&\frac{M_i(u^n_{j-1}) + 2M_i(u^n_j) + M_i(u^n_{j+1})}{4} + \frac{sgn(\lambda_i)}{4}(M_i(u^n_{j-1}) - M_i(u^n_{j+1})) \\
&-\frac{f^i_{j-1,n} + 2f^i_{j,n} + f^i_{j+1,n}}{4} - \frac{sgn(\lambda_i)}{4}(f^i_{j-1,n} - f^i_{j+1,n}).
\end{aligned} \tag{14}$$

To complete the definition of scheme 10 coupled with conditions 11-13 it is still necessary to choose some free parameters, such as for the 2×2 case $\mathcal{B}_0^{1,2}(\cdot)$ and $\beta_0^{1,2}$. For both cases considered, such parameters can be defined by applying monotonicity conditions to the scheme. We refer to [1] for more details.

For all tests, we focus our attention on the numerical error as a function of time: we plot the error $e(t) = \|(u^H - U^h)(t)\|_{L^\infty}$ as the time $t = n\Delta t$ increases, where u^H is the reference solution obtained by the ROE scheme 14, with $\Delta x = \mathcal{O}(10^{-4})$.

Then, given different numerical approximations U^h, we look for a constant C and γ which best fit the equality

$$e(t) = \|(u^H - U^h)(t)\|_{L^\infty} = Ct^{-\gamma}. \tag{15}$$

Given N data points $(t_i, e(t_i))_{i=1,N}$, we shall define γ and C as the solution of the following least squares problem,

$$\min_{C,\gamma} \sum_{i=1}^{N} |\ln(e(t_i)) - \ln(Ct^{-\gamma})|^2.$$

For all schemes, we fix the steps ratio ρ to verify all the CFL conditions. Since all schemes are of first order approximation, to emphasize the good behaviour of TAHO compared to the others schemes, we compute the numerical solutions U^h by using a quite big grid step $\Delta x = \mathcal{O}(10^{-1})$.

All numerical results we present show that for standard approximations, such as STD and ROE, the absolute error $e(t)$, for a fixed space step, decays as

$$e_u(t) = O(t^{-1/2}),\ e_z(t) = O(t^{-1}),$$

while for the TAHO scheme, it improves of $t^{-1/2}$ on the previous schemes.

4.1. **Results for the Jin–Xin 2×2 system.** We fix $q = \lambda$ and we compare for the 2×2 case the TAHO scheme coupled with monotonicity assumptions, with ROE and STD scheme. We shall consider as initial datum the function

$$u_0 = \chi_{[-1,1]}\left(-x^2 + 1\right), \quad z_0 = \frac{1}{\lambda}F(u_0(x)),$$

and we fix

$$F(u) = a\left(u - u^2\right).$$

The numerical results 1 show a better performance of the TAHO scheme; for both conservative and dissipative variable, the numerical solution obtained by TAHO fit better the benchmark curve. Again, the decay of the errors $e_u(t)$ and $e_z(t)$, Figure 1-(c)-(d), goes faster for the TAHO scheme, as confirmed by Table 1. There the decay parameter γ is numerically computed for all three schemes. The value obtained for the TAHO scheme improve of $t^{-1/2}$ on the others. We stress on that the numerical solutions are computed with quite big step $\Delta x = 0.1$.

4.2. **Results for the 3×3 system.** As initial data, we take the smooth function u_0 defined by

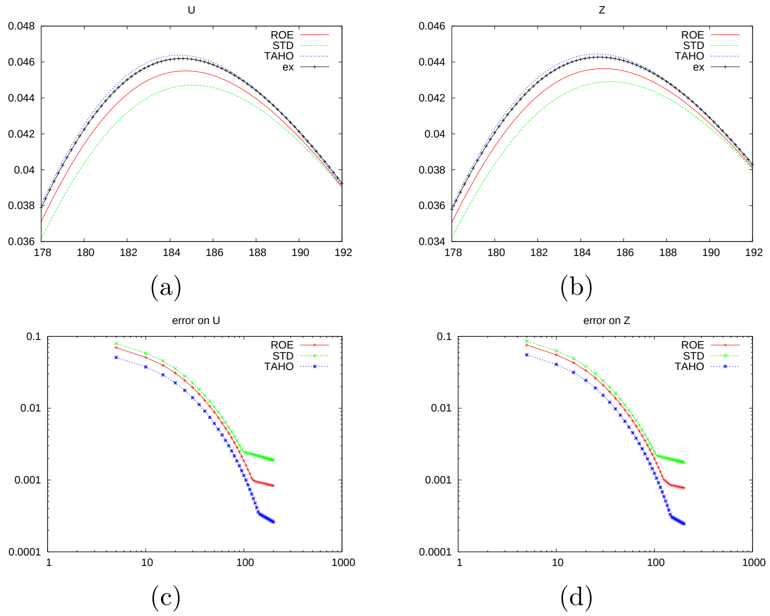
$$u_0(x) = \chi_{[-1,1]} \exp\left(1 - \frac{1}{1 - x^2}\right).$$

Then we set $f_0(x) = M(u_0(x))$. We choose $a = 1$, $\lambda = 2.1$, $\beta = (\alpha - a)/\lambda = 0.1$. The discretization parameters are $\Delta x = 0.1$, $\rho = \frac{1}{2\lambda}$, which satisfy all monotonicity requirements, see [1].

The numerical results show as in the 2×2 case a better performance of the TAHO scheme. In Figure 2-(a)-(b), we plot the time evolution of the l^∞ errors $e_u(t)$ and $e_z(t)$. They show how for the TAHO scheme both errors decay as time increases more quickly than other. This result is also confirmed by Table 2, where the values of γ and C are computed. Looking at the different values of γ, it is clear that for the TAHO approximation the decay velocity of the absolute error improves of $t^{-1/2}$ on the previous schemes.

scheme	C_u	γ_u	C_z	γ_z
STD	0.013797	0.374708	0.010744	0.341554
ROE	0.004874	0.333634	0.007850	0.439996
TAHO	0.111380	1.151517	0.495480	1.451030

TABLE 1. The 2×2 Test, see subsection 4.1. Evaluation of constants γ and C for $e_u(t) = C_u t^{-\gamma_u}$ and $e_z(t) = C_z t^{-\gamma_z}$ defined in 15. For standard approximation STD and ROE, the absolute error decays as $e_{u,z}(t) \approx O(t^{-1/2})$; while, for the TAHO scheme it improves of $t^{-1/2}$.



(a)

(b)

(c)

(d)

FIGURE 1. The 2×2 Test, see subsection 4.1. (a)-(b) Zoom on the solutions u and z respectively obtained by the different schemes at final time T. The plot show that TAHO scheme gives better results than others with a quite big step $\Delta x = 0.1$. (c)-(d) Time evolution of the l^∞ errors $e_u(t)$ and $e_z(t)$ defined in 15 for the different schemes. As expected by our asymptotic analysis, for the TAHO scheme the absolute errors $e_{u,z}(t)$ decay faster as the time increases. This result is confirmed in Table 1, where we compute the decay parameters γ of absolute errors previously plotted.

REFERENCES

[1] D. Aregba-Driollet, M. Briani and R. Natalini, *Time Asymptotic High Order schemes for dissipative BGK hyperbolic systems*, preprint, arXiv:math.NA/1207.6279v1

[2] D. Aregba-Driollet, M. Briani and R. Natalini, *Asymptotic high-order schemes for 2X2 dissipative hyperbolic systems*, SIAM J. Numer. Anal. **46** (2008), no.2, 869–894

scheme	C_u	γ_u	C_z	γ_z
STD	0.0052	0.54	0.0064	1.1
ROE	0.0027	0.66	0.0036	1.2
TAHO	0.006	1	0.012	1.62

TABLE 2. The 3×3 Test, see section 4.2. Evaluation of constants γ and C for $e_u(t) = C_u t^{-\gamma_u}$ and $e_z(t) = C_z t^{-\gamma_z}$ defined in 15. For STD and ROE approximations, the numerical results show that the absolute error decays as $e_u(t) = O(t^{-1/2})$ and $e_z(t) = O(t^{-1})$; while, for TAHO's it improves of $t^{-1/2}$.

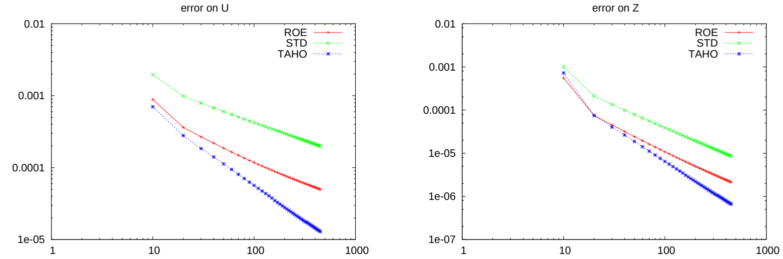


FIGURE 2. The 3×3 Test, see section 4.2. Absolute errors for u component (left) and Z components (right) with respect to time.

[3] S. BIANCHINI, B. HANOUZET AND R. NATALINI, *Asymptotic behavior of smooth solutions for partially Dissipative hyperbolic systems with a convex entropy*, Communications Pure Appl. Math. **60** (2007), 1559–1622.

[4] F. Bouchut, *Construction of BGK models with a family of kinetic entropies for a given system of conservation law*, J. Statist. Phys. **95** (1999), 113–170.

[5] B. Hanouzet, R. Natalini, *Global existence of smooth solutions for partially dissipative hyperbolic systems with a convex entropy*, Arch. Ration. Mech. Anal. **169** (2003), no. 2, 89–117.

[6] Y. Shizuta, S. Kawashima, *Systems of equations of hyperbolic–parabolic type with applications to the discrete Boltzmann equation*, Hokkaido Math. J. 14 (1984) 435–457.

E-mail address: m.briani@iac.cnr.it
E-mail address: Denise.Aregba@math.u-bordeaux1.fr
E-mail address: r.natalini@iac.cnr.it

ON THE ASYMPTOTICS OF SOLUTIONS TO RESONATOR EQUATIONS

Buğra Kabil

Institute of Applied Analysis and Numerical Simulation
Department of Mathematics
University of Stuttgart
Pfaffenwaldring 57, 70569 Stuttgart, Germany

Abstract. In this paper, we consider a system of micro-beam resonators within the thermoelastic theory of Lord and Shulman. It is a particular case of a thermoelastic system given by a coupling of a plate equation to a hyperbolic heat equation arising from Cattaneo's law of heat conduction. In a bounded domain, the system can be damped by a term such that it is exponentially stable. In the whole space, one can determine the decay rates for the system.

1. **Introduction.**

Resonators are systems which naturally oscillate at some frequencies, called its resonant frequencies. There are many kinds of resonators. We consider mechanical resonators. Microresonators have high sensitivity at room temperature. Thermoelastic damping is one of the reasons for the dissipation or loss of energy from the system to its surroundings, see [8, 1, 2].

We model the problem of thermoelastic damping in micro-resonators by coupling the plate equation to a modified heat equation with one relaxation parameter proposed by Lord and Shulman [4]. One can find a good review of the relevance of the thermoelastic damping and the derivation for the one dimensional case in [8].

We consider the system in the dimensionless form. The system of equations reads as

$$a\Delta^2 u + \Delta\theta + u_{tt} = F, \tag{1}$$

$$\Delta\theta - m\theta + d\Delta\hat{u}_t = c\hat{\theta}_t + G, \tag{2}$$

where $\hat{f} = f + \tau f_t$.

We assume first that a, d, c and τ are positive constants. The constant m may be non-negative. F and G correspond to the external force and heat supply. First, we consider a bounded domain $B \subset \mathbb{R}^n$ whose boundary satisfies the assumptions of the divergence theorem. In this paper solutions $(u, \theta) = (u(x,t), \theta(x,t))$ with $x \in B$, $t \geq 0$ are considered. The initial conditions are given by

$$u(x,0) = u_0(x), \quad u_t(x,0) = u_1(x), \quad \theta(x,0) = \theta_0(x), \quad \theta_t(x,0) = \theta_1(x) \tag{3}$$

and the boundary conditions read

$$u(x,t) = \Delta u(x,t) = \theta(x,t) = 0, \quad x \in \partial B, \quad t \in [0,\infty) \tag{4}$$

2000 *Mathematics Subject Classification.* 35L30.

Key words and phrases. Hyperbolic models in thermoelasticity, mechanical resonator, exponential stability, decay rates.

or

$$u(x,t) = \nabla u(x,t) \cdot n(x) = \theta(x,t) = 0, \quad x \in \partial B, \quad t \in [0,\infty). \tag{5}$$

In this context $n(x)$ is the outer normal vector to ∂B at a certain $x \in \partial B$.

2. **Exponential Stability of a damped System.** The system (1), (2) in a bounded domain with the initial conditions (3) and boundary conditions(4) or (5) was partly considered in [5]. Unfortunately, [5] includes a mistake which was corrected in [6]. It is shown in [6] that the associated semigroup $\left\{e^{tA}\right\}_{t\geq 0}$ for $\tau > 0$ is **not** exponentially stable. This effect can also be observed in some Timoshenko systems, see [7].

Next, we want to introduce an additional term, called damping, to assure for the exponential stability. The damped system has the form

$$\begin{aligned} a\Delta^2 u + \Delta\theta + u_{tt} + \gamma u_t &= 0, &(6)\\ \Delta\theta - m\theta + d\Delta\hat{u}_t &= c\hat{\theta}_t, &(7)\end{aligned}$$

where $\gamma > 0$ is a damping factor. The natural energy is given by

$$E(t) = \int\limits_B (d\hat{u}_t^2 + ad|\Delta\hat{u}|^2 + c\hat{\theta}^2 + \tau(|\nabla\theta|^2 + m\theta^2))dB.$$

One can show that

$$\frac{d}{dt}E(t) = -2\int\limits_B (|\nabla\theta|^2 + m\theta^2)dB - 2\gamma d\int\limits_B \hat{u}_t^2 dB. \tag{8}$$

Our aim is to determine a suitable Lyapunov functional which is equivalent to the energy. First of all we define

$$\eta(x,t) := \int\limits_0^t \theta(x,s)ds. \tag{9}$$

Letting Q be the solution to

$$\Delta Q - mQ = [c\theta_0 + c\tau\theta_1 - d\Delta u_0 - d\tau\Delta u_1] \tag{10}$$

with homogenous boundary conditions $Q = 0$ on ∂B, we observe that $\beta := \eta + Q$ satisfies the homogenized equation

$$\Delta\beta - m\beta = c\hat{\theta} - d\Delta\hat{u}. \tag{11}$$

Now we define the Lyapunov functional

$$F(t) := NE(t) + \int\limits_B \left(\frac{1}{2}(|\nabla\beta|^2 + m\beta^2) + \tau\nabla\beta\nabla\beta_t + m\tau\beta\beta_t + d\hat{u}\hat{u}_t + \frac{\gamma d}{2}\hat{u}^2\right) dB,$$

for a constant $N > 0$ being arbitrary for a moment. It yields

$$\begin{aligned}\frac{d}{dt}F(t) &= -2N\int\limits_B(|\nabla\theta|^2+m\theta^2)dB - 2N\gamma\int\limits_B d\hat{u}_t^2 dB - \int\limits_B c\hat{\theta}^2 dB\\ &- \int\limits_B ad|\Delta\hat{u}|^2 dB + \int\limits_B d\hat{u}_t^2 dB + \tau\int\limits_B(|\nabla\theta|^2+m\theta^2)dB\\ &= \left(\frac{-2N+\tau}{\tau}\right)\tau\int\limits_B(|\nabla\theta|^2+m\theta^2)dB + (-2N\gamma+1)\int\limits_B d\hat{u}_t^2 dB\\ &- \int\limits_B c\hat{\theta}^2 dB - \int\limits_B ad|\Delta\hat{u}|^2 dB.\end{aligned}$$

We choose N to satisfy

$$\frac{-2N}{\tau}+1<0 \qquad \text{and} \qquad -2N\gamma+1<0. \tag{12}$$

Thus, we obtain

$$\frac{d}{dt}F(t) \le -\underbrace{\min\left\{\frac{2N}{\tau}-1, 2N\gamma-1, 1\right\}}_{=:C}\underbrace{\int\limits_B(d\hat{u}_t^2+ad|\Delta\hat{u}|^2+c\hat{\theta}^2+\tau(|\nabla\theta|^2+m\theta^2))dB}_{=E(t)}$$

meaning

$$\frac{d}{dt}F(t) \le -CE(t), \qquad C>0.$$

Using the homogenized equation (11), one can show the equivalence of the functional and the energy, i.e.,

$$\exists c_1, c_2 > 0: \quad c_1E(t) \le F(t) \le c_2E(t).$$

In particular, we can show that there is a positive constant C such that

$$|F(t) - NE(t)| \leqslant CE(t).$$

Altogether we obtain for generic positive constant C

$$\frac{d}{dt}F(t) \le -CF(t).$$

Gronwall's lemma implies

$$F(t) \le e^{-Ct}F(0).$$

This yields the exponential stability of the damped system.

Theorem 2.1. *For the energy of the damped system (6), (7), there exist constants $\tilde{c}_1 > 0$ and $\tilde{c}_2 > 0$ independent from the initial data such that*

$$E(t) \le \tilde{c}_1E(0)e^{-\tilde{c}_2 t} \tag{13}$$

holds for all $t \ge 0$.

Remark 1. One can similarly show that the following damped system is also exponentially stable:

$$\begin{aligned}a\Delta^2u+\Delta\theta+u_{tt}-\gamma\Delta u_t &= 0, &(14)\\ \Delta\theta - m\theta + d\Delta\hat{u}_t &= c\hat{\theta}_t. &(15)\end{aligned}$$

3. **Asymptotics and Decay Rates.** The resonator equations in the whole space have not been studied in the literature before. Here, we want to study the asymptotic behaviour of the solutions to the Cauchy problem

$$\begin{aligned} a\Delta^2 u + \Delta\theta + u_{tt} &= 0, && (16)\\ \Delta\theta - m\theta + d\Delta\hat{u}_t &= c\hat{\theta}_t, && (17) \end{aligned}$$

where $t \in \mathbb{R}^+$ and $x \in \mathbb{R}^n$. Our aim is to determine the decay rates of this system. Now we have to asssume $m \neq 0$. The initial conditions are

$$u(x,0) = u_0(x), \quad u_t(x,0) = u_1(x), \quad \theta(x,0) = \theta_0(x), \quad \theta_t(x,0) = \theta_1(x)$$

for $x \in \mathbb{R}^n$. We assume the existence of smooth solutions which can be shown by using the Fourier transform. Application of the Fourier transform $(Fu(x,t))(\xi) =: v(\xi,t)$, $(F\theta(x,t))(\xi) =: w(\xi,t)$ implies

$$\begin{aligned} a\rho^2 v(\xi,t) - \rho w(\xi,t) + v_{tt}(\xi,t) &= 0, && (18)\\ \rho w(\xi,t) - m w(\xi,t) - d\rho\hat{v}_t(\xi,t) &= c\hat{w}_t(\xi,t), && (19) \end{aligned}$$

where $\rho := |\xi|^2$. It is easy to see that both v and w satisfy the following fourth-order ordinary differential equation

$$\begin{aligned} v_{tttt} + \frac{1}{\tau}v_{ttt} + \frac{1}{c\tau}\left((ac+d)\tau\rho^2 + m + \rho\right)v_{tt} \\ + \frac{1}{c\tau}\rho^2(ac+d)\,v_t + \frac{a}{c\tau}\rho^2(\rho+m)v = 0. \end{aligned} \qquad (20)$$

The characteristic polynomial of equation (20) is given by

$$P_\rho(\lambda) = \lambda^4 + \frac{1}{\tau}\lambda^3 + \frac{1}{c\tau}(\rho + m + \tau(ac+d)\rho^2)\lambda^2 + \frac{1}{c\tau}(ac+d)\rho^2\lambda + \frac{a}{c\tau}(\rho^3 + m\rho^2). \qquad (21)$$

First, we study the behaviour of the roots for $\rho \to 0$. By a straightforward analysis, we obtain the following asymptotic properties of the roots which give information about the decay rates.

Proposition 1. *The roots of the characteristic polynomial have the following properties for $\xi \to 0$ and $\rho := |\xi|^2$:*

$$\begin{aligned} \lambda_1 &= -\tfrac{d}{2m}\rho^2 + \mathcal{O}(\rho^3) + i\sqrt{a}\rho + i\mathcal{O}(\rho^2),\\ \lambda_2 &= -\tfrac{d}{2m}\rho^2 + \mathcal{O}(\rho^3) - i\sqrt{a}\rho + i\mathcal{O}(\rho^2),\\ \lambda_3 &= -\tfrac{1}{2\tau} + \tfrac{1}{2}\left(\tfrac{1}{\tau^2} - \tfrac{4m}{c\tau}\right)^{1/2} - \frac{1}{c\tau\left(\frac{1}{\tau^2} - \frac{4m}{c\tau}\right)^{1/2}}\rho + \mathcal{O}(\rho^2),\\ \lambda_4 &= -\tfrac{1}{2\tau} - \tfrac{1}{2}\left(\tfrac{1}{\tau^2} - \tfrac{4m}{c\tau}\right)^{1/2} + \frac{1}{c\tau\left(\frac{1}{\tau^2} - \frac{4m}{c\tau}\right)^{1/2}}\rho + \mathcal{O}(\rho^2). \end{aligned}$$

One can numerically verify these properties. One can also determine the asymptotic behaviour of the roots for $\rho \to \infty$. The following proposition describes the asymptotic behaviour of the roots for $\rho \to \infty$.

Proposition 2. *The roots of the characteristic polynomial have the following properties for $\xi \to \infty$ and $\rho := |\xi|^2$. There are positive constants $c_1, c_2, c_3, c_4 > 0$ such* *that*

$$\begin{aligned}\lambda_{1,2} &= -c_1\frac{1}{\rho}+\mathcal{O}\left(\frac{1}{\rho^2}\right)\pm c_2\rho i+i\mathcal{O}(\rho^2),\\ \lambda_{3,4} &= -c_3+\mathcal{O}\left(\frac{1}{\rho}\right)\pm c_4\sqrt{\rho}i+i\mathcal{O}(\rho).\end{aligned}$$

3.1. **Decay Rates for $m \neq 0$.** In this section C stands for a generic positive constant. We want to determine the decay rates for the homogenous system (1), (2) for the case $m \neq 0$. A fundamental system for equation (20) is given by

$$\left\{e^{\lambda_1 t}, e^{\lambda_2 t}, e^{\lambda_3 t}, e^{\lambda_4 t}\right\}.$$

Every solution of equation (20) has the form

$$v(t,\xi) = a_1(\xi)e^{\lambda_1 t}+a_2(\xi)e^{\lambda_2 t}+a_3(\xi)e^{\lambda_3 t}+a_4(\xi)e^{\lambda_4 t}.$$

We can explicitly determine the coefficients $a_j(\xi)$ as

$$a_j(\xi) = \sum_{k=0}^{3} b_j^k v_k \qquad \text{for} \quad j=1,2,3,4,$$

where b_j^k are given by

$$b_j^0 = \frac{-\prod\limits_{l\neq j}\lambda_l}{\prod\limits_{l\neq j}\lambda_j-\lambda_l}, \qquad b_j^1 = \frac{\sum\limits_{l\neq j\neq i}\lambda_i\lambda_l}{\prod\limits_{l\neq j}\lambda_j-\lambda_l},$$

$$b_j^2 = \frac{-\sum\limits_{l\neq j}\lambda_l}{\prod\limits_{l\neq j}\lambda_j-\lambda_l}, \qquad b_j^3 = \frac{1}{\prod\limits_{l\neq j}\lambda_j-\lambda_l}.$$

The following lemma describes the asymptotic behaviour of the coeffiecients $a_j(\xi)$.

Lemma 3.1. *$\exists C>0$, $\exists R_1>0$, $\forall \rho \leq R_1$:*

$$\begin{aligned}|a_3(\xi)e^{\lambda_3(\xi)t}| &\leq C\left(|v_0(\xi)|+|v_1(\xi)|+|v_2(\xi)|+|v_3(\xi)|\right)e^{\text{Re}\lambda_3 t},\\ |a_4(\xi)e^{\lambda_4(\xi)t}| &\leq C\left(|v_0(\xi)|+|v_1(\xi)|+|v_2(\xi)|+|v_3(\xi)|\right)e^{\text{Re}\lambda_4 t}.\end{aligned}$$

We can also estimate the first part of the solution for small ρ.

Lemma 3.2. *$\exists C>0$, $\exists R_1>0$, $\forall \rho \leq R_1$:*

$$\begin{aligned}&|a_1(\xi)e^{\lambda_1(\xi)t}+a_2(\xi)e^{\lambda_2(\xi)t}|\\ \leq\ & C\left(|v_0(\xi)|+C\frac{|\sin(\rho t)|}{\rho}\left(|v_1(\xi)|+|v_2(\xi)|+|v_3(\xi)|\right)\right)e^{\text{Re}\lambda_1 t}.\end{aligned}$$

As a corollary we obtain

Corollary 1. *$\exists C>0$, $\exists R_1>0$, $\forall \rho \leq R_1$:*

$$|a_i(\xi)e^{\lambda_i(\xi)t}| \leq C|v_0(\xi)|e^{\text{Re}\lambda_i t}+C\frac{1}{\rho}\left(|v_1(\xi)|+|v_2(\xi)|+|v_3(\xi)|\right)e^{\text{Re}\lambda_i t},$$

where $i=1,2$.

For large ρ, we have

Lemma 3.3. *$\exists R_2>0$, $\exists C>0$, $\forall \rho \geq R_2$:*

$$|a_i e^{\lambda_i t}| \leq Ce^{\text{Re}\lambda_i t}(|v_0|+|v_1|+|v_2|+|v_3|),$$

where $i=1,2,3,4$.

These results lead to the estimate

$$\begin{aligned}|v_t| &= |\lambda_1 a_1 e^{\lambda_1 t} + \lambda_2 a_2 e^{\lambda_2 t} + \lambda_3 a_3 e^{\lambda_3 t} + \lambda_4 a_4 e^{\lambda_4 t}| \\ &\leq |\lambda_1||a_1 e^{\lambda_1 t}| + |\lambda_2||a_2 e^{\lambda_2 t}| + |\lambda_3||a_3 e^{\lambda_3 t}| + |\lambda_4||a_4 e^{\lambda_4 t}|.\end{aligned}$$

Let $\tilde{u}$ be the Fourier transformed of v, yielding

$$\begin{aligned}|\tilde{u}_t| &= C|\int\limits_{\mathbb{R}^n} e^{ix\xi} v_t(\xi,t)d\xi| \\ &\leq C \int\limits_{|\xi|\leq R_1} |v_t(\xi,t)|d\xi + C \int\limits_{R_1\leq|\xi|\leq R_2} |v_t(\xi,t)|d\xi + C \int\limits_{|\xi|\geq R_2} |v_t(\xi,t)|d\xi\end{aligned}$$

Altogether, these results lead to the following estimate

$$\int\limits_{|\xi|\geqslant R_2} |v_t(\xi,t)|d\xi \leq C\left(||v_0||_{3n+3,\infty} + ||v_1||_{3n+3,\infty} + ||v_2||_{3n+3,\infty} + ||v_3||_{3n+3,\infty}\right) \times \\ \times \left(e^{-Ct} + t^{-n}\right).$$

We obtain for all $t \geq 0$:

$$|\tilde{u}_t| \leqslant C(1+t)^{-n/4}\left(||\tilde{u}_0||_{3n+3,1} + ||\tilde{u}_1||_{3n+3,1} + ||\tilde{u}_2||_{3n+3,1} + ||\tilde{u}_3||_{3n+3,1}\right).$$

We can also get an estimate for the solution in the L^2-norm since the system is dissipative. Now we can use interpolation techniques to describe the asymptotic behaviour of the solutions.

Theorem 3.4. *Let $2 \leqslant q \leqslant \infty$, $1/p + 1/q = 1$, $m \neq 0$, $N_p > (1-2/q)(3n+3)$. Then $\exists c = c(n,q) > 0$ $\forall V_0 \in W^{N_p,p}(\mathbb{R}^n)$ $\forall t \geq 0$:*

$$||V_t(t)||_q \leqslant c(1+t)^{-\frac{n}{4}(1-\frac{2}{q})}||V_0||_{N_p,p},$$

where $V(t) = (\hat{u}(t), \hat{u}_t(t), \theta(t), \theta_t(t))$ and (u,θ) the solutions of the system (16), (17).

It should be mentioned that $m \neq 0$ has been assumed. We will see that the decay rates differ from the case $m = 0$. Compared to the classical linear thermoelastic plate equations $\tau = 0$, i.e.,

$$\begin{aligned}a\Delta^2 u + \Delta\theta + u_{tt} &= 0, \\ c\theta_t - \Delta\theta - d\Delta u_t &= 0,\end{aligned}$$

we also observe different decay rates. Namely, the decay rates in the system of linear thermoelastic plate equation has the form $t^{-n/2}$. We remark that the decay rate $t^{-n/4}$ for the case $m = 0$ is not the decay rate of the classical plate equation

$$u_{tt} + \Delta^2 u = 0$$

having the rate of $t^{-n/2}$ corresponding to that of the heat equation. We compare the decay rates of the systems, i.e., we estimate the vector $V(t)$ in some time derivatives to the decay rate. In this case you can do this for first time derivative.

3.2. **Decay Rates for $m = 0$.** Letting $m = 0$, we observe that the homogenized system

$$a\Delta^2 u + \Delta\theta + u_{tt} = 0, \tag{22}$$
$$\Delta\theta - m\theta + d\Delta\hat{u}_t = c\hat{\theta}_t \tag{23}$$

can be rewritten as

$$a\Delta^2 u + \Delta\theta + u_{tt} = 0, \tag{24}$$
$$c\theta_t + \nabla' q - d\Delta u_t = 0, \tag{25}$$
$$\tau q_t + q + \nabla\theta = 0, \tag{26}$$

where q is the heat flux. The system is given by a coupling of the plate equation to the heat equation after Cattaneo's law. The last equation represents the Cattaneo's law. One can easily get for $\tau = 0$ and $m = 0$ the classical thermoelastic plate equations.

As mentioned before, the system changes its decay rates for $m = 0$, because the roots of the characteristic polynomial change their behaviour. We assume $a = c = d = 1$ and $m = 0$. By a similar analysis we obtain the following proposition.

Proposition 3. *There holds for $m = 0$, $\rho \to 0$:*

$$\begin{aligned}\lambda_1 &= \frac{1}{2}(c_1 - 1)\rho + \mathcal{O}(\rho^2) + \left(\mathcal{O}(\rho^2)\right)^{1/2},\\ \lambda_2 &= \frac{1}{2}(c_1 - 1)\rho + \mathcal{O}(\rho^2) - \left(\mathcal{O}(\rho^2)\right)^{1/2},\\ \lambda_3 &= -c_1\rho + \mathcal{O}(\rho^2),\\ \lambda_4 &= -\frac{1}{\tau} + \rho + \mathcal{O}(\rho^2),\end{aligned}$$

where

$$c_1 \approx 0,56984, \quad c_1 - 1 \approx -0,43015.$$

The behaviour of the roots for $\rho \to \infty$ does not change, so we have the same result as before for $\rho \to \infty$. Analogously, we can study the asymptotics of the system for $m = 0$. By a similar analysis of the solution as in the section before, we can obtain an estimate for the second time derivative of the solution. So we have the following theorem.

Theorem 3.5. *Let $2 \leqslant q \leqslant \infty$, $1/p + 1/q = 1$, $m = 0$, $N_p > (1 - 2/q)(3n + 5)$. Then $\exists c = c(n, q) > 0 \ \ \forall V_0 \in W^{N_p,p}(\mathbb{R}^n) \ \ \forall t \geq 0$:*

$$||V_{tt}(t)||_q \leqslant c(1 + t)^{-\frac{n}{2}(1-\frac{2}{q})}||V_0||_{N_p,p},$$

where $V(t) = (\hat{u}(t), \hat{u}_t(t), \theta(t), \theta_t(t))$.

One can get estimates for the vector $V(t)$ putting some conditions on the space dimension. These theorems are presented next without proofs which can be done analogously.

Theorem 3.6. *Let $2 \leqslant q \leqslant \infty$, $1/p + 1/q = 1$, $m \neq 0$, $N_p > (1 - 2/q)(3n + 1)$, $n \geqslant 3$. Then $\exists c = c(n, q) > 0 \ \ \forall V_0 \in W^{N_p,p}(\mathbb{R}^n) \ \ \forall t \geqslant 0$:*

$$||V(t)||_q \leqslant c(1 + t)^{-\frac{n-2}{4}(1-\frac{2}{q})}||V_0||_{N_p,p}.$$

Theorem 3.7. *Let* $2 \leqslant q \leqslant \infty$*,* $1/p + 1/q = 1$*,* $m = 0$*,* $N_p > (1 - 2/q)(3n + 1)$*,* $n \geqslant 5$*. Then* $\exists c = c(n,q) > 0 \quad \forall V_0 \in W^{N_p,p}(\mathbb{R}^n) \quad \forall t \geqslant 0$*:*

$$||V(t)||_q \leqslant c(1+t)^{-\frac{n-4}{2}(1-\frac{2}{q})}||V_0||_{N_p,p}.$$

REFERENCES

[1] A. I. Akhiezer and V.B. Beretetskii, *Quantum Electrodynamics*, Interscience Publishers, 1965, 23–25.

[2] E. A. S. Bahaa and T. C. Malvin, *Grundlagen der Photonik*, Wiley-VHC, 2008.

[3] J. E. Lagnese, J. L. Lions, *Modelling Analysis and Control of Thin Plates*, Masson, Paris, **RMA 6** (1988).

[4] H. W. Lord, Y. Shulman, *A generalized dynamical theory of thermoelasticity*, J. Mech. Phy. Solids, **15** (1967), 299–309.

[5] R. Racke, R. Quintanilla, *Qualitative aspects in resonators*, Arch. Mech., **60** (2008), 345–360.

[6] R. Racke, R. Quintanilla, *Addendum to: Qualitative aspects in resonators*, Arch. Mech., **63** (2011), 429–435.

[7] R. Racke, H. F. Sare *On the stability of damped Timoshenko systems - Cattaneo versus Fourier law*, Arch. Rational Mech. Anal., **194** (2009), 221–251.

[8] Y. Sun, D. Fang, A. K. Soh, *Thermoelastic damping in micro-beam resonators*, Int. J. Solids Structures, **43** (2006), 3213–3229.

E-mail address: Bugra.Kabil@mathematik.uni-stuttgart.de

ADAPTIVE MESH REFINEMENT FOR SPECTRAL WENO SCHEMES FOR EFFICIENT SIMULATION OF POLYDISPERSE SEDIMENTATION PROCESSES

Raimund Bürger

CI^{2}MA and Departamento de Ingeniería Matemática
Facultad de Ciencias Físicas y Matemáticas, Universidad de Concepción
Casilla 160-C, Concepción, Chile

Pep Mulet

Departament de Matemàtica Aplicada, Universitat de València
Universitat de València, Av. Dr. Moliner 50, E-46100 Burjassot, Spain

Luis M. Villada

CI^{2}MA and Departamento de Ingeniería Matemática
Facultad de Ciencias Físicas y Matemáticas, Universidad de Concepción
Casilla 160-C, Concepción, Chile

Abstract. The sedimentation of a polydisperse suspension with N particle size classes can be described by a system of N nonlinear scalar first-order conservation laws. For its numerical solution, Bürger et al. [J. Comput. Phys. **230**, 2322–2344 (2011)] proposed a spectral weighted essentially non-oscillatory (WENO) scheme based on a hyperbolicity analysis. It is demonstrated that this scheme becomes more efficient by the technique of Adaptive Mesh Refinement (AMR), which concentrates computational effort on zones of strong variation.

1. **Introduction.** The sedimentation of a polydisperse suspension of small rigid equal-density spheres of N size classes can be described by a spatially one-dimensional system of first-order nonlinear conservation laws

$$\partial_t \Phi + \partial_x \boldsymbol{f}(\Phi) = 0, \quad \Phi = (\phi_1, \dots, \phi_N)^{\mathrm{T}}, \quad \boldsymbol{f}(\Phi) = \bigl(f_1(\Phi), \dots, f_N(\Phi)\bigr)^{\mathrm{T}}, \qquad (1)$$

where $t > 0$ and $x \in I \subset \mathbb{R}$. The unknowns are the volume fractions (concentrations) ϕ_i of species i, $i = 1, \dots, N$, as functions of depth x and time t. The flux density functions are $f_i(\Phi) = \phi_i v_i(\Phi)$, where the settling velocities v_i are given functions of Φ. The model (1) is widely used in engineering and other applications, and a very similar model describes multi-class traffic flow. See [5, 7, 10] for references.

Typical solutions of (1) include discontinuities (kinematic shocks) separating areas of different composition. The accurate numerical approximation of these solutions is a challenge since closed-form eigenvalues and eigenvectors of the flux Jacobian $\mathcal{J}_{\boldsymbol{f}}(\Phi) = (f_{ij}(\Phi))_{1 \le i,j \le N} := (\partial f_i(\Phi)/\partial \phi_j)_{1 \le i,j \le N}$ are usually not available. Some of these sedimentation models, including the widely used one by Masliyah, Lockett and Bassoon (MLB model, cf. [5]), lead to flux Jacobians that can be analyzed by a convenient hyperbolicity criterion that has become known as the “secular equation” [1, 6]. When this approach applies, hyperbolicity can be ensured under easily verifiable conditions and the eigenstructure of the Jacobian can be computed numerically, so that efficient shock capturing schemes may be applied for (1).

Adaptive techniques, as the Adaptive Mesh Refinement (AMR) algorithm [4], aim to reduce the computational cost of these schemes, by using a higher resolution near shocks, heads and tails of rarefactions, etc., while employing a coarse mesh near smooth regions of the flow. We herein apply the AMR technique to a WENO scheme implemented in a component-wise fashion combined with global Lax-Friedrichs flux vector splitting (denoted by "COMP-GLF"), and alternatively, to a WENO scheme applied in a characteristic-wise (spectral) fashion (denoted by "SPEC-INT") [7]. The scheme COMP-GLF does not rely on characteristic information, is much easier to implement than SPEC-INT, and on a fixed uniform grid is several times faster than SPEC-INT. However, SPEC-INT is more accurate than COMP-GLF, and more efficient than COMP-GLF in terms of reduction of numerical error per CPU time [7]. It turns out that equipping both versions with AMR produces substantial gains in computational efficiency when compared with the corresponding non-adaptive version, and that the adaptive versions based on SPEC-INT are consistently more efficient than those relying on COMP-GLF.

Any kind of adaptativity that permits to restrict the use of a high-resolution scheme on a fine grid to a portion of the computational domain will make computations more efficient (cf., e.g., [8, 9, 17]). AMR is a grid adaptation technique, introduced by Berger and Oliger [4] for hyperbolic conservation laws, which is based not so much on the reduction of the number of cells on the grid as on the reduction of the overall number of applications of the integration algorithm. This algorithm in very time-consuming especially for high-resolution shock capturing schemes. The AMR algorithm is a two-fold adaptive method. The goal of allowing arbitrary grid resolution is attained by the definition of a set of overlapping grids of different resolutions –a grid hierarchy– being the grid at each resolution level defined only on the part of the domain that is foreseen to require such a resolution. The way in which the grids are overlapped allows to refine also in time, in the sense that each grid is integrated with temporal steps adapted to its spatial grid size. This time refinement further improves the overall performance of the algorithm [3, 4].

2. **Sedimentation of polydisperse suspensions.** The MLB model arises from the continuity and linear momentum balance equations for the solid species and the fluid through suitable constitutive assumptions and simplifications (cf. [5]). For particles that have the same density, the MLB velocities $v_1, \dots, v_N$ are given by

$$v_i(\Phi) := \frac{(\varrho_{\mathrm{s}} - \varrho_{\mathrm{f}}) g d_1^2}{18\mu_{\mathrm{f}}} (1-\phi) V(\phi) (\delta_i - \boldsymbol{\delta}^{\mathrm{T}} \Phi), \quad i = 1, \dots, N, \tag{2}$$

where $d_1 > d_2 > \cdots > d_N$ are the respective species diameters, $\delta_i := d_i^2/d_1^2$, $\boldsymbol{\delta} := (\delta_1, \dots, \delta_N)^{\mathrm{T}}$, ϱ_{s} and ϱ_{f} are the solid and fluid densities, g is the acceleration of gravity, μ_{f} is the fluid viscosity, $\phi := \phi_1 + \cdots + \phi_N$, and $V(\phi)$ is a hindered settling factor that should satisfy $V(0) = 1$, $V(\phi_{\max}) = 0$ and $V'(\phi) \leq 0$ for $\phi \in [0, \phi_{\max}]$, where $\phi_{\max}$ denotes the maximum total solids concentration. A standard choice is

$$V(\phi) = (1-\phi)^{n_{\mathrm{RZ}}-2} \quad \text{if } \Phi \in \mathcal{D}_{\phi_{\max}},\ n_{\mathrm{RZ}} > 2; \quad V(\phi) = 0 \quad \text{otherwise}, \tag{3}$$

where $\mathcal{D}_{\phi_{\max}} := \{\Phi \in \mathbb{R}^n \,|\, \phi_1 \geq 0, \dots, \phi_N \geq 0, \phi \leq \phi_{\max}\}$. The components $f_1(\Phi), \dots, f_N(\Phi)$ of the flux vector $\boldsymbol{f}(\Phi)$ of the MLB model are given by

$$f_i(\Phi) := v_1(\mathbf{0}) \phi_i (1-\phi) V(\phi) (\delta_i - \boldsymbol{\delta}^{\mathrm{T}} \Phi), \quad i = 1, \dots, N. \tag{4}$$

3. **Secular equation and hyperbolicity analysis.** For general kinematic models with $v_i = v_i(\phi_1, \dots, \phi_N)$, the spectral information of $\mathcal{J}_{\boldsymbol{f}}(\Phi)$ cannot be readily obtained. However, when $v_i = v_i(p_1, \dots, p_m)$ and $p_l = p_l(\Phi)$ for $i = 1, \dots, N$ and $l = 1, \dots, m \ll N$, then $\mathcal{J}_{\boldsymbol{f}}(\Phi)$ is a rank-m perturbation of $\boldsymbol{D} := \operatorname{diag}(v_1, \dots, v_N)$ of the form $\mathcal{J}_{\boldsymbol{f}} = \boldsymbol{D} + \boldsymbol{B}\boldsymbol{A}^{\mathrm{T}}$, where

$$\boldsymbol{B} := (B_{il}) = \left(\frac{\phi_i \partial v_i}{\partial p_l}\right), \ \boldsymbol{A} := (A_{jl}) = \left(\frac{\partial p_l}{\partial \phi_j}\right), \quad 1 \leq i, j \leq N, \ 1 \leq l \leq m. \tag{5}$$

The hyperbolicity analysis is then based on the following theorem.

Theorem 3.1 (The secular equation, [1, 10])**.** *Assume that $v_i > v_j$ for $i < j$, and that $\boldsymbol{A}$ and $\boldsymbol{B}$ have the formats specified in (5). Let $\lambda \neq v_i$ for $i = 1, \dots, N$. Then λ is an eigenvalue of $\boldsymbol{D} + \boldsymbol{B}\boldsymbol{A}^{\mathrm{T}}$ if and only if*

$$R(\lambda) := \det\bigl(\boldsymbol{I} + \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{D} - \lambda \boldsymbol{I})^{-1}\boldsymbol{B}\bigr) = 1 + \sum_{i=1}^{N} \frac{\gamma_i}{v_i - \lambda} = 0,$$

where γ_i can be effectively computed from v_i and determinants of submatrices of A, B. The relation $R(\lambda) = 0$ is known as the secular equation [1].

When $m \leq 2$, one may easily compute $\gamma_1, \dots, \gamma_N$, and the hyperbolicity analysis via Theorem 3.1 is less involved than discussing the zeros of $\det(\mathcal{J}_{\boldsymbol{f}}(\Phi) - \lambda \boldsymbol{I})$. For the MLB model with equal-density spheres, v_i depends on $p_1 := \phi$ and $p_2 := \boldsymbol{\delta}^{\mathrm{T}}\Phi$. Thus, $m = 2$ and $\gamma_i = -v_1(\boldsymbol{0})(n-1)(1-\phi)^{n-2}\phi_i\delta_i > 0$ if $\phi_i > 0$ and $\phi < 1$.

The proof of the following corollary follows from Theorem 3.1 by a discussion of the poles of $R(\lambda)$ and its asymptotic behavior as $\lambda \to \pm\infty$, see [6].

Corollary 1 ([6])**.** *With the notation of Theorem 3.1, assume that $\gamma_i \cdot \gamma_j > 0$ for $i, j = 1, \dots, N$. Then $\boldsymbol{D} + \boldsymbol{B}\boldsymbol{A}^{\mathrm{T}}$ is diagonalizable with real eigenvalues $\lambda_1, \dots, \lambda_N$. If $\gamma_1, \dots, \gamma_N < 0$, the following so-called* interlacing property *holds:*

$$v_{N+1} := v_N + \gamma_1 + \dots + \gamma_N < \lambda_N < v_N < \lambda_{N-1} < \dots < v_2 < \lambda_1 < v_1. \tag{6}$$

As a consequence, we see that the model (1) with the flux vector $\boldsymbol{f}(\Phi)$ of the MLB model given by (4) is strictly hyperbolic if $\phi_1 > 0, \dots, \phi_N > 0$ and $\phi < \phi_{\max} < 1$.

4. **SPEC-INT and COMP-GLF schemes.** For grid points $x_i = (i + 1/2)\Delta x$, $t_n = n\Delta t$, a conservative scheme for $\Phi_i^n \approx \Phi(x_i, t_n)$ is given by

$$\Phi_i^{n+1} = \Phi_i^n - \frac{\Delta t}{\Delta x}\bigl(\hat{\boldsymbol{f}}_{i+1/2} - \hat{\boldsymbol{f}}_{i-1/2}\bigr), \quad \hat{\boldsymbol{f}}_{i+1/2} = \hat{\boldsymbol{f}}\left(\Phi_{i-s+1}^n, \dots, \Phi_{i+s}^n\right)$$

for $i = 0, \dots, M-1$ along with $\hat{\boldsymbol{f}}_{-1/2} = \hat{\boldsymbol{f}}_{M-1/2} = \boldsymbol{0}$ (zero-flux boundary conditions). The key point is the design of the numerical flux $\hat{\boldsymbol{f}}_{i+1/2}$ so that the resulting scheme is (at least formally second-order) accurate and stable. The most common approach for this task is to solve Riemann problems, either exactly or approximately. For polydisperse sedimentation, exact Riemann solvers are out of reach, since the eigenstructure of the flux Jacobian is hard to compute.

In [7] we used Shu-Osher's technique [16] along with the information provided by the secular equation to get efficient schemes for polydisperse sedimentation models. We here briefly describe this scheme, which is based on applying the third-order TVD Runge-Kutta method of [16] to spatially semi-discretized equations. For the discretization of the flux derivative we use local characteristic projections. Local characteristic information to compute $\hat{\boldsymbol{f}}_{i+1/2}$ is provided by the eigenstructure of

$\mathcal{J}_{\boldsymbol{f}}(\Phi_{i+1/2})$, where $\Phi_{i+1/2} := \frac{1}{2}(\Phi_i + \Phi_{i+1})$, given by the right and left eigenvectors, $\boldsymbol{r}_{i+1/2,j}$ and $\boldsymbol{l}_{i+1/2,j}$, respectively, that form the respective matrices

$$\boldsymbol{R}_{i+1/2} = \begin{bmatrix} \boldsymbol{r}_{i+1/2,1} & \dots & \boldsymbol{r}_{i+1/2,N} \end{bmatrix}, \quad \bigl(\boldsymbol{R}^{-1}_{i+1/2}\bigr)^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{l}_{i+1/2,1} & \dots & \boldsymbol{l}_{i+1/2,N} \end{bmatrix}.$$

From a local flux-splitting $\boldsymbol{f}^{\pm,k}$ (we omit its dependency on $i+1/2$) given by $\boldsymbol{f}^{-,k} + \boldsymbol{f}^{+,k} = \boldsymbol{f}$, where $\pm\lambda_k(\mathcal{J}_{\boldsymbol{f}^{\pm,k}}(\Phi)) \geq 0$, $\Phi \approx \Phi_{i+1/2}$ and λ_k is the k-th eigenvalue, we can define the k-th characteristic flux as $g_j^{\pm,k} = \boldsymbol{l}^{\mathrm{T}}_{i+1/2,k} \cdot \boldsymbol{f}^{\pm,k}(\Phi_j)$, $k = 1, \dots, N$. If $\mathcal{R}^+$ and $\mathcal{R}^-$ denote upwind-biased reconstructions (in our experiments we use the fifth-order WENO method introduced in [13]), then

$$\begin{aligned} \hat{g}_{i+1/2,k} &= \mathcal{R}^+\bigl(g^{+,k}_{i-s+1}, \dots, g^{+,k}_{i+s-1}; x_{i+1/2}\bigr) + \mathcal{R}^-\bigl(g^{-,k}_{i-s+2}, \dots, g^{-,k}_{i+s}; x_{i+1/2}\bigr), \\ \hat{\boldsymbol{f}}_{i+1/2} &= \boldsymbol{R}_{i+1/2}\hat{\boldsymbol{g}}_{i+1/2} = \hat{g}_{i+1/2,1}\boldsymbol{r}_{i+1/2,1} + \dots + \hat{g}_{i+1/2,n}\boldsymbol{r}_{i+1/2,n}. \end{aligned}$$

If we do not want to use local characteristic information, we can use the previous formula with $\boldsymbol{R}_{i+1/2} = \boldsymbol{I}_N$, where $\boldsymbol{I}_N$ denotes the $N \times N$ identity matrix, and a global flux splitting $\boldsymbol{f}^- + \boldsymbol{f}^+ = \boldsymbol{f}$, where $\pm\lambda_k(\mathcal{J}_{\boldsymbol{f}^\pm}(\Phi)) \geq 0$ for all k. With this choice, and denoting by $\boldsymbol{e}_k$ the kth unit vector, we get $g_j^{\pm,k} = \boldsymbol{e}_k^{\mathrm{T}}\boldsymbol{f}^\pm(\Phi_j) = f_k^\pm(\Phi_j)$, i.e., $g_j^{\pm,k}$ are the components of the split fluxes, and the numerical flux is computed component by component by reconstructing the split fluxes component by component, i.e., $\hat{\boldsymbol{f}}_{i+1/2} = (\hat{f}_{i+1/2,1}, \dots, \hat{f}_{i+1/2,N})^{\mathrm{T}}$, where

$$\hat{f}_{i+1/2,k} = \mathcal{R}^+\bigl(g^{+,k}_{i-s+1}, \dots, g^{+,k}_{i+s-1}; x_{i+1/2}\bigr) + \mathcal{R}^-\bigl(g^{-,k}_{i-s+2}, \dots, g^{-,k}_{i+s}; x_{i+1/2}\bigr)$$

for $k = 1, \dots, N$. This scheme will be referred to as COMP-GLF and it is a high-order extension of the Lax-Friedrichs scheme.

We now explain the SPEC-INT scheme. If $\lambda_k(\mathcal{J}_{\boldsymbol{f}}(\Phi)) > 0$ (respectively, < 0) for all $\Phi \in \Gamma_i := [\Phi_i, \Phi_{i+1}]$, where $\Gamma_i \subset \mathbb{R}^N$ denotes the segment joining both states, then we upwind (since then there is no need for flux splitting):

$$\boldsymbol{f}^{+,k} = \boldsymbol{f},\ \boldsymbol{f}^{-,k} = 0 \quad \text{if } \lambda_k(\mathcal{J}_{\boldsymbol{f}}(\Phi)) > 0, \quad \boldsymbol{f}^{+,k} = 0,\ \boldsymbol{f}^{-,k} = \boldsymbol{f} \quad \text{if } \lambda_k(\mathcal{J}_{\boldsymbol{f}}(\Phi)) < 0.$$

On the other hand, if $\lambda_k(\mathcal{J}_{\boldsymbol{f}}(\Phi))$ changes sign on Γ_i, then we use a Local Lax-Friedrichs flux splitting given by $\boldsymbol{f}^{\pm,k}(\Phi) = \boldsymbol{f}(\Phi) \pm \alpha_k\Phi$, where the numerical viscosity parameter α_k should satisfy $\alpha_k \geq \max_{\Phi\in\Gamma_i} |\lambda_k(\mathcal{J}_{\boldsymbol{f}}(\Phi))|$. The usual choice of the numerical viscosity $\alpha_k = \max\{|\lambda_k(\mathcal{J}_{\boldsymbol{f}}(\Phi_i))|, |\lambda_k(\mathcal{J}_{\boldsymbol{f}}(\Phi_{i+1}))|\}$ produces oscillations in the numerical solution indicating that the amount of numerical viscosity is insufficient. Usually $\max_{\Phi\in\Gamma_i} |\lambda_k(\mathcal{J}_{\boldsymbol{f}}(\Phi))|$ cannot be evaluated exactly. However, in the case of the MLB model, we have $\gamma_k < 0$ (see [6, 10]) and may employ the interlacing property (6) to obtain efficiently computable bounds

$$\max_{\Phi\in\Gamma_i} |\lambda_k(\Phi)| \leq \alpha_k := \max\Bigl\{\max_{\Phi\in\Gamma_i} \bigl|v_k(\Phi)\bigr|, \max_{\Phi\in\Gamma_i} \bigl|v_{k+1}(\Phi)\bigr|\Bigr\}, \quad k = 1, \dots, N. \tag{7}$$

(This property also holds for other models, under appropriate circumstances [6].) We denote by "SPEC-INT" the scheme for which $\alpha_1, \dots, \alpha_N$ are defined by (7).

5. **Adaptive Mesh Refinement (AMR).** We now outline the main building blocks of the AMR algorithm and refer to [2] for details. We denote by $G_0, \dots, G_L$ a 1D grid hierarchy composed of $L+1$ grids, such that, except for the coarsest grid G_0, cells of a given grid are obtained by the subdivision of cells of the immediately coarser grid into r parts (we assume $r = 2$). The unit interval is thus divided into $N_0, \dots, N_L$ subintervals of length $h_l = 1/N_l$, with $N_l = 2^l N_0$, $l = 0, \dots, L$, whose centers will be denoted by $x_j^l = (j + 1/2)h_l$, $j = 0, \dots, N_l - 1$, $l = 0, \dots, L$. A "mesh" G_l at resolution level l is just a subset of the index set $\{0, \dots, N_l - 1\}$

whose "extent", the union of the cells indexed by elements of G_l, is denoted by $\Omega_l(G_l)$. We consider only "nested" grid hierarchies, i.e., $\Omega_l(G_l) \subseteq \Omega_{l-1}(G_{l-1})$ for $1 \leq l \leq L$ is assumed to hold along with $\Omega_0(G_0) = \Omega$.

The meshes will be dynamically updated so that they adapt to the features of the solution, and we denote by $G_l^{t_l}$ the mesh that corresponds to the resolution level l and time t_l. Over each mesh we consider a numerical solution defined by a discrete function $\Phi_l^{t_l} = (\Phi_{l,j}^{t_l})$, with $\Phi_{l,j}^{t_l} \approx \Phi(x_j^l, t_l)$ and $j \in G_l^{t_l}$.

The algorithm can be described by the time evolution of the meshes and their associated numerical solutions, starting with $t_l = 0$, $l = 0, \dots, L$ and ending at $t_l = T$, $l = 0, \dots, L$, for some $T > 0$. The main building blocks of the AMR algorithm — integration and adaptation of the grids and projection from fine to coarse grids — are described next. The time step Δt_0 to integrate G_0 is selected to comply with a CFL condition that takes into account the maximal characteristic speed, computed from the spectral radius (or a estimate of it in case of the COMP-GLF scheme) of all Jacobian matrices $\mathcal{J}_{\boldsymbol{f}}(\Phi)$ appearing in the grid hierarchy. The time steps for the rest of the grids are taken by $\Delta t_l = \Delta t_{l-1}/2$ for $l = 1, \dots, L$, so the equivalent CFL condition holds for each grid. The grids are integrated according to the order dictated by the following condition: $t_{l'} \leq t_l \leq t_{l'} + \Delta t_l$ if $l \leq l'$. At some step of this time evolution, $(\Phi_l^{t_l + k\Delta t_l}, G_l^{t_l})$, $k = 1, 2$, are sequentially computed from $(\Phi_l^{t_l}, G_l^{t_l})$, supplemented by boundary conditions at a band surrounding $\Omega_l(G_l^{t_l})$ obtained by MUSCL-type interpolation from $(\Phi_{l-1}^{t_l}, G_{l-1}^{t_l})$ and $(\Phi_{l-1}^{t_l+2\Delta t_l}, G_{l-1}^{t_l})$, which must have been computed in previous steps. Once $(\Phi_l^{t_l+2\Delta t_l}, G_l^{t_l})$ is computed, there is data that overlay $\Omega_l(G_l^{t_l})$ at different resolution levels. A suitable projection of the data at the fine resolution level should be applied to modify the values $\Phi_{l-1,j}^{t_l+2\Delta t_l}$ of the immediately coarser grid function that correspond to cells overlaid by cells at $G_l^{t_l}$ and adjacent to them as well. This can be achieved by modifying the coarse numerical fluxes so that discrete conservation is maintained.

The update of the grids is performed by marking some cells to be refined following the following criteria: Let $\mathcal{I}(\Phi_{l-1}^t, x)$ be an interpolation operator defined on the data $\Phi_{l-1}^t = \{\Phi_{l-1,i}^t\}_{i \in G_{l-1}^t}$, then the cell centered at $x_{j/2}^{l-1}$ is selected for refinement if $|\Phi_{l,j}^t - \mathcal{I}(\Phi_{l-1}^t, x_j^l)| > \tau_p \cdot \max_{l,j} |\Phi_{l,j}^t - \mathcal{I}(\Phi_{l-1}^t, x_j^l)|$, where τ_p is a given tolerance. Further, we also include a cell in the refinement list if the modulus of the discrete gradient, computed in the coarser grid, exceeds some large threshold, so that shock formation can be detected from steepened data.

Once the cells that will compose the refined grid have been selected we add a certain number of extra cells forming a band around each marked cell to ensure that the cells adjacent to a singularity are refined. This device of creating "safety points" follows the spirit of [12, 14, 15]. These extra cells will avoid singularities to escape from the fine grid during one coarse time step and provide smooth data for accurate interpolation to the next finer grid. This refinement procedure is performed from fine to coarse resolution levels to ensure that $\Omega_l(G_l^t) \subseteq \Omega_{l-1}(G_{l-1}^t)$ for all t.

Once the new grid $\hat{G}_l$ is computed such that $\Omega_l(\hat{G}_l^t) \subseteq \Omega_{l-1}(G_{l-1}^t)$, one sets

$$\hat{\Phi}_{l,j}^t = \mathcal{I}(\Phi_{l-1}^t, x_j^l) \quad \text{if } j \in \hat{G}_l^t \setminus G_l^t, \quad \hat{\Phi}_{l,j}^t = \Phi_{l,j}^t \quad \text{if } j \in G_l^t,$$

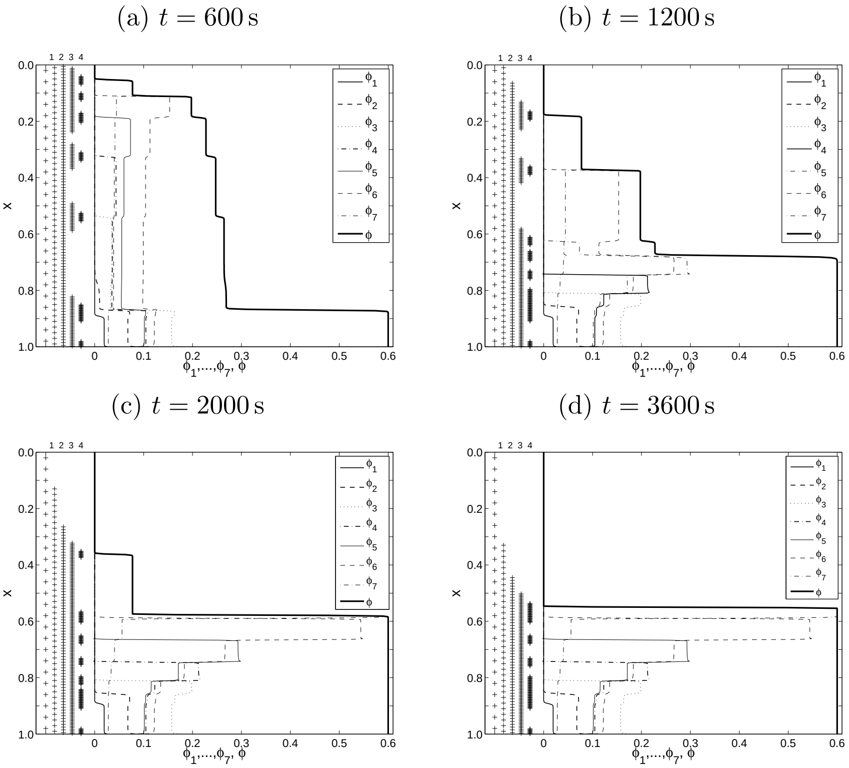
i.e., the value at the j-th cell is interpolated from data at the next coarser level for cells not in G_l^t. The refined grid is therefore defined by $(\hat{G}_l^t, \hat{\Phi}_l^t)$. Discrete boundary conditions are also applied if the grid overlaps the domain boundary.

6. Numerical example.

This example is based on experimental data from [11], where a suspension in a column of height $h = 0.227\,\mathrm{m}$ is considered and which

i	1	2	3	4	5	6	7
ϕ_i^0 $[10^{-2}]$	0.21285	0.99351	3.21012	3.48984	5.43924	9.80982	3.84462
d_i $[10^{-5}\,\mathrm{m}]$	280	240	200	150	110	80	40
δ_i	1.000000	0.734693	0.510204	0.286989	0.154336	0.081632	0.020481

TABLE 1. Particle sizes d_i and normalized squared particle sizes δ_i.


(a) $t = 600\,\mathrm{s}$

(b) $t = 1200\,\mathrm{s}$

(c) $t = 2000\,\mathrm{s}$

(d) $t = 3600\,\mathrm{s}$

FIGURE 1. Numerical solutions obtained with SPEC-INT-AMR with $L+1 = 5$ levels with coarsest grid of $N_0 = 50$ subintervals.

is characterized by (2) with $N = 7$, $\varrho_{\mathrm{s}} = 2790\,\mathrm{kg/m^3}$, $\varrho_{\mathrm{f}} = 1208\,\mathrm{kg/m^3}$, $\mu_{\mathrm{f}} = 0.02416\,\mathrm{Pa\,s}$, $g = 9.8\,\mathrm{m/s^2}$. Initial concentrations ϕ_i^0, the diameters d_i and normalized diameters $\delta_i = d_i/d_1$ are given in Table 1. The maximum total concentration is $\phi_{\max} = 0.6$ and $V(\phi)$ is given by (3) with the exponent $n_{\mathrm{RZ}} = 5$.

We simulate the process until the phenomenon enters in a steady state. Figure 1 shows the numerical solution obtained by SPEC-INT-AMR at four different times together with the corresponding grid hierarchy. We have used $L+1 = 5$ levels with a coarsest grid of $N_0 = 50$ points so that results are comparable with those for a fixed grid of $N_4 = 800$ points. We have tested for different choices for the threshold value τ_p and observed that $\tau_p = 10^{-2}$ is the most efficient choice. The plotted positions indicate that the adaptive mesh refinement technique works correctly, in the sense that the scheme detects the shock formation and refines these areas.

In Table 2, we present the percentages of storage space, number of integrations and CPU time required by AMR with respect to schemes on the uniform finest mesh.

	SPEC-INT-AMR at $t = 600\,\text{s}$			SPEC-INT-AMR at $t = 2000\,\text{s}$		
N_L	%Int's	%Memory	%CPU time	%Int's	%Memory	%CPU time [s]
800	30.33	27.28	25.54	27.14	27.96	28.11
1600	17.22	15.33	14.02	15.20	15.43	14.98
3200	8.88	8.21	7.54	7.88	8.23	8.05
6400	4.67	4.43	4.00	4.14	4.63	4.18

Table 2. Percentage of storage space (memory), number of integrations and CPU time of the adaptive algorithm with respect to the fixed grid algorithm with $\tau_p = 10^{-2}$ at two simulated times t for a hierarchy of $L+1 = 5$ levels and four different values of N_0.

SPEC-INT							SPEC-INT-AMR						
	$t = 600\,\text{s}$			$t = 2000\,\text{s}$				$t = 600\,\text{s}$			$t = 2000\,\text{s}$		
M	error	cr	cpu [s]	error	cr	cpu [s]	M	error	cr	cpu [s]	error	cr	cpu [s]
400	221.5	—	141.9	404.4	—	337.7	25	241.9	–	58.6	424.5	–	161.4
800	110.1	1.00	564.6	204.5	0.98	1316.9	50	130.3	0.89	144.2	217.1	0.96	370.2
1600	49.10	1.16	2263.9	69.6	1.55	5280.8	100	63.1	1.04	317.3	84.8	1.35	791.0
3200	25.70	0.93	8942.3	36.6	0.92	20859.2	200	28.7	1.13	674.5	40.3	1.07	1678.6
6400	12.81	1.00	36216.5	18.3	1.00	85522.9	400	14.4	0.99	1450.1	20.0	1.00	3574.8

Table 3. Total approximate L^1 errors ($\times 10^{-5}$), convergence rates and CPU times at two different times for SPEC-INT on a fixed grid and SPEC-INT-AMR with $L+1 = 5$ levels of refinament.

The indicated percentages represent the average memory load over all iterations. The values of Table 2 correspond to coarsest grids of $N_0 = 50$, 100, 200 and 400 subintervals and $L+1 = 5$ levels of refinement. CPU times and the percentages of memory allocated by SPEC-INT-AMR decrease as N_0 increases, as expected.

Table 3 and Figure 2 show approximate L^1 errors and CPU times at two different times for the method SPEC-INT-AMR using a grid hierarchy for different levels, corresponding to $N_0 = 25$, 50, 100, 200 and 400, and for the method SPEC-INT using a fixed uniform grid corresponding to $N_0 = 400$, 800, 1600, 3200 and 6400.

For a fixed L^1 error, the CPU time is smaller for the AMR technique than for the equivalent fixed-grid computation. In many cases the AMR technique is around twice faster at least for short simulated times.

Acknowledgments. RB is supported by Fondecyt project 1130154, Basal project PFB03 CMM, U. de Chile and Centro de Investigación en Ingeniería Matemática (CI^{2}MA), U. de Concepción, and CONICYT project Anillo ACT1118 (ANANUM). PM is supported by Spanish MICINN projects MTM 2008-00974 and MTM 2011-22741. LMV is supported by MECESUP project UCO0713.

REFERENCES

[1] J. Anderson, *A secular equation for the eigenvalues of a diagonal matrix perturbation*, Lin. Alg. Appl., **246** (1996), 49–70.

[2] A. Baeza, A. Martínez-Gavara and P. Mulet, *Adaptation based on interpolation errors for high order mesh refinement methods applied to conservation laws*, Appl. Numer. Math., **62** (2012), 278–296.

[3] M.J. Berger and P. Colella, *Local adaptive mesh refinement for shock hydrodynamics*, J. Comput. Phys., **82** (1989), 64–84.

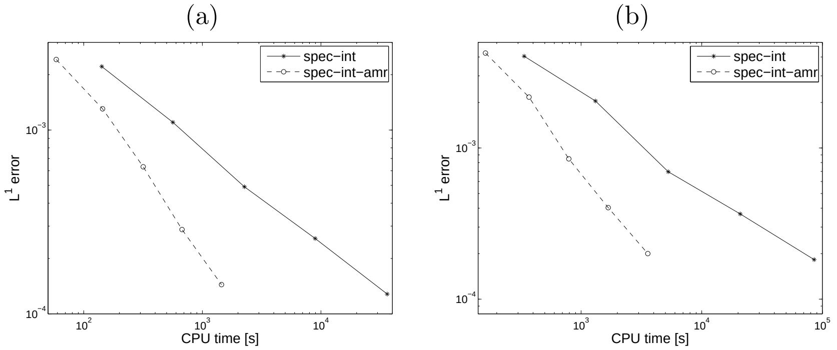


FIGURE 2. Approximate L^1 errors versus CPU time for SPEC-INT-AMR at simulated times (a) at $t = 600\,\mathrm{s}$, (b) $t = 2000\,\mathrm{s}$. .

[4] M.J. Berger and J. Oliger, *Adaptive mesh refinement for hyperbolic partial differential equations*, J. Comput. Phys., **53** (1984), 484–512.

[5] S. Berres, R. Bürger, K.H. Karlsen and E.M. Tory, *Strongly degenerate parabolic-hyperbolic systems modeling polydisperse sedimentation with compression*, SIAM J. Appl. Math., **64** (2003), 41–80.

[6] R. Bürger, R. Donat, P. Mulet and C.A. Vega, *Hyperbolicity analysis of polydisperse sedimentation models via a secular equation for the flux Jacobian*, SIAM J. Appl. Math., **70** (2010), 2186–2213.

[7] R. Bürger, R. Donat, P. Mulet and C.A. Vega, *On the implementation of WENO schemes for a class of polydisperse sedimentation models*, J. Comput. Phys., **230** (2011), 2322–2344.

[8] G. Chiavassa and R. Donat, *Point-value multiscale algorithms for 2D compressible flows*, SIAM J. Sci. Comput., **20** (2001), 805–823.

[9] A. Cohen, S. Kaber, S. Müller and M. Postel, *Fully adaptive finite volume schemes for conservation laws*, Math. Comp., **72** (2003), 183–225.

[10] R. Donat and P. Mulet, *A secular equation for the Jacobian matrix of certain multi-species kinematic flow models*, Numer. Methods Partial Differential Equations, **26** (2010), 159–175.

[11] R.M. Dorrell, A.J. Hogg, E.J. Sumner and P.J. Talling, *The structure of the deposit produced by sedimentation of polydisperse suspensions*, J. Geophys. Res., **116** (2011), paper F01024.

[12] A. Harten, *Multiresolution algorithms for the numerical solution of hyperbolic conservation laws*, Comm. Pure Appl. Math., **48** (1995), 1305–1342.

[13] G.S. Jiang and C.-W. Shu, *Efficient implementation of weighted ENO schemes*, J. Comput. Phys., **126** (1996), 202–228.

[14] J. Liandrat, *The wavelet transform; some applications to fluid dynamics and turbulence*, Eur. J. Mech. B/Fluids, **9** (1990), 1–19.

[15] J.J. Quirk, "An Adaptive Grid Algorithm for Computational Shock Hydrodynamics," PhD Thesis, Cranfield Institute of Technology, 1991.

[16] C.-W. Shu and S. Osher, *Efficient implementation of essentially non-oscillatory shock-capturing schemes*, J. Comput. Phys., **77** (1988), 439–471.

[17] H. Tang and T. Tang, *Adaptive mesh methods for one- and two-dimensional hyperbolic conservation laws*, SIAM J. Numer. Anal., **41** (2004), 487–515.

E-mail address: rburger@ing-mat.udec.cl
E-mail address: mulet@uv.es
E-mail address: lmvillada@ing-mat.udec.cl

A NONLINEAR MOVING-BOUNDARY PROBLEM OF PARABOLIC-HYPERBOLIC-HYPERBOLIC TYPE ARISING IN FLUID-MULTI-LAYERED STRUCTURE INTERACTION PROBLEMS

Sunčica Čanić

University of Houston, 4800 Calhoun Rd. Houston
TX 77204, USA

Boris Muha

University of Zagreb, Bijenička 30
10000 Zagreb, Croatia

Abstract. Motivated by modeling blood flow in human arteries, we study a fluid-structure interaction problem in which the structure is composed of multiple layers, each with possibly different mechanical characteristics and thickness. In the problem presented in this manuscript the structure is composed of two layers: a thin layer modeled by the 1D wave equation, and a thick layer modeled by the 2D equations of linear elasticity. The flow of an incompressible, viscous fluid is modeled by the Navier-Stokes equations. The thin structure is in contact with the fluid thereby serving as a fluid-structure interface with mass. The coupling between the fluid and the structure is nonlinear. The resulting problem is a nonlinear, moving-boundary problem of parabolic-hyperbolic-hyperbolic type. We show that the model problem has a well-defined energy, and that the energy is bounded by the work done by the inlet and outlet dynamic pressure data. The spaces of weak solutions reveal that the presence of a thin fluid-structure interface with mass regularizes solutions of the coupled problem. This opens up a new area withing the field of fluid-structure interaction problems, possibly revealing properties of FSI solutions that have not been studied before.

1. **Motivation.**

Fluid-structure interaction (FSI) problems arise in many applications. They include multi-physics problems in engineering such as aeroelasticity and propeller turbines, as well as biofluidic application such as self-propulsion organisms, fluid-cell interactions, and the interaction between blood flow and cardiovascular tissue. In biofluidic applications, such as the interaction between blood flow and cardiovascular tissue, the density of the structure (arterial walls) is roughly equal to the density of the fluid (blood). In such problems the energy exchange between the fluid and the structure is significant, leading to a highly nonlinear FSI coupling which is responsible for the instabilities in loosely coupled partitioned algorithms [3]. Despite a significant progress within the past decade [1, 2, 7, 8, 10, 13, 14, 6, 5, 4, 11, 15],

2000 *Mathematics Subject Classification.* Primary: 74F10, 35Q30, 76D05; Secondary: 74B10, 74K25.

Key words and phrases. Nonlinear moving-boundary problem, fluid-structure interaction.

The first author was supported by NSF grants DMS-1109189, DMS-0806941, and by the Texas Higher Education Board ARP grant 003652-0023-2009. The second author's Post-doctoral position was supported by the Texas Higher Education Board ARP grant 003652-0023-2009.

a comprehensive study of these problems remains to be a challenge due to their strong nonlinearity and multi-physics nature. In the blood flow application, the problems are further exacerbated by the fact that arterial walls of major arteries are composed of several layers, each with different mechanical characteristics. The main layers are the tunica intima, the tunica media, and the tunica adventitia. They are separated by the thin elastic laminae, see Figure 1, left. To this date, there are no fluid-structure interaction models or computational solvers in hemodynamics that take into account the multi-layered structure of arterial walls. In this manuscript we take a first step in this direction by proposing a benchmark problem in fluid-multi-layered-structure interaction. The proposed problem is a nonlinear moving-boundary problem of parabolic-hyperbolic-hyperbolic type for which the questions of well-posedness and numerical simulation are wide open. This opens up a new area within the field of FSI problems, in which the structure is composed of multiple layers, each with possibly different mechanical characteristics and thickness.

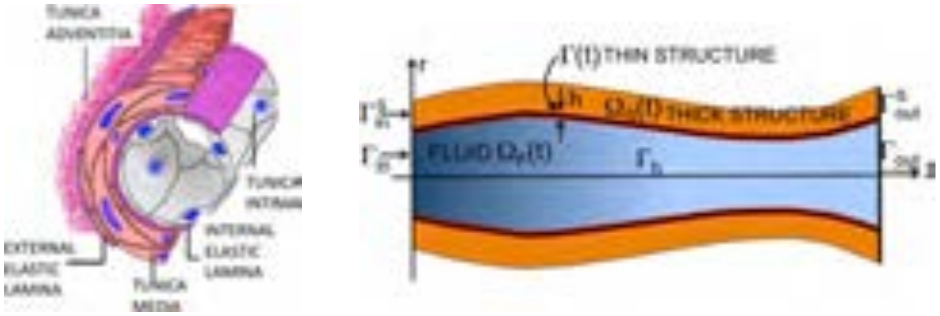


Figure 1. Left: Arterial wall structure. Right: Domain sketch.

2. **The benchmark problem.** We study a FSI problem in which the structure consists of two layers: a "thin" structural layer (modeled, e.g., by the linearly elastic Koiter shell equations), and a "thick" layer (modeled, e.g., by the equations of 2D/3D elasticity). To simplify matters, we will be assuming that the elastodynamics of the thin structure is modeled by the 1D linear wave equation. The wave equation model retains the main difficulties associated with the study of solutions to the more general elastodynamics models mentioned above. The thin structural layer is in contact with the flow of an incompressible, viscous fluid, modeled by the Navier-Stokes equations. From an application point of view, it is of interest to study this fluid-multi-structure interaction problem on a cylindrical domain, with the flow driven by the time-dependent dynamic pressure data, see Figure 1, right. The Navier-Stokes equations are defined in a time-dependent fluid domain $\Omega_F(t)$, which is not known *a priori*:

$$\textbf{FLUID}: \qquad \left.\begin{array}{rcl} \rho_F(\partial_t \mathbf{u} + \mathbf{u}\cdot\nabla\mathbf{u}) & = & \nabla\cdot\boldsymbol{\sigma}, \\ \nabla\cdot\mathbf{u} & = & 0, \end{array}\right\} \text{ in } \Omega_F(t),\ t\in(0,T), \qquad (1)$$

where ρ_F denotes the fluid density; $\mathbf{u}$ the fluid velocity; $\boldsymbol{\sigma} = -p\mathbf{I} + 2\mu_F\mathbf{D}(\mathbf{u})$ is the fluid Cauchy stress tensor; p is the fluid pressure; μ_F is the dynamic viscosity coefficient; and $\mathbf{D}(\mathbf{u}) = \frac{1}{2}(\nabla\mathbf{u} + \nabla^T\mathbf{u})$ is the symmetrized gradient of $\mathbf{u}$.

We assume that the reference fluid domain is a cylinder of radius R and length L, denoted by Ω_F, with the lateral boundary denoted by Γ. To fix ideas, consider the fluid domain to be a subset of $\mathbb{R}^2$ with z and r denoting the axial (horizontal) and radial (vertical) coordinates. The cylinder wall is assumed to be compliant and consisting of two layers: a thin layer, whose location at time t is denoted by $\Gamma(t)$, and a thick structural layer, whose location at time t is denoted by $\Omega_S(t)$, as shown in Figure 1, right. The **thin layer** $\Gamma(t)$ is modeled by the 1D linear wave equation

$$\textbf{THIN STRUCTURE}: \quad m\partial_{tt}\eta = T\partial_{zz}\eta + f, \quad \text{on } \Gamma \times (0,T). \tag{2}$$

Here $\eta := \eta(t,z)$ denotes the radial (transverse) displacement from the reference position $\Gamma = \{(z,R)|z \in (0,L)\}$, f is the source term (the radial component), m is mass per unit length, and T is tension. The elastodynamics of the **thick structural layer** is governed by the 2D equations of linear elasticity:

$$\textbf{THICK STRUCTURE}: \quad \rho\partial_{tt}\mathbf{d} = \nabla \cdot \mathbf{S} \quad \text{in } \Omega_S,\ t \in (0,T). \tag{3}$$

Here $\mathbf{d} := (d_r(t,z,r), d_z(t,z,r))$ describes the displacement of a thick elastic structure with respect to a fixed, reference configuration Ω_S, and $\mathbf{S}$ is the first Piola-Kirchoff stress tensor $\mathbf{S} = 2\mu\mathbf{D}(\mathbf{d}) + \lambda(\nabla \cdot \mathbf{d})\mathbf{I}$, with the Lamé constants λ and μ, where $\mathbf{D}(\mathbf{d})$ is the symmetrized gradient of $\mathbf{d}$, and ρ is the mass density.

To capture a full **two-way coupling** between the fluid and the structure, and between the two structural layers, two sets of boundary conditions need to be prescribed: the kinematic and dynamic coupling conditions. The kinematic condition provides information about the kinematic quantities, such as velocity. We adopt the no-slip condition requiring continuity of velocities at both the fluid-structure interface and at the structure-structure interface. The dynamic coupling condition, on the other hand, describes the second Newton's Law of motion. This condition states that the rate of change of (radial) momentum $\partial_{tt}\eta$ of the interface with mass is a result of the balancing of all the forces exerted onto $\Gamma(t)$, which includes the radial component of the trace of normal stress $\boldsymbol{\sigma}\mathbf{n}$ exerted by the fluid onto $\Gamma(t)$, the trace of the radial component of the normal Piola-Kirchoff stress $\mathbf{S}\mathbf{e}_r$ exerted by the thick structure onto $\Gamma(t)$, and the action of the elastic forces associated with $\Gamma(t)$. Therefore, the coupling conditions are given by:

$$\textbf{COUPLING}: \quad \left.\begin{array}{rcl} \mathbf{u}|_{\Gamma(t)} & = & (\partial_t\eta, 0)^T \\ \mathbf{d}|_\Gamma & = & (\eta, 0)^T\ (or\ \partial_t\mathbf{d}|_\Gamma = (\partial_t\eta,0)^T), \\ m\partial_{tt}\eta - T\partial_{zz}\eta & = & -J\boldsymbol{\sigma}\mathbf{n}|_{\Gamma(t)}\cdot\mathbf{e}_r + \mathbf{S}|_\Gamma\mathbf{e}_r\cdot\mathbf{e}_r, \end{array}\right\} \begin{array}{l}\text{on } \Gamma \\ \times(0,T),\end{array} \tag{4}$$

where $J = \sqrt{1+(\partial_z\eta)^2}$ is the Jacobian of the transformation between the Lagrangian coordinates used in the formulation of the structure problem and the Eulerian coordinates used in the formulation of the fluid problem. Vector $\mathbf{e}_r$ is the unit normal to the reference cylinder Γ, while u_r and d_r denote the vertical components of the velocity and displacement of the thick structure, respectively. Notation $\mathbf{u}|_{\Gamma(t)} = (\partial_t\eta, 0)^T$ means $\mathbf{u}(t,z,R+\eta(t,z)) = (\partial_t\eta(t,z), 0)^T$ on $\Gamma \times (0,T)$.

We supplement this problem by the initial and boundary conditions. For example, let the inlet and outlet boundary data for the fluid be given in terms of the dynamic pressure ($p + |\mathbf{u}|^2/2 = P_{in/out}(t)$ on $\Gamma_{in/out}$) and assume that the fluid is entering and leaving the domain parallel to the axis of symmetry ($u_r = 0$ on $\Gamma_{in/out}$). Furthermore, assume that the displacement of both structures is equal to zero at the in/out boundaries ($\eta = d_r = d_z = 0$ on $\Gamma_{in/out}$), and that $\mathbf{S}\mathbf{e}_r = 0$ at the external wall of the thick structure. We can also introduce the symmetry boundary

$\Gamma_b = \{(z,0)|z \in (0,L)\}$ with the symmetry boundary conditions $u_r = \partial_r u_z = 0$, and consider the problem only in the upper half-domain.

The resulting fluid-multi-structure-interaction problem can be summarized as follows (for simplicity we take all the parameters in the problem equal to 1, i.e., $m = T = \rho = \lambda = \mu = \rho_F = \mu_F = 1$): find $\mathbf{u}$, η and $\mathbf{d}$ such that

$$\left.\begin{array}{rcl} (\partial_t \mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u}) & = & \nabla\cdot\boldsymbol{\sigma} \\ \nabla\cdot\mathbf{u} & = & 0 \end{array}\right\} \text{ in } \Omega_F(t),\ t\in(0,T), \tag{5}$$

$$\partial_{tt}\mathbf{d} = \nabla\cdot\mathbf{S} \qquad \text{on } \Omega_S\times(0,T), \tag{6}$$

$$\left.\begin{array}{rcl} \mathbf{u}|_{\Gamma(t)} & = & (\partial_t\eta, 0)^T, \\ \mathbf{d}|_{\Gamma} & = & (\eta,0)^T, \\ \partial_{tt}\eta - \partial_{zz}\eta & = & -J\boldsymbol{\sigma}\mathbf{n}|_{\Gamma(t)}\cdot\mathbf{e}_r + \mathbf{S}|_\Gamma\mathbf{e}_r\cdot\mathbf{e}_r \end{array}\right\} \text{ on } \Gamma\times(0,T), \tag{7}$$

$$\left.\begin{array}{rcl} p + |\mathbf{u}|^2/2 & = & P_{in/out}(t) \\ u_r & = & 0 \end{array}\right\} \text{ on } \Gamma^f_{in/out}\times(0,T), \tag{8}$$

$$\begin{array}{rcl} \eta & = & 0 \text{ on } \partial\Gamma \\ \mathbf{d} & = & \mathbf{0} \text{ on } \Gamma^S_{in/out} \end{array} \tag{9}$$

$$\mathbf{S}\mathbf{e}_r = 0 \text{ on } \Gamma_{ext}, \tag{10}$$

$$\left.\begin{array}{rcl} u_r & = & 0 \\ \partial_r u_z & = & 0 \end{array}\right\} \text{ on } \Gamma_b\times(0,T), \tag{11}$$

with $\mathbf{u}(0,\cdot) = \mathbf{u}_0, \eta(0,\cdot) = \eta_0, \partial_t\eta(0,\cdot) = v_0, \mathbf{d}(0,\cdot) = \mathbf{d}_0, \partial\mathbf{d}(0,\cdot) = \mathbf{V_0}$.

Problem (5)-(7) defines a nonlinear, moving boundary problem of mixed, parabolic-hyperbolic-hyperbolic type. The nonlinearity appears both in the equations, as well as in the coupling conditions (7) via the composite function $\mathbf{u}|_{\Gamma(t)} := \mathbf{u}(t,z,R+\eta(t,z))$. The hyperbolic problem in (7) (3rd equation) serves as a lateral boundary condition for both the fluid problem (5) and for the thick-structure problem (6).

Lemma 2.1. *Problem* (5)-(10) *satisfies the following energy inequality*

$$\frac{d}{dt}\left(E_{kin}(t) + E_{el}(t)\right) + D(t) \le C(P_{in}(t), P_{out}(t)), \tag{12}$$

where

$$\begin{array}{rcl} E_{kin}(t) & := & \|\mathbf{u}\|^2_{L^2(\Omega(t))} + \|\partial_t\eta\|^2_{L^2(\Gamma)} + \|\partial_t\mathbf{d}\|^2_{L^2(\Omega_S)}, \\ E_{el}(t) & := & \|\partial_z\eta\|^2_{L^2(\Gamma)} + 2\|\mathbf{D}(\mathbf{d})\|^2_{L^2(\Omega_S)} + \|\nabla\cdot\mathbf{d}\|^2_{L^2(\Omega_S)}, \end{array}$$

denote the kinetic and elastic energy of the coupled problem, respectively, and the term $D(t)$ captures dissipation $D(t) := \|\mathbf{D}(\mathbf{u})\|^2_{L^2(\Omega(t))}$. The bound $C(P_{in}(t), P_{out}(t))$ depends only on the inlet and outlet pressure data.

Proof. To show that (12) holds,multiply the first equation in (5) by $\mathbf{u}$, integrate over $\Omega_F(t)$, and formally integrate by parts to obtain:

$$\int_{\Omega_F(t)}(\partial_t\mathbf{u}\cdot\mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u}\cdot\mathbf{u}) + 2\int_{\Omega_F(t)}|\mathbf{D}\mathbf{u}|^2 - \int_{\partial\Omega_F(t)}(-p\mathbf{I} + 2\mathbf{D}(\mathbf{u}))\mathbf{n}(t)\cdot\mathbf{u} = 0. \tag{13}$$

To deal with the inertia term we first recall that $\Omega_F(t)$ is moving in time and that the velocity of the lateral boundary is given by $\mathbf{u}|_{\Gamma(t)}$. The transport theorem applied to the first term on the left hand-side of the above equation then gives:

$$\int_{\Omega_F(t)}\partial_t\mathbf{u}\cdot\mathbf{u} = \frac{1}{2}\frac{d}{dt}\int_{\Omega_F(t)}|\mathbf{u}|^2 - \frac{1}{2}\int_{\Gamma(t)}|\mathbf{u}|^2\mathbf{u}\cdot\mathbf{n}(t).$$

To deal with the nonlinear advection term in (13) we integrate by parts, and use the divergence-free condition to obtain:

$$\int_{\Omega_F(t)} (\mathbf{u}\cdot\nabla)\mathbf{u}\cdot\mathbf{u} = \frac{1}{2}\int_{\partial\Omega_F(t)} |\mathbf{u}|^2\mathbf{u}\cdot\mathbf{n}(t) = \frac{1}{2}\Big(\int_{\Gamma(t)} |\mathbf{u}|^2\mathbf{u}\cdot\mathbf{n}(t) - \int_{\Gamma_{in}} |\mathbf{u}|^2 u_z + \int_{\Gamma_{out}} |\mathbf{u}|^2 u_z\Big).$$

These two terms added together give

$$\int_{\Omega_\eta(t)} \partial_t\mathbf{u}\cdot\mathbf{u} + \int_{\Omega_\eta(t)} (\mathbf{u}\cdot\nabla)\mathbf{u}\cdot\mathbf{u} = \frac{1}{2}\frac{d}{dt}\int_{\Omega_\eta(t)} |\mathbf{u}|^2 - \frac{1}{2}\int_{\Gamma_{in}} |\mathbf{u}|^2 u_z + \frac{1}{2}\int_{\Gamma_{out}} |\mathbf{u}|^2 u_z. \tag{14}$$

Notice the importance of nonlinear advection in canceling the cubic term $\int_{\Gamma(t)} |\mathbf{u}|^2\mathbf{u}\cdot\mathbf{n}(t)$!

To deal with the boundary integral over $\partial\Omega_F(t)$ of the normal stress in (13) we first employ the boundary condition $u_r = 0$ form (8) in combination with the divergence-free condition to obtain $\partial_z u_z = -\partial_r u_r = 0$. Now, using the fact that the normal to $\Gamma_{in/out}$ is $\mathbf{n} = (\mp 1, 0)$, we get:

$$\int_{\Gamma_{in/out}} (-p\mathbf{I} + 2\mathbf{D}(\mathbf{u}))\mathbf{n}\cdot\mathbf{u} = \int_{\Gamma_{in}} P_{in}u_z - \int_{\Gamma_{out}} P_{out}u_z. \tag{15}$$

In a similar way, using the symmetry boundary condition (11), we obtain

$$\int_{\Gamma_b} (-p\mathbf{I} + 2\mathbf{D}(\mathbf{u}))\mathbf{n}\cdot\mathbf{u} = 0.$$

What is left is to integrate the normal stress over $\Gamma(t)$. For this purpose we consider the wave equation (2), multiply it by $\partial_t\eta$, and integrate by parts to obtain

$$\int_\Gamma f\partial_t\eta = \frac{1}{2}\frac{d}{dt}\|\partial_t\eta\|^2_{L^2(\Gamma)} + \frac{1}{2}\frac{d}{dt}\|\partial_z\eta\|^2_{L^2(\Gamma)} \tag{16}$$

Furthermore, we consider the elasticity equation (6), multiply it by $\partial_t\mathbf{d}$ and integrate by parts over Ω_S to obtain:

$$\frac{1}{2}\frac{d}{dt}\ \Big(\|\partial_t\mathbf{d}\|^2_{L^2(\Omega_S)} + 2\|\mathbf{D}(\mathbf{d})\|^2_{L^2(\Omega_S)} + \|\nabla\cdot\mathbf{d}\|^2_{L^2(\Omega_S)}\Big) = -\int_\Gamma \mathbf{S}\mathbf{e}_r\cdot\partial_t\mathbf{d}. \tag{17}$$

By enforcing the dynamic and kinematic coupling conditions (7) we obtain

$$-\int_{\Gamma(t)} \sigma\mathbf{n}(t)\cdot\mathbf{u} = -\int_\Gamma J\sigma\mathbf{n}\cdot\mathbf{u} = \int_\Gamma (f - \mathbf{S}\mathbf{e}_r)\partial_t\eta. \tag{18}$$

Finally, by combining (18) with (16), (17), and by adding the remaining contributions to the energy of the FSI problem one obtains the following **energy equality**:

$$\frac{1}{2}\frac{d}{dt}\int_{\Omega_F(t)} |\mathbf{u}|^2 + \frac{1}{2}\frac{d}{dt}\|\partial_t\eta\|^2_{L^2(0,1)} + 2\int_{\Omega_F(t)} |\mathbf{D}\mathbf{u}|^2 + \frac{1}{2}\frac{d}{dt}\|\partial_z\eta\|^2_{L^2(0,1)}$$
$$+\frac{1}{2}\frac{d}{dt}\Big(\|\partial_t\mathbf{d}\|^2_{L^2(\Omega_S)} + 2\|\mathbf{D}(\mathbf{d})\|^2_{L^2(\Omega_S)} + \|\nabla\cdot\mathbf{d}\|^2_{L^2(\Omega_S)}\Big) = \pm P_{in/out}(t)\int_{\Gamma_{in/out}} u_z$$

By using the trace inequality and Korn inequality one can estimate:

$$|P_{in/out}(t)\int_{\Gamma_{in/out}} u_z| \leq C|P_{in/out}|\|\mathbf{u}\|_{H^1(\Omega_F(t))}$$

$$\leq \frac{C}{2\varepsilon}|P_{in/out}|^2 + \frac{\varepsilon C}{2}\|\mathbf{D}(\mathbf{u})\|^2_{L^2(\Omega_F(t))}.$$

By choosing ε such that $\frac{\varepsilon C}{2} \leq 1$ we get the energy inequality (12). □

3. **Weak Solutions.** To define weak solutions of the moving-bounday problem (5)-(11) we introduce the following notation. We use a_S to denote the following bilinear form associated with the elastic properties of the thick structure:

$$a_S(\mathbf{d}, \boldsymbol{\psi}) := \int_{\Omega_S} 2\mathbf{D}(\mathbf{d}) : \mathbf{D}(\boldsymbol{\psi}) + (\nabla \cdot \mathbf{d})\ (\nabla \cdot \boldsymbol{\psi}). \tag{19}$$

Here $A : B := \mathrm{tr}\left[AB^T\right]$. Furthermore, we use b to denote the following trilinear form corresponding to the (symmetrized) nonlinear advection term in the Navier-Stokes equations:

$$b(t, \mathbf{u}, \mathbf{v}, \mathbf{w}) := \frac{1}{2}\int_{\Omega_F(t)} (\mathbf{u}\cdot\nabla)\mathbf{v}\cdot\mathbf{w} - \frac{1}{2}\int_{\Omega_F(t)} (\mathbf{u}\cdot\nabla)\mathbf{w}\cdot\mathbf{v}. \tag{20}$$

Finally, we define a linear functional which associates the inlet and outlet dynamic pressure boundary data to a test function $\mathbf{v}$ in the following way:

$$\langle F(t), \mathbf{v}\rangle_{\Gamma_{in/out}} = P_{in}(t)\int_{\Gamma_{in}} v_z - P_{out}(t)\int_{\Gamma_{out}} v_z.$$

To define a weak solution to problem (5)-(11) we introduce the following function spaces. For the fluid velocity we would like to work with the classical function space. However, due to the moving fluid-structure interface which is modeled by the wave equation, the lateral boundary of the fluid domain is not necessarily a Lipshitz function. Namely, from the energy inequality (12) we see that $\eta \in H^1(0,1)$. The Sobolev embedding then implies that $\eta \in C^{0,1/2}(0,1)$, which means that $\Omega_F(t)$ is not necessarily a Lipshitz domain. However, $\Omega_F(t)$ is locally a sub-graph of a Hölder continuous function. In that case one can define a"Lagrangian" trace

$$\begin{aligned} &\gamma_{\Gamma(t)} : C^1(\overline{\Omega_F(t)}) \to C(\Gamma), \\ &\gamma_{\Gamma(t)} : v \mapsto v(t, z, r + \eta(t, z)). \end{aligned} \tag{21}$$

Furthermore, it was shown in [4, 11, 16] that the trace operator $\gamma_{\Gamma(t)}$ can be extended by continuity to a linear operator from $H^1(\Omega_F(t))$ to $H^s(\Gamma)$, $0 \leq s < \frac{1}{4}$. Therefore, we define the fluid velocity solution space to be the closure in $H^1(\Omega_F(t))$ of the set $\{\mathbf{u} = (u_z, u_r) \in C^1(\overline{\Omega_F(t)})^2 : \nabla\cdot\mathbf{u} = 0, u_z = 0 \text{ on } \Gamma(t),\ u_r = 0 \text{ on } \Omega_F(t)\setminus\Gamma(t)\}$. Using the fact that $\Omega_F(t)$ is locally a sub-graph of a Hölder continuous function we can get the following characterization of the velocity solution space $\mathcal{V}_F(t)$: (see [4, 11])

$$\begin{aligned} \mathcal{V}_F(t) \quad = \quad &\{\mathbf{u} = (u_z, u_r) \in H^1(\Omega_\eta(t))^2 : \nabla\cdot\mathbf{u} = 0, \\ &u_z = 0 \text{ on } \Gamma(t),\ u_r = 0 \text{ on } \Omega_\eta(t)\setminus\Gamma(t)\}. \end{aligned} \tag{22}$$

The function space associated displacement of the thin structural layer is

$$\mathcal{V}_K = H^1_0(\Gamma), \tag{23}$$

and the function space associated with displacement of the thick structural layer is

$$\mathcal{V}_S = \{\boldsymbol{d} = (d_z, d_r) \in H^1(\Omega_S)^2 : d_z = 0 \text{ on } \Gamma,\ \boldsymbol{d} = 0 \text{ on } \Gamma^s_{in/out}\cup\Gamma_{ext}\}. \tag{24}$$

Motivated by the energy inequality (12) we also define the corresponding evolution spaces for the fluid and structure sub-problems, respectively:

$$\mathcal{W}_F(0,T) = L^\infty(0,T;L^2(\Omega_F(t)) \cap L^2(0,T;\mathcal{V}_F(t)), \tag{25}$$

$$\mathcal{W}_K(0,T) = W^{1,\infty}(0,T;L^2(\Gamma)) \cap L^2(0,T;\mathcal{V}_K), \tag{26}$$

$$\mathcal{W}_S(0,T) = W^{1,\infty}(0,T;L^2(\Omega_S)) \cap L^2(0,T;\mathcal{V}_S). \tag{27}$$

Finally, we are in a position to define the solution space for the coupled fluid-multi-layered-structure interaction problem. This space must involve the kinematic coupling condition. The dynamic coupling condition will be enforced in a weak sense, through integration by parts in the weak formulation of the problem. Thus, we define

$$\begin{gathered}\mathcal{W}(0,T) = \{(\mathbf{u},\eta,\boldsymbol{d}) \in \mathcal{W}_F(0,T)\times\mathcal{W}_K(0,T)\times\mathcal{W}_S(0,T) : \\ \mathbf{u}(t,z,R+\eta(t,z)) = \partial_t\eta(t,z)\mathbf{e}_r,\ \boldsymbol{d}(t,z,R) = \eta(t,z)\mathbf{e}_r\}.\end{gathered} \tag{28}$$

The equality $\mathbf{u}(t,z,R+\eta(t,z)) = \partial_t\eta(t,z)\mathbf{e}_r$ is taken in a sense of operator $\gamma_{\Gamma(t)}$, defined in (21). The corresponding test space will be denoted by

$$\begin{gathered}\mathcal{Q}(0,T) = \{(\mathbf{q},\psi,\boldsymbol{\psi}) \in C_c^1([0,T);\mathcal{V}_F\times\mathcal{V}_K\times\mathcal{V}_S) : \\ \mathbf{q}(t,z,R+\eta(t,z)) = \psi(t,z)\mathbf{e}_r = \boldsymbol{\psi}(t,z,R)\}.\end{gathered} \tag{29}$$

Definition 3.1. (Weak Solution) We say that $(\mathbf{u},\eta,\boldsymbol{d}) \in \mathcal{W}(0,T)$ is a weak solution of problem (5)-(11) if for every $(\mathbf{q},\psi,\boldsymbol{\psi}) \in \mathcal{Q}(0,T)$ the following holds:

$$\begin{gathered}-\int_0^T\int_{\Omega_F(t)}\mathbf{u}\cdot\partial_t\mathbf{q} + \int_0^T b(t,\mathbf{u},\mathbf{u},\mathbf{q}) + 2\int_0^T\int_{\Omega_F(t)}\mathbf{D}(\mathbf{u}):\mathbf{D}(\mathbf{q}) \\ -\frac{1}{2}\int_0^T\int_\Gamma(\partial_t\eta)^2\psi - \int_0^T\int_\Gamma\partial_t\eta\partial_t\psi + \int_0^T\int_\Gamma\partial_z\eta\partial_z\psi \\ -\int_0^T\int_{\Omega_S}\partial_t\boldsymbol{d}\cdot\partial_t\boldsymbol{\psi} + \int_0^T a_s(\boldsymbol{d},\boldsymbol{\psi}) = \int_0^T\langle F(t),\mathbf{q}\rangle_{\Gamma_{in/out}} \\ +\int_{\Omega_F(0)}\mathbf{u}_0\cdot\mathbf{q}(0) + \int_\Gamma v_0\psi(0) + \int_{\Omega_S}\mathbf{V}_0\cdot\boldsymbol{\psi}(0).\end{gathered} \tag{30}$$

In deriving the weak formulation we used integration by parts in a classical way, and the following equalities which hold for smooth functions:

$$\begin{aligned}\int_{\Omega_F(t)}(\mathbf{u}\cdot\nabla)\mathbf{u}\cdot\mathbf{q} = {}& \frac{1}{2}\int_{\Omega_F(t)}(\mathbf{u}\cdot\nabla)\mathbf{u}\cdot\mathbf{q} - \frac{1}{2}\int_{\Omega_F(t)}(\mathbf{u}\cdot\nabla)\mathbf{q}\cdot\mathbf{u} \\ & +\frac{1}{2}\int_\Gamma(\partial_t\eta)^2\psi \pm \frac{1}{2}\int_{\Gamma_{out/in}}|u_r|^2v_r,\end{aligned}$$

$$\int_0^T\int_{\Omega_F(t)}\partial_t\mathbf{u}\cdot\mathbf{q} = -\int_0^T\int_{\Omega_F(t)}\mathbf{u}\cdot\partial_t\mathbf{q} - \int_{\Omega_F(0)}\mathbf{u}_0\cdot\mathbf{q}(0) - \int_0^T\int_\Gamma(\partial_t\eta)^2\psi.$$

4. **Conclusions.** The energy estimate (12) and the spaces of weak solutions show that the presence of a fluid-structure interface with mass regularizes the solution of this fluid-structure interaction problem. If we had a FSI problem between an incompressible, viscous fluid and a thick structure only, the trace of the displacement of the structure would not have been defined at the fluid-structure interface, and the evolution of the fluid-structure interface could not be controlled by the energy estimates. In problem (5)-(11) not only that the trace of the displacement and the axial derivative of the displacement of the fluid-structure interface are well defined, but the time-derivative of the displacement of the fluid-structure interface is controlled by the energy estimate. The kinetic energy term $\|\partial_t \eta\|^2$ in the energy estimate (12), which is responsible for the control of the evolution of the fluid-structure interface, appears in (12) due to the inertia of the fluid-structure interface with mass. Our preliminary results indicate that this will play a crucial role in proving existence of a weak solution to this fluid-multi-structure interaction problem [17]. Namely, in a problem in which viscoelasticity of the structure is lacking, the inertia of the fluid-structure interface with mass provides a new regularizing mechanism for a weak solution to exist. This is reminiscent of the results by Hansen and Zuazua [12] in which the presence of a point mass at the interface between two linearly elastic strings with solutions in asymmetric spaces (different regularity on each side) allowed the proof of well-posedness due to the regularizing effects by the point mass. Further research in this direction for problem (5)-(11) is under way [17].

REFERENCES

[1] V. Barbu, Z. Grujić, I. Lasiecka, and A. Tuffaha. *Smoothness of weak solutions to a nonlinear fluid-structure interaction model.* Indiana Univ. Math. J., **57** (2008), 1173–1207.

[2] H. Beirão da Veiga. *On the existence of strong solutions to a coupled fluid-structure evolution problem.* J. Math. Fluid Mech., **6** (2004) 21–52.

[3] P. Causin, J. Gerbeau, and F. Nobile. *Added-mass effect in the design of partitioned algorithms for fluid-structure problems.* Comput. Methods Appl. Mech. Eng., **194** (2005), 4506–4527.

[4] A. Chambolle, B. Desjardins, M. J. Esteban, and C. Grandmont. *Existence of weak solutions for the unsteady interaction of a viscous fluid with an elastic plate.* J. Math. Fluid Mech., **7** (2005), 368–404.

[5] C. H. A. Cheng, D. Coutand, and S. Shkoller. *Navier-Stokes equations interacting with a nonlinear elastic biofluid shell.* SIAM J. Math. Anal., **39** (2007), 742–800.

[6] C. H. A. Cheng and S. Shkoller. *The interaction of the 3D Navier-Stokes equations with a moving nonlinear Koiter elastic shell.* SIAM J. Math. Anal., **42** (2010), 1094–1155.

[7] D. Coutand and S. Shkoller. *Motion of an elastic solid inside an incompressible viscous fluid.* Arch. Ration. Mech. Anal., **176** (2005), 25–102.

[8] D. Coutand and S. Shkoller. *The interaction between quasilinear elastodynamics and the Navier-Stokes equations.* Arch. Ration. Mech. Anal., **179** (2006), 303–352.

[9] B. Desjardins, M. J. Esteban, C. Grandmont, and P. Le Tallec. *Weak solutions for a fluid-elastic structure interaction model.* Rev. Mat. Comput., **14** (2001), 523–538.

[10] Q. Du, M. D. Gunzburger, L. S. Hou, and J. Lee. *Analysis of a linear fluid-structure interaction problem.* Discrete Contin. Dyn. Syst., **9** (2003), 633–650.

[11] C. Grandmont. *Existence of weak solutions for the unsteady interaction of a viscous fluid with an elastic plate.* SIAM J. Math. Anal., **40** (2008), 716–737.

[12] S. Hansen and E. Zuazua. *Exact controllability and stabilization of a vibrating string with an interior point mass.* SIAM J. Cont. Optim., **33** (1995), 1357–1391.

[13] I. Kukavica and A. Tuffaha. *Solutions to a fluid-structure interaction free boundary problem.* DCDS-A, **32** (2012) 1355–1389.

[14] J. Lequeurre. *Existence of strong solutions to a fluid-structure system.* SIAM J. Math. Anal., **43** (2011), 389–410.

[15] B. Muha and S. Čanić. *Existence of a weak solution to a nonlinear fluid-structure interaction problem modeling the flow of an incompressible, viscous fluid in a cylinder with deformable walls.* Arch. Ration. Mech. Anal., **207** (2013), 919–968.

[16] B. Muha. *A note on the trace theorem for domains which are locally subgraphs of a Hölder continuous functions.* to appear in Networks and Heterogeneous Media

[17] B. Muha and S. Čanić. *Existence of a solution to a fluid-multi-layered-structure interaction problem.* Submitted 2013. `arXiv:1305.5310`

E-mail address: canic@math.uh.edu
E-mail address: barism@math.hr

REDUCTION ON CHARACTERISTICS FOR CONTINUOUS SOLUTIONS OF A SCALAR BALANCE LAW

Giovanni Alberti

Università di Pisa, largo Pontecorvo 5, 56127 Pisa, IT

Stefano Bianchini

SISSA, via Bonomea 265, 34136 Trieste, IT

Laura Caravenna

OxPDE, Mathematical Institute
24-29 St Giles', OX1 3LB Oxford, UK

Abstract. We consider continuous solutions u to the balance equation

$$\partial_t u(t,x) + \partial_x \left[f(u(t,x))\right] = g(t,x) \qquad f \in C^2(\mathbb{R}),\ g \in L^\infty(\mathbb{R}^+ \times \mathbb{R})$$

for a bounded source term g. Continuity improves to Hölder continuity when f is *uniformly* convex, but it is not more regular in general. We discuss the reduction to ODEs on characteristics, mainly based on the joint works [5, 1]. We provide here local Lipschitz regularity results holding in the region where $f'(u)f''(u) \neq 0$ and only in the simpler case of autonomous sources $g = g(x)$, but for solutions $u(t,x)$ which may depend on time. This corresponds to a local Lipschitz regularity result, in that region, for the system of ODEs

$$\begin{cases} \dot\gamma(t) = f'(u(t,\gamma(t))) \\ \frac{d}{dt}u(t,\gamma(t)) = g(\gamma(t)). \end{cases}$$

1. Introduction.

In the context of classical solutions, the balance law

$$\partial_t u(t,x) + \partial_x \left[f(u(t,x))\right] = g(t,x), \qquad f \in C^2(\mathbb{R}), \tag{1.1}$$

can be reduced to ordinary differential equations along characteristic curves, defined as those curves $t \mapsto (t,\gamma(t))$ satisfying $\dot\gamma(t) = f'(u(t,\gamma(t)))$. Indeed,

$$\begin{aligned} g(t,\gamma(t)) &= \partial_t u(t,\gamma(t)) + \partial_x \left[f(u(t,\gamma(t)))\right] \\ &= \partial_t u(t,\gamma(t)) + f'(u(t,\gamma(t)))\partial_x u(t,\gamma(t)) \\ &= \partial_t u(t,\gamma(t)) + \dot\gamma(t)\partial_x u(t,\gamma(t)) \qquad\qquad = \frac{d}{dt}u(t,\gamma(t)). \end{aligned}$$

This more generally allows a parallel between the Cauchy problem for a scalar quasilinear first order PDE and for a system of ODEs, which is known as the method of characteristics (see for instance [10], where it is also provided an application to determine local existence).

2000 *Mathematics Subject Classification.* Primary: 35-06; Secondary: 34A12, 35L60.

Key words and phrases. Method of characteristics, Lagrangian formulation, One-dimensional ODEs, balance law, Peano phenomenon.

L.C. acknowledges the UK EPSRC Science and Innovation award to the Oxford Centre for Nonlinear PDE (EP/E035027/1) and the ERC Starting Grant CONSLAW.

If one interprets $f'(u)$ as a velocity, this is just the change of variable from the Eulerian (PDE) to the Lagrangian (ODEs) formulation.

We discuss here what remains of this equivalence when u is just continuous and g is bounded. We prove then in Section 2 that when g depends only on x, but not on the time t, then $u(t,x)$ is locally Lipschitz continuous on the open set where $f'(u)f''(u)$ is nonvanishing. This is sensibly better than the general case, where u is only Hölder continuous. It is based on proving the corresponding result for the system of ODEs. As we are discussing local issues, we will fix for simplicity the domain $\mathbb{R}^2$ and we will assume u bounded.

1.1. A motivation for a different setting.

The development of Geometric Measure Theory in the context of the sub-Riemannian Heisenberg group $\mathbb{H}^n$ brought the attention to *continuous* solutions to the equation

$$\partial_t u(t,x) + \partial_x \left[\frac{u^2(t,x))}{2}\right] = g(t,x). \tag{1.2}$$

Continuity is natural from the fact that u parametrizes a surface. As one studies surfaces that have differentiability properties in the intrinsic structure of the Heisenberg group, but not in the Euclidean structure, then it is not natural assuming more regularity of u than continuity [13], which for bounded sources improves to 1/2-Hölder continuity [4, 5]. Notice that with u continuous the second term of the equation cannot even be rewritten as $u\partial_x u$, because $\partial_x u$ is only a distribution and u is not a suitable test function.

The PDE arises if one wants to show the equivalence between a pointwise, metric notion of differentiability and a distributional one: for $n = 1$ the distributional definition is precisely (1.2), while for $n > 1$ it is a related multi-D system of PDEs. The correspondence was introduced first in [3, 4] for intrinsic regular hypersurfaces, which are the analogue of what are C^1-hypersurfaces in the Euclidean setting. It was extended in [5, 7] when considering intrinsic Lipschitz hypersurfaces, analogue of Lipschitz hypersurfeces in the Euclidean setting. The source term g, in $\mathbb{H}^1$, turns out to be what is called the intrinsic gradient of u, which is the counterpart of the gradient in Euclidean geometry; in $\mathbb{H}^n$ it is one if its components: u locally parametrizes an intrinsic regular hypersurface if and only if (1.2) holds locally with g continuous; it parametrizes an intrinsic regular hypersurface if and only if (1.2) holds locally with g bounded. As the notion of differentiability they provide in the intrinsic structure of $\mathbb{H}^n$ is closer to the Lagrangian formulation, the equivalence between Lagrangian and Eulerian formulation arises as intermediate step of this characterization.

When considering intrinsic Lipschitz hypersurfaces the fact that g is only bounded gives rise to new subtleties. In particular, one already knows by an intrinsic Rademacher theorem [11] that the intrinsic differential exists and it is unique $\mathcal{L}^2$-a.e. However, for the ODE formulation this is not enough: as one needs to restrict this L^∞ function on curves, a precise representative is needed also at points where u is not intrinsically differentiable. Viceversa, if one chooses badly the representative of the source of the ODE formulation a priori it differs on a positive measure set from the source of the ODE. There is however a canonical choice for defining the two sources, which makes the formulations equivalent when the inflection points of f are negligible.

1.2. **Summary of the equivalence.** When u is Lipschitz, the ODEs

$$\begin{cases}\dot\chi(t,x) = f'(u(t,\chi(t,x))) \\ \chi(0,x) = x\end{cases} \qquad x \in \mathbb{R}, f \in C^2$$

provides a local diffeomorphism by the classical theory on ODE. If u is instead continuous, Peano's theorem ensures local existence of solutions, but more characteristics may start at one point and characteristics from different points may collapse (see in [5] the classical example of the square-root). This makes clearly impossible to have a local diffeomorphism, or even having a Lagrangian flow in the sense by Ambrosio-DiPerna-Lions [9, 2]. A recent result about this can be obtained for u not depending on time [6], but it is clearly not our assumption. Dropping out injectivity, it is however possible to construct a continuous change of variables with bounded variation.

Let u be a continuous, bounded function.

Lemma 1.1. *There exists a continuous function $\chi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that*

- *$\tau \mapsto \chi(t,\tau)$ is nondecreasing for every t and surjective;*
- *$\partial_t \chi(t,\tau) = f'(u(t,\tau))$.*

We call it Lagrangian parameterization. This function is not unique.

See [1, 5] for the proof. See also [12] for a similar change of variable, for a $1D$-system. In general one cannot have that χ is SBV [1].

Consider now u continuous distributional solution to (1.1) with g bounded.

Lemma 1.2. *Assume that $\mathcal{L}^1(\mathrm{clos}(\{\text{Inflection points of } f\})) = 0$. Then u is Lipschitz continuous along every characteristic curve.*

The proof follows a computation by Dafermos [8]. For general fluxes, there are cases when u is not Lipschitz along some Lagrangian parameterization [1]. The counterexample holds also for continuous autonomous sources $g(t,x) = g_0(x)$. What we find more striking is the following.

Theorem 1.3. *Assume that $\mathcal{L}^1(\mathrm{clos}(\{\text{Inflection points of } f\})) = 0$. Then there exists a pointwise defined function $\hat g(t,x)$*

$$\frac{d}{dt}u(t,\gamma(t)) = \hat g(t,\gamma(t)) \qquad \text{in } \mathcal{D}'(\mathbb{R}) \text{ for every characteristic curve } \gamma.$$

The proof is based on a selection theorem as a technical device, but $\hat g$ is essentially uniquely defined as the derivative of u along some characteristic.

Remark 1.4. There is a substantial difference between the uniformly convex and the strictly convex cases: in the former at almost every (t,x) there exists a unique value for $\frac{d}{dt}u(t,\gamma(t))$, $\gamma(t) = x$, and it does not depend on which characteristic $\gamma(s)$ one has chosen. That value is the most natural choice of $\hat g$ at those points, and this a.e. defined function $\hat g$ identifies the same distribution as the source term g. Without uniform convexity $\frac{d}{dt}u(t,\gamma(t))$ may not exist on a set of positive $\mathcal{L}^2$-measure, independently of which characteristic γ one choses through the point. The correspondence between distributional and Lagrangian sources gets more complicated with non-covexity.

The converse also holds. We give here a weaker statement without the negligibility condition on the inflection points. As mentioned identifying sources is delicate, we refer for it to the more extensive work [1].

Theorem 1.5. *Assume that a continuous function u has a Lagrangian parameterization χ for which there exists a bounded function $\tilde{g}$ s.t. it satisfies*

$$\frac{d}{dt}u(t,\chi(t,\tau)) = \tilde{g}(t,\chi(t,\tau)) \qquad \text{in } \mathcal{D}'(\mathbb{R}) \text{ for every } \tau \in \mathbb{R}. \tag{1.3}$$

Then there exists a function $g(t,x)$ s.t. (1.1) holds.

Viceversa, if (1.1) holds then there exists a *Lagrangian parameterization χ and function $\tilde{g}$ s.t. (1.3) holds.*

We finally mention that continuous distributional solutions to this simple equation do not dissipate entropy.

Theorem 1.6. *Let u be a continuous distributional solution to (1.1) with bounded source g. Then for every smooth function η and q satisfying $q' = \eta' f'$*

$$\partial_t\left[\eta(u(t,x))\right] + \partial_x\left[q(u(t,x))\right] = \eta'(u(t,x))g(t,x).$$

2. **Some Local Regularity with Autonomous Sources.** We mention a local regularity result holding in the case of autonomous sources: the continuous function $u(t,x)$ is locally Lipschitz continuous in the (open) complementary of the 0-level set of the product $f'(u)f''(u)$. For $f(u) = u^2/2$, this means $u \neq 0$. When the source is not autonomous, then this fails to be true, indeed characteristics may bifurcate also at points where u is not vanishing.

We remind [1] that when f has inflection points of positive measure, then a priori u may not be Lipschitz along some characteristics, even with $g = g(x)$.

Lemma 2.1. *There may be locally multiple solutions to the ordinary differential equation*

$$\begin{cases}\dot\gamma(t) = u(t,\gamma(t)) \\ \ddot\gamma(t) = g(\gamma(t))\end{cases} \qquad \gamma(\bar t) = \bar x \qquad u(t,x) \text{ continuous, } g(x) \text{ bounded}$$

only if $u(\bar t,\bar x) = 0$ but it does not identically vanish in a whole neighborhood.

Remark 2.2. We are not stating existence. The lemma is however still not obvious because we do not have differentiability properties of u, which follow a posteriori by the next corollary in the region where u does not vanish. As a consequence, we do not have now the differentiability of the map $\gamma(t)$ w.r.t. the initial data of the ODE. The lemma asserts indeed the continuity in this variable in that region, provided it exists. We remind that when g depends on t bifurcations may easily occur also if $u \neq 0$.

Proof. We just prove that if u does not vanish at some point $(\bar t,\bar x)$, at that point there is at most one solution of the ODE, as an effect of the autonomous source. The reason is that if $u(\bar t,\bar x)$ does not vanish, then any Lipschitz characteristic $x = \gamma(t)$, with $\bar x = \gamma(\bar t)$, is a diffeomorphism in some neighborhood of $(\bar t,\bar x)$, and we can invert it. This allows to have the space variable as a parameter: the characteristic can be expressed as $t = \theta(x)$. However, the second order relation $\ddot\gamma(t) = g(\gamma(t))$, once expressed in the x variable, can be integrated determining the function θ.

By elementary arguments, it suffices to show that there exists (locally) only one characteristic passing through $(\bar t,\bar x) = (0,0)$ with slope $u(0,0) = 1$. Focus the attention on a neighborhood U of the origin where u is bigger than some $\varepsilon > 0$. Let

$x = \gamma(t)$ be any Lipschitz continuous solution of the ODE. Since $\dot\gamma(0) = u(0,0) > 0$, by the inverse function theorem there exists $\delta > 0$ and a function

$$\theta : (\gamma(-\delta), \gamma(\delta)) \to (-\delta, \delta) \quad : \quad \theta(\gamma(t)) = t,\ \gamma(\theta(x)) = x.$$

Moreover, it is continuously differentiable with derivative

$$\dot\theta(x) = \frac{1}{\dot\gamma(\theta(x))} = \frac{1}{u(\theta(x), x)} \in \left[\frac{1}{\max|u|}, \frac{1}{\varepsilon}\right]. \tag{2.1}$$

From the Lipschitz continuity of $u(t, \gamma(t))$ and the fact that γ is a local diffeomorphism with inverse θ we deduce that the composite function $u(\theta(x), x)$ is Lipschitz continuous. At points $X \subset U$ of differentiability by the classical chain rule

$$\begin{aligned}\lim_{h\downarrow 0} &\frac{\dot\gamma(\theta(x+h)) - \dot\gamma(\theta(x))}{h} \\ &= \frac{\dot\gamma(\theta(x+h)) - \dot\gamma(\theta(x))}{\theta(x+h) - \theta(x)} \frac{\theta(x+h) - \theta(x)}{h} = \ddot\gamma(\theta(x))\dot\theta(x)\end{aligned}$$

and by (2.1) we have that $\dot\theta$ is differentiable at $x \in X$ with derivative

$$\ddot\theta(x) = -\frac{\ddot\gamma(\theta(x))\dot\theta(x)}{[\dot\gamma(\theta(x))]^2} = -\frac{g(\theta(x))}{u^3(\theta(x), x)} \qquad \Leftrightarrow \qquad -\frac{\ddot\theta(x)}{[\dot\theta(x)]^3} = g(x).$$

For those $x \in X$, the differential equation may be rewritten as

$$\frac{d}{dx}\left[\frac{1}{2[\dot\theta(x)]^2}\right] = g(x) \qquad \Leftrightarrow \qquad \frac{d}{dx}\frac{u^2(\theta(x), x)}{2} = g(x).$$

The explicit ODE for $\theta(x)$, with initial data $\theta(0) = 0$, $[\dot\theta(0)]^{-1} = u(0,0) = 1$ is easily solved locally by

$$u^2(\theta(x), x) = \frac{1}{\dot\theta^2(x)} = 1 + 2\int_0^x g(z)dz. \tag{2.2}$$

This shows that the slope of every characteristic through the origin, which is a local diffeomophism, is fixed at each x independently of the characteristic we have chosen: therefore there can be only one characteristic, precisely (in the space parameterization)

$$\theta(x; \bar t, \bar x) = \bar t + \int_{\bar x}^x \frac{1}{\sqrt{u^2(\bar t, \bar x) + 2\int_{\bar x}^w g(z)dz}}dw. \tag{2.3}$$

Notice finally that if u vanishes in a neighborhood, being $\dot\gamma(t) \equiv 0$ there characteristics must be vertical (in that region of the (x,t)-plane). □

Lemma 2.3. *Under the hypothesis of Lemma 2.1, if $g(x)$ is continuous it should also vanish at points where there are more characteristics, but it must not identically vanish in a neighborhood.*

Proof. We show that not only u, but also g must vanish. The argument shows that when two characteristics meet and have both second derivative with the same value, this value must be 0. For simplifying notations, consider two characteristics $\gamma_1(t) \le \gamma_2(t)$ for arbitrarily small $t > 0$ with $\gamma_1(0) = \gamma_2(0) = 0$. If $\gamma_1(t_k)\gamma_2(t_k) \le 0$ for $t_k \downarrow 0$, then

$$0 \le \ddot\gamma_2(0) = g(0) = \ddot\gamma_1(0) \le 0,$$

thus g vanishes. If instead e.g. $g > 0$ near the origin, having excluded the above case there exists $\delta > 0$ such that $0 < \gamma_1(t) \le \gamma_2(t)$ for $t \in [0, \delta]$. Then (2.2) implies that the two curves coincide: having $\dot\gamma_1(t_k) = 0$ or $\dot\gamma_2(t_k) = 0$ for a sequence $|t_k| \downarrow 0$ would contradict the positivity of g, therefore for small $t > 0$ necessarily $\dot\gamma_1(t) > 0$, $\dot\gamma_2(t) > 0$ and therefore

$$u^2(\gamma_1^{-1}(x), x) + 2\int_x^0 g(z)dz = \dot\gamma_1^2(0)$$
$$= 0 = \dot\gamma_2^2(0) = u^2(\gamma_2^{-1}(x), x) + 2\int_x^0 g(z)dz.$$

Being $\dot\gamma_i(t) = u(t,\gamma_i(t))$, $i = 1,2$, by the differential relation, this shows that $\dot\gamma_1(t) \equiv \dot\gamma_2(t)$ for small times. This implies that the two curves coincide.

Finally, suppose g vanishes in a neighborhood. Then, as $\ddot\gamma(t) = 0$ in that neighborhood, characteristics are straight lines. As by the continuity of u characteristics may only intersect with the same derivative, they must be parallel lines and therefore bifurcation of characteristics does not occur. □

We now show that in case u does not vanish, in the above lemma much more regularity holds.

Lemma 2.4. *If for every $(\bar t,\bar x) \in \Omega$ open in $\mathbb{R}^2$ there exists a curve γ s.t.*

$$\begin{cases}\dot\gamma(t) = u(t,\gamma(t)) \\ \ddot\gamma(t) = g(\gamma(t))\end{cases} \qquad \gamma(\bar t) = \bar x \qquad u(t,x) \text{ continuous, } g(x) \text{ bounded}$$

then $u(t,x)$ is locally Lipschitz in the open set $\{(t,x):\ u(t,x) \neq 0\} \subset \Omega$.

Corollary 2.5. *If u is not locally Lipschitz where nonvanishing then the system in Lemma 2.1 cannot have solutions through each point of the plane. In particular, u cannot be a continuous solution to*

$$\partial_t u(t,x) + \partial_x\left[f(u(t,x))\right] = g(x).$$

Proof. By Lemma 2.1 there is a unique characteristic starting at each point $(\bar t,\bar x) \in \Omega = \{(t,x):\ u(t,x) \neq 0\}$, which is given by (2.3). We start comparing the value of u at two points $(0,0)$, $(-t,0)$, $t > 0$, in a ball B compactly contained in Ω. In particular, there exists $\delta(B)$ s.t. the two characteristics starting from the points we have chosen do not intersect if $0 < x < \delta(B)$, as there u does not vanish. For such small x one has by (2.3)

$$\int_0^x \frac{1}{\sqrt{\lambda_1^2 + 2\int_0^w g(z)dz}}dw > -t + \int_0^x \frac{1}{\sqrt{\lambda_2^2 + 2\int_0^w g(z)dz}}dw, \tag{2.4}$$

where we defined $\lambda_1 = u(0,0)$ and $\lambda_2 = u(-t,0)$. Equivalently

$$t > \int_0^x \left\{\frac{1}{\sqrt{\lambda_2^2 + 2\int_0^w g(z)dz}} - \frac{1}{\sqrt{\lambda_1^2 + 2\int_0^w g(z)dz}}\right\}dw.$$

Suppose $\lambda_1 > \lambda_2$. By convexity of the graph of $r \mapsto \frac{1}{\sqrt{r}}$, the RHS is more than

$$\begin{aligned}
&\int_0^x \frac{d}{dr}\left\{\frac{1}{\sqrt{r}}\right\}\Big|_{r=\lambda_1^2+2\int_0^w g(z)dz}(\lambda_2^2-\lambda_1^2)dw\\
&=\left\{\frac{\lambda_2+\lambda_1}{-2}\int_0^x \frac{1}{(\lambda_1^2+2\int_0^w g(z)dz)^{3/2}}dw\right\}(\lambda_2-\lambda_1)\\
&\geq \left\{\frac{\lambda_2+\lambda_1}{2(\lambda_1^2+2Gx)^{3/2}}x\right\}(\lambda_1-\lambda_2)
\end{aligned}$$

The argument in the last brackets is uniformly continuous and as $t \downarrow 0$ it is more than x/λ_1^2. As the inequalities hold for every positive t, $x < \delta = \delta(B)$, the non-intersecting condition (2.4) implies

$$t > \left(\frac{\lambda_1^2}{\delta}+\varepsilon\right)^{-1}(\lambda_1-\lambda_2) \quad\Rightarrow\quad u(0,0)-u(t,0)=\lambda_1-\lambda_2 \leq \left(\frac{\lambda_1^2}{\delta}+\varepsilon\right)t,$$

which is half the Lipschitz inequality at the points $(0,0)$, $(-t,0)$. The other half, for $\lambda_1 < \lambda_2$ is similarly obtained considering small negative x.

For comparing two generic close points (t,x) and $(0,0)$, by the finite speed of propagation one can combine the Lipschitz regularity along characteristics and the Lipschitz regularity along vertical lines. □

Corollary 2.6. *Let $u(t,x)$ be a continuous solution to the balance equation*

$$\partial_t u(t,x) + \partial_x\left[f(u(t,x))\right] = g(x), \qquad g \in L^\infty(\mathbb{R}).$$

The function $u(t,x)$ is locally Lipschitz in the open set

$$\left\{(t,x):\ f'(u(t,x))\cdot f''(u(t,x)) \neq 0\right\}.$$

Proof. We first consider the case of quadratic flux $f(u) = u^2/2$. By Theorem 1.3, there exists a function $\hat{g}(t,x)$ such that we can apply Lemma 2.1, which gives the thesis. If $g \in L^\infty$ they may a priori differ on an $\mathcal{L}^2$-negligible set, but one can prove that $\hat{g}(t,x) = \hat{g}(x)$.

Being u an entropy solution by Theorem 1.6, $f'(u)$ solves the equation

$$\left[f'(u)\right]_t + \left[\frac{f'(u)^2}{2}\right]_x = f''(u)g.$$

By the previous case then $f'(u)$ is Lipschitz in the open set where it does not vanish. If moreover $f''(u)$ does not vanish, then the regularity of u can be proved just by inverting f'. □

REFERENCES

[1] G. Alberti, S. Bianchini, L. Caravenna, *Eulerian and Lagrangian continuous solutions to a balance law with non convex flux*, Forthcoming.

[2] L. Ambrosio, *Transport equation and Cauchy problem for BV vector fields*, Invent. Math. 158 (2004) 227–260.

[3] L. Ambrosio, F. Serra Cassano, D. Vittone, *Intrinsic Regular Hypersurfaces in Heisenberg Groups,* J. Geom. Anal. 16 (2006), no. 2, 187–232.

[4] F. Bigolin, F. Serra Cassano, *Intrinsic regular graphs in Heisenberg groups vs. weak solutions of non-linear first-order PDEs*, Adv. Calc. Var. 3 (2010), 69-97

[5] F. Bigolin, L. Caravenna, F. Serra Cassano, *Intrinsic Lipschitz graphs in Heisenberg groups and continuous solutions of a balance equation*, arXiv:1202.3083.

[6] G. Crippa, *Lagrangian flows and the one dimensional Peano phenomenon for ODEs*, J. Differential Equations 250 (2011), no. 7, 3135–3149.

[7] G. Citti, M. Manfredini, A. Pinamonti, F. Serra Cassano, *Approximation, area formula and characterization of intrinsic Lipschitz functions in Heisenberg groups*, Calc. Var. and PDEs, DOI: 10.1007/s00526-013-0622-8.

[8] C. M. Dafermos, *Continuous solutions for balance laws,* Ricerche di Matematica 55 (2006), 79–91.

[9] R.J. DiPerna, P.-L. Lions, *Ordinary differential equations, transport theory and Sobolev spaces*, Invent. Math. 98 (1989) 511–547.

[10] L. Evans, *Partial differential equations.* Second edition. Graduate Studies in Mathematics, 19. American Mathematical Society, Providence, RI, 2010.

[11] B. Franchi, R. Serapioni, F. Serra Cassano, *Differentiability of intrinsic Lipschitz Functions within Heisenberg groups*, J. Geom. Anal. 21 (2011), no. 4, 1044–1084.

[12] H. Holden, R. Xavier, *Global semigroup of conservative solutions of the nonlinear variational wave equation*, Arch. Ration. Mech. Anal. 201(3):871–964 (2011).

[13] B. Kirchheim, F. Serra Cassano, *Rectifiability and parametrization of intrinsic regular surfaces in the Heisenberg group*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (5) III (2004), 871–896.

E-mail address: galberti1@dm.unipi.it
E-mail address: Stefano.Bianchini@sissa.it
E-mail address: Laura.Caravenna@maths.ox.ac.uk

DETONATION WAVE PROBLEMS: MODELING, NUMERICAL SIMULATIONS AND LINEAR STABILITY

Filipe Carvalho

Instituto Politécnico de Viana do Castelo
Centro de Matemática da Universidade do Minho, Portugal

Ana Jacinta Soares

Centro de Matemática da Universidade do Minho
Departamento de Matemática e Aplicações
Universidade do Minho, Portugal

Abstract. Traveling waves arising in detonation physics are described by the reactive Euler equations obtained in the fluid dynamical limit of the Boltzmann equation for a binary reactive mixture. The hydrodynamic linear stability of the detonation wave solution is investigated with a normal mode analysis. Numerical simulations are performed for both the detonation wave solution and its linear stability.

1. **Introduction.** Detonation waves are combustion fronts triggered by a strong shock and sustained by a chemical reaction [1, 2]. They can be mathematically modeled by the reactive Euler equations, which includes conservation laws of momentum and total energy (kinetic and chemical) of the mixture as well as reaction rate equation for the constituents. Experimental studies show that the detonation waves tend to be structurally unstable and a first attempt to understand and describe the instabilities is a hydrodynamic stability analysis based on the linearization of the governing Euler equations and a normal-mode representation of the perturbations [3]. It is well known that the numerical analysis of detonation waves and its hydrodynamic stability is a rich and challenging problem with many engineering applications [2, 3]. We investigate this problem starting by considering a binary mixture modeled by the Boltzmann equation for the constituent distribution functions, with both elastic scattering and reactive collision terms. Then we pass to the fluid dynamical limit for an Eulerian regime and use the resulting macroscopic reactive equations to investigate the existence of detonation wave solutions. The analysis presented in this paper includes the modeling of the detonation waves, its hydrodynamic stability and the numerical treatment of these problems.

2. **The model for the explosive reactive mixture.** We consider an idealized explosive mixture of two constituents, denoted by A and B, whose particles undergo a reversible symmetric chemical reaction of type A + A $\rightleftharpoons$ B + B. The molecules have binding energies E_A and E_B, the same mass m and equal diameter d.

2000 *Mathematics Subject Classification.* Primary: 35L67, 80A25, 76E99; Secondary: 80A32, 82C40.

Key words and phrases. Reactive Euler equations, Detonation wave solutions, Hydrodynamic linear stability.

In one-space dimension, the macroscopic physical observables of the reactive mixture are the constituent number densities n_A, n_B, the mixture mean velocity v and the mixture temperature T. Neglecting diffusion and heat fluxes, as well as shear and non-equilibrium stresses, the governing equations of the mixture are the reactive Euler equations, namely the rate equations of the constituents together with the conservation laws of momentum and total energy of the whole mixture. They are given by

$$\frac{\partial n_\alpha}{\partial t} + \frac{\partial}{\partial x}\left(n_\alpha v\right) = \tau_\alpha, \qquad \alpha = A, B \tag{1}$$

$$\frac{\partial}{\partial t}(\varrho v) + \frac{\partial}{\partial x}\left(\varrho v^2 + p\right) = 0 \tag{2}$$

$$\begin{aligned}&\frac{\partial}{\partial t}\Big(\frac{1}{2}\varrho v^2 + \frac{3}{2}nkT + n_A E_A + n_B E_B\Big)\\ &\qquad + \frac{\partial}{\partial x}\Big[pv + \Big(\frac{1}{2}\varrho v^2 + \frac{3}{2}nkT + n_A E_A + n_B E_B\Big)v\Big] = 0\end{aligned} \tag{3}$$

where τ_α represents the reaction rate of the $\alpha-$constituent, such that $\tau_B = -\tau_A$ as predicted by the chemical law. Moreover, ϱ, p and n are the mass density, pressure and number density of the whole mixture, with

$$n = n_A + n_B, \qquad \varrho = mn, \qquad p = nkT \tag{4}$$

The number density n_α represents a measure of the concentration of the constituent α and thus Eq. 1 constitutes the rate equation of the considered reactive mixture. The term $n_A E_A + n_B E_B$ appearing in Eq. 3 represents the chemical bond energy of the mixture, and k is the Boltzmann constant.

The above governing equations 1-3 have been derived from a kinetic theory based on the Boltzmann equation extended to the considered reactive mixture, see Ref. [4, 5]. In particular, a chemical regime of slow reactive process is assumed and an appropriate scalling is introduced in terms of the Knudsen number associated to elastic scattering. The corresponding fluid dynamic limit is obtained by means of the Chapman-Enskog method [6] which leads to a distribution function containing the non-equilibrium effects associated to the chemical reaction. The procedure leads to the explicit computation of the reaction rate τ_α, which follows an Arrhenius-type law, given by

$$\begin{aligned}\tau_B = -\tau_A, \quad \tau_A = -4n_A^2 \mathsf{d}_r^2 \sqrt{\frac{\pi kT}{m}}\, e^{-\varepsilon_A^\star}\Big[1 + \varepsilon_A^\star + \frac{n_A^2}{128n^2}\left(\frac{\mathsf{d}}{\mathsf{d}_r}\right)^2 Q_R^\star\\ \times\left(1 + Q_R^\star + Q_R^\star\varepsilon_A^\star + \varepsilon_A^\star - 2\varepsilon_A^{\star 2}\right)\left(4\varepsilon_A^{\star 3} - 8\varepsilon_A^{\star 2} - \varepsilon_A^\star - 1\right)e^{-\varepsilon_A^\star}\Big]\end{aligned} \tag{5}$$

where $\varepsilon_A^\star$ is the activation energy of the forward chemical reaction in units of kT and $Q_R^\star = 2(E_B - E_a)/kT$ is the reaction heat of the chemical reaction, also in units of kT. The details of the passage to the fluid dynamic limit are given in Ref. [7] and revisited in Refs. [4, 5].

The qualitative properties of the Euler equations are well known in literature, see for example Refs. [8, 9, 10]. In particular they constitute an hyperbolic set of non-linear PDE's and admit shock profile solutions.

In addition, when a reactive gaseous mixture is considered, like the one previously introduced in this section, an interesting and relevant type of shock solutions may arise, namely steady traveling detonation wave solutions.

3. **Detonation waves.** Physically, detonation wave solutions represent a combustion front in which a strong shock wave ignites the explosive mixture and the burning keeps the shock advancing and proceeding to equilibrium in the reaction zone behind the shock. The configuration of such solutions is well described in the literature of the detonation phenomenon and a good and accepted model for such solutions is the well known ZND model developed by Zeldovich, von Neumann and Doering, the founders of the modern detonation theory, see Refs. [1, 2]. According to the ZND model, the structure of the detonation wave solution consists of a leadind planar non-reactive shock propagating with constant velocity, followed by a finite reaction zone where the chemical process evolves.

3.1. **Steady detonation solutions.** We investigate one-dimensional ZND steady detonation wave solutions propagating in an explosive mixture described by the model of Section 2. Mathematically, these solutions are traveling waves for the reactive Euler equations 1-3.

We consider a planar shock wave propagating in the $x-$direction with constant velocity D from left to the right. Ahead the shock front, we consider the initial quiescent mixture at rest, where the rate of the chemical reaction is negligible. Such *initial state* is labeled as $I = (n_A^+, n_B^+, 0, T^+)$. The passage of the shock raises the density and temperature above the ignition values, so that the chemical reaction is suddenly activated. The state just behind the shock wave, where the chemical reaction is triggered, is the *von Neumann state* which is labeled as $N = (n_A, n_B, v, T)$. The shock wave is followed by a reaction zone with a finite length, where the chemical reaction continuously proceeds from the state N to a *final state* F of chemical equilibrium. All states inside the reaction zone are *intermediate states* R of partial reaction.

Traveling detonation waves with velocity D are determined as solutions depending on the normalized steady variable

$$x_s = \frac{x - Dt}{Dt_c}, \qquad \text{with} \qquad t_c = \frac{1}{4n^+\mathsf{d}^2}\sqrt{\frac{m}{\pi k T^+}} \tag{6}$$

where t_c is a characteristic time. The Euler equations 1-3 are transformed to the steady frame attached to the shock wave and re-written in terms of the variable x_S. They transform to a system of four ODE's for the unknowns n_A, n, v, T, which can be writen in the more conservative form

$$\frac{d}{dx}\Big[(v - D)\, n_A\Big] = Dt_c\tau_A \tag{7}$$

$$\frac{d}{dx}\Big[(v - D)\, n\Big] = 0 \tag{8}$$

$$\frac{d}{dx}\Big[(v - D)\, \varrho v + nkT\Big] = 0 \tag{9}$$

$$\frac{d}{dx}\left[(v - D)\left(\frac{3}{2}nkT + \frac{\varrho v^2}{2} + E_A n_A + E_B n_B\right) + nkTv\right] = 0 \tag{10}$$

where we have written x in place of x_S. The explicit expression of the reaction rate is given by expression 5.

Von Neumann state. State N, just behind the shock wave, is the solution of a non-reactive shock problem. Euler equations 7-10 hold true with vanishing reaction rate and should be taken in a weak integrated sense between the initial state and the von Neumann state. The integration leads to the algebraic Rankine-Hugoniot jump conditions connecting the state N to the state I, namely

$$n_A\,(v-D) = -n_A^+ D \tag{11}$$

$$n\,(v-D) = -n^+ D \tag{12}$$

$$\varrho v\,(v-D) + nkT = kn^+T^+ \tag{13}$$

$$\left(\frac{3}{2}nkT + \frac{\varrho v^2}{2} + E_A n_A + E_B n_B\right)(v-D) + nkTv = -\left(\frac{3}{2}kn^+T^+ + E_A n_A^+ + E_B n_B^+\right)D \tag{14}$$

Intermediate and final states. States R and F, in the reaction zone, are obtained solving an initial value problem for the number density n_A, with initial condition assigned at the von Neumann state. The chemical reaction is the dominant process within this problem since the evolution of the gaseous mixture within the reaction zone is determined by the reactive process. More in detail, conservative Eqs. 8-10 are integrated between the initial state and an arbitrary state within the reaction zone. The resulting three algebraic Rankine-Hugoniot conditions allow to express the state variables n_B, v and T in terms of n_A and reduce the system to the ODE

$$\frac{dn_A}{dx} = \frac{Dt_c\tau_A}{v - D + n_A\frac{dv}{dn_A}} \tag{15}$$

Equation 15 represents the rate law in the shock attached frame and specifies the chemical composition of the explosive mixture in the reaction zone. The integration of Eq. 15 and further computation of the remaining state variables through the three algebraic Rankine-Hugoniot conditions characterize the thermodynamical state of the mixture in the reaction zone. The final state of chemical equilibrium, is obtained when the reaction rate τ_A vanishes and n_A becomes constant, so that the chemical concentrations of constituents A and B remain unchanged.

3.2. **Numerical solutions.** Numerical solutions of the detonation wave problem can be determined, characterizing the states N, R and F for different values of the velocity D, reaction heat Q_R^* and activation energy $\varepsilon_A^\star$. Some numerical simulations are performed for one elementary reaction of a theoretical detonating mixture. The initial state are assumed as follows

$$m = 0.01\,Kg/mol, \quad n_A^+ = 0.35\,mol/l, \quad n_B^+ = 0\,mol/l$$
$$T^+ = 298.15\,K, \quad E_A = 2400\,K, \quad \varepsilon_A^\star = 6 \tag{16}$$

Some representative profiles are shown in Figure 1 for the mixture pressure, in dependence of the algebraic distance behind the shock wave. The left frame of Figure 1 is obtained for a fixed detonation velocity and refers to exothermic chemical reactions with reaction heat $Q_R^* = -2$ and $Q_R^* = -1$. It shows that the pressure profile and the thickness of the reaction zone decrease for the hight absolute value of the reaction heat. The right frame refers to exothermic reaction with $Q_R^* = -1$ and detonation velocity $D = 1600\,ms^{-1}$ and $D = 1700\,ms^{-1}$. It shows that the pressure profile increases and the thickness of the reaction zone decreases for the hight value of the detonation velocity. The results are in a good agreement with the analytical studies and numerical predictions known from the literature on the

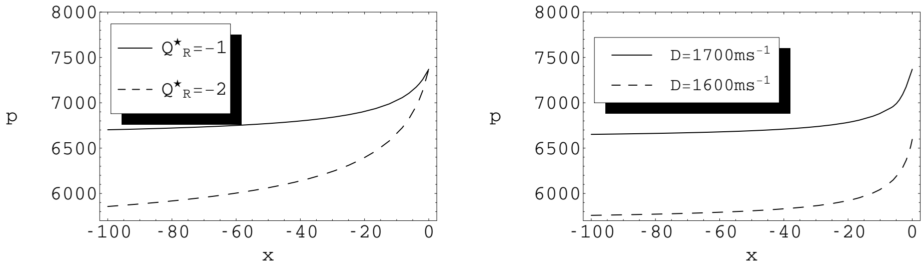


FIGURE 1. Mixture pressure profile. *Left:* detonation velocity $D = 1700\,ms^{-1}$, exothermic reaction with $Q_R^\star = -1$ (solid line) and $Q_R^\star = -2$ (dashed line). *Right:* reaction heat $Q_R^\star = -1$ (exothermic reaction), detonation velocity $D = 1700\,ms^{-1}$ (solid line) and $D = 1600\,ms^{-1}$ (dashed line).

detonation phenomenon. See, for example, Refs. [1, 2]. A more detailed discussion about these results can be seen in Ref. [5].

4. **Hidrodynamic linear stability.** It is well known from theoretical studies and experimental investigations, see Refs. [1, 2], that the detonation wave solution tends to be structurally unstable and can degenerate into an oscillatory solution in the long-time limit. Such oscillatory configuration exhibits complex three-dimensional non-linear perturbations, so that its characterization results to be a very complex and difficult problem, from either analytical or numerical point of view.

As a first step of a formal treatment, one studies the problem of hydrodynamical stability of the steady solution, formulated as follows: one assumes that a small rear boundary perturbation is instantaneously assigned, inducing a deviation on the shock wave position; as a consequence, small perturbations are induced on the state variables and one is interested in their evolution in the reaction zone. The pertinent question is to investigate if all perturbations decay with time or if any perturbation grow with time. In the first case, the steady solution becomes stable and in the latter it becomes unstable.

4.1. **Stability analysis.** From the mathematical point of view, the stability problem requires first the transformation to the perturbed wave coordinate

$$\overline{x} = x - \psi(t), \qquad \text{with} \qquad \psi(t) = Dt + \widetilde{\psi}(t) \tag{17}$$

where $\psi(t)$ represents the location of the perturbed shock wave and $\widetilde{\psi}(t)$ the displacement of the wave from the unperturbed position.

Normal mode approach. Since the perturbations are small, we then linearize the governing equations and Rankine-Hugoniot conditions about the steady detonation solution, assuming the following expansions for the the state variables and shock distortion with exponential time dependence,

$$z(x,t) = z^*(x) + e^{at}\,\overline{z}(x), \qquad \psi(t) = e^{at}, \qquad a \in \mathbb{C} \tag{18}$$

where $z = [n_A\ n_B\ v\ p]^T$, $z^*(x)$ is the steady solution, $\overline{z}(x)$ is the vector of complex eigenfunctions representing the unknown spatially disturbances, and a is the complex eigenvalue, with $\mathrm{Re}\,a$ and $\mathrm{Im}a$ being the disturbance growth rate and frequency, respectively.

Linearized stability equations. The transformation to the perturbed shock and the linearization by means of the expansions 18 lead to the stability equations

$$Da\overline{n}_\alpha+(v^*-D)\,\frac{d\overline{n}_\alpha}{dx}+\frac{dn^*_\alpha}{dx}\,(\overline{v}-aD)+\frac{dv^*}{dx}\overline{n}_\alpha+n^*_\alpha\frac{d\overline{v}}{dx}=\overline{\tau}_\alpha, \quad \alpha=A,B \tag{19}$$

$$\varrho^*aD\overline{v}+\frac{d\overline{p}}{dx}+\varrho^*\frac{dv^*}{dx}\,(\overline{v}-aD)+(v^*-D)\,\frac{dv^*}{dx}\overline{\varrho}+\varrho^*\,(v^*-D)\,\frac{d\overline{v}}{dx}=0 \tag{20}$$

$$Da\overline{p}+\frac{5}{3}\left(p^*\frac{d\overline{v}}{dx}+\overline{p}\frac{dv^*}{dx}\right)+(v^*-D)\,\frac{d\overline{p}}{dx}+(\overline{v}-aD)\,\frac{dp^*}{dx}=\frac{Q^*_RDt_c\overline{\tau}_A}{3} \tag{21}$$

where the linearized reaction rate $\overline{\tau}_\alpha$ is given by

$$\overline{\tau}_A=-4\mathsf{d}_r^2\sqrt{\frac{\pi k}{m}}e^{-\epsilon^\star}\left[\left(2n^*_A\overline{n}_A\sqrt{T^*}+\frac{\overline{p}+\frac{\overline{n}}{n^*}p^*}{2n^*k\sqrt{T^*}}\,{n^*_A}^2\right)\left(1+\epsilon^\star+\Gamma {x^*_A}^2\right)\right.$$
$$\left.+\,2\sqrt{T^*}\,\frac{{n^*_A}^3}{n^*}\,(-n^*_A\overline{n}_B+n^*_B\overline{n}_A)\right], \qquad\qquad \overline{\tau}_B=-\overline{\tau}_A$$

Equations 19-21 constitute a system of eight first-order homogeneous linear ordinary differential equations with spatially varying coefficients, for the real and imaginary parts of the complex perturbations.

Initial conditions. The initial conditions for the stability equations are obtained from the Rankine-Hugoniot relations 11-14, after transforming to the wave coordinate and linearizing around the steady state. The resulting conditions are

$$\overline{n}_\alpha(0)=\frac{(n^*_\alpha-n^+_\alpha)\,a-n^*_\alpha\overline{v}(0)}{v^*-D}\;, \qquad \alpha=A,B \tag{22}$$

$$\overline{v}(0)=\frac{3\varrho^+{v^*}^2+\frac{3}{2}\,(p^*-p^+)-\frac{3}{2}D\varrho^+v^*+2E_An^++Q^*_Rn^+_B}{-\varrho^*\,(v^*-D)^2+\frac{5}{2}p^*}\;a \tag{23}$$

$$\overline{p}(0)=-\varrho^+av^*-(v^*-D)\,\varrho^*\overline{v}(0) \tag{24}$$

Closure condition. Equations 19-21 and their initial conditions 22-24 involve the complex perturbation parameter a and therefore the stability system is not closed. The closure condition is given by the dispersion relation of the normal modes 18 formulated at the final state F, that is

$$\overline{v}_F+a=\frac{-1}{\gamma\varrho^*_{eq}c^*_{eq}}\,\overline{p}_F, \tag{25}$$

where γ is the ratio of specific heats, c^*_{eq} and ϱ^*_{eq} the isentropic sound speed and gas density at the equilibrium final state.

Stability problem. The linear stability problem consists in the *eight* ordinary differential equations 19-21 for the complex eigenfunction disturbances $\overline{z}(x)$ and complex eigenvalue parameter a, with initial conditions 22-24 and closure condition 25.

The objective is to determine the instability modes which correspond to a positive growth rate $\mathrm{Re}\, a$. Moreover, since these modes occur in conjugate pairs, they are searched in the upper-right quarter of the complex plane.

The stability problem is solved numerically and its solution gives valuable predictions about the stability of the steady detonation solutions.

4.2. **Numerical technique.** For a given set of thermodynamical and chemical parameters characterizing the steady detonation wave solution, eigenfunctions $\overline{z}(x)$ and eigenvalues a are determined. We apply a numerical technique which combines the iterative shooting method proposed by Lee and Stewart in paper [3] with the argument principle used by Erpenbeck in paper [11]. The basic idea is the following, see Ref. [5]: *(i)* we start with a trial value of a in a fixed bounded domain $\mathcal{R}$ of the complex plane; *(ii)* then we use a fourth order Runge-Kutta routine to integrate the stability equations 19-21 in the reaction zone, say for $x \in]x_F, 0[$, with initial conditions 22-24 at $x=0$; *(iii)* finally we enquire if the approximate solution $\overline{z}(x)$, $x \in]x_F, 0]$ and related parameter a satisfy the boundary condition 25; *(iv)* if this is not the case, we iterate the procedure on the trial value a until condition 25 is satisfied. Since, in general, a trial value of a does not produce a stability solution, in the sense that condition 25 is not verified, the key preliminary step consists in searching appropriate trial values for a. To do this, we introduce the residual function

$$\mathscr{H}(a) = \overline{v}(x_F) + a + \frac{1}{\gamma \varrho_{eq}^* c_{eq}^*}\, \overline{p}(x_F)\,, \quad a \in \mathcal{R}, \tag{26}$$

and search the zeros of $\mathscr{H}$ in the considered region $\mathcal{R}$. To count the number Z of zeros we use the argument principle and estimate the quantity

$$Z = \frac{1}{2\pi i} \int_k^\ell \frac{\mathscr{H}'(\zeta(t))}{\mathscr{H}(\zeta(t))} \parallel \zeta'(t) \parallel dt \tag{27}$$

where $\zeta : [k, \ell] \to \mathbb{C}$ is a path smooth by parts, describing the contour of $\mathcal{R}$ in the positive direction. Such estimation requires a rather involved numerical technique which is explained in detail in Refs. [4, 5]. Finally successive refinements of $\mathcal{R}$ are considered and a three-dimensional plot of $|\mathscr{H}|$ in the last refinement is used to determine the location of the zeros.

Some numerical simulations are performed in order to investigate the response of the steady detonation wave solution to the rear boundary perturbations.

4.3. **Stability results.** The stability problem is solved numerically for the set of thermodynamical and chemical parameters given in Eq. 16. We choose a rectangular region $\mathcal{R}$ in the upper-right complex plane such that $0.001 < Re(a) < 0.02$ and $0.001 < Im(a) < 0.1$. The reaction heat and the detonation velocity are varying in the ranges $-2 \leq Q_R^* \leq 2$ and $1278\, ms^{-1} \leq D \leq 1896\, ms^{-1}$, respectively. Table 1 shows the number of instability modes in the region $\mathcal{R}$, for different values of the detonation velocity, and for fixed reaction heat and activation energy, $Q_R^\star = -1$ and $\varepsilon^\star = 6.5$. One can see that the number of instability modes in the region $\mathcal{R}$ is zero when $D \geq 1645\, ms^{-1}$, and that it increases for lower values of D.

Detonation velocity	Number of modes	Detonation velocity	Number of modes
$1896\, ms^{-1}$	0	$1518\, ms^{-1}$	17 to 28
$1700\, ms^{-1}$	0	$1391\, ms^{-1}$	57 to 120
$1645\, ms^{-1}$	0	$1328\, ms^{-1}$	250 to 334
$1581\, ms^{-1}$	1 to 3	$1278\, ms^{-1}$	442 to 493

TABLE 1. Number of instability modes in the region $\mathcal{R}$, for different values of the detonation velocity D. The reaction is exothermic with reaction heat $Q_R^\star = -1$ and activation energy $\varepsilon^\star = 6.5$.

Figure 2 shows the stability boundary in the parameter plane $Q^*_R - \varepsilon^\star_A$, for detonation velocity $D = 1700\, ms^{-1}$. A pair $(Q^*_R, \varepsilon^\star_A)$ in the stability zone indicates that for the corresponding values of Q^*_R and $\varepsilon^\star_A$, no instability modes have been found in the domain $\mathcal{R}$. Conversely, a pair in the instability zone indicates that one instability mode, at least, has been found in $\mathcal{R}$.

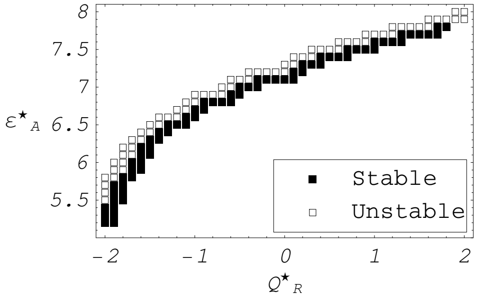


Figure 2. Stability boundary in the $Q^\star_R - \varepsilon^\star$ plane, for detonation velocity $D = 1700\, ms^{-1}$ and for the considered region $\mathcal{R}$.

The results of Table 1 can be compared to those of Figure 2 considering, simultaneously, $Q^\star_R = -1$, $\varepsilon^\star = 6.5$ and $D = 1700\, ms^{-1}$. The results in both representations indicate a stable solution since no instability modes are found.

Acknowledgments. This research was partially supported by the FCT Phd Grant SFRH/BD/28795/2006, by the Research Centre of Mathematics of the University of Minho through the FEDER Funds "Programa Operacional Factores de Competitividade ?COMPETE", and by the Portuguese Funds through FCT - 'Fundação para a Ciência e Tecnologia" within the Project PEst-C/MAT/UI0013/2011.

REFERENCES

[1] W. Fickett, *"Introduction to Detonation Theory"* University of California Press, Berkeley, 1986.

[2] J. H. S. Lee, *"The Detonation Phenomenon"*, Cambridge University Press, Cambridge, 2008.

[3] H. I. Lee and D. S. Stewart, *Calculation of linear detonation stability: one dimensional instability of plane detonation*, J. Fluid Mech., **216** (1990), 103–132.

[4] F. Carvalho and Ana J. Soares, *On the dynamics and linear stability of one-dimensional steady detonation waves*, J. Phys. A: Math. Theor., **45** (2012), 255501, 1–23.

[5] F. Carvalho, "Mathematical methods for the Boltzmann equation in the context of chemically reactive gases", Ph.D thesis, University of Minho, 2012. http://hdl.handle.net/1822/24430

[6] S. Chapman and T. G. Cowling, *"Mathematical Theory on Non-Uniform Gases"*, Cambridge University Press, Cambridge, 1999.

[7] G. M. Kremer and Ana J. Soares, *Effect of reaction heat on Maxwellian distribution functions and rate of reactions*, J. Stat. Mech., **P12003** (2007), 1–16.

[8] C. Cercignani, *"Theory and Application of the Boltzmann Equation"*, Scottish Academic Press, Edinburgh, 1975.

[9] G. A. Bird, *"Molecular Gas Dynamics and the Direct Simulations of Gas Flow"*, Claredon Press, Oxford, 1994.

[10] G. M. Kremer, *"An introduction to the Boltzmann equation and transport processes in gases"*, Springer, Berlin, 2010.

[11] J. J. Erpenbeck, *Stability of idealized one-reaction detonations*, Phys. Fluids., **7** (1964), 684–696.

E-mail address: filipecarvalho@esce.ipvc.pt
E-mail address: ajsoares@math.uminho.pt

ON SINGULAR POINTS FOR CONVEX PERMEABILITY MODELS

Pablo Castañeda and Dan Marchesin

Instituto Nacional de Matemática Pura e Aplicada
Estrada Dona Castorina, 110
Rio de Janeiro, 22460-320 RJ, Brazil

Frederico Furtado

Departament of Mathematics University of Wyoming
82071-3036 Laramie, WY, USA

Abstract. We focus on a system of two conservation laws representing a large class of models relevant for petroleum engineering, the domain of which possesses singular points. It has been conjectured that the structure of the Riemann solution in the saturation triangle is strongly influenced by the nature of the umbilic point. In the current work we show that features originally related to umbilic points actually belong to a distinct point, the new Equal-Speed Shocks point.

Even though the location of the umbilic point is known, for the first time, we relate the umbilic point to a physical property, namely, the minimum of the total mobility for any Corey model.

1. **Introduction.** We are interested in the study of injection problems for 2×2 systems of conservation laws; a survey may be found in [2, 4, 7, 12] and references therein. The solution construction for the injection of water and gas is presented in [2] for the case of quadratic Corey models.

We discuss the location of the umbilic point in the interior of the triangle and the new "Equal-Speed Shocks" (ESS) point, which arises in these more general non-symmetric models. Analyses on umbilic points were made in the last few years [8, 9, 14]. The special case of quadratic Corey models is discussed in [1].

We consider models for reservoirs that may contain three fluids, for concreteness, we call them water, gas, and oil; although they could be any three fluids that are immiscible with each other. For simplicity, we assume that the three phases are incompressible, that gravitational segregation and capillary effects are negligible, and that there is no mass transfer among the phases. The flow occurs in one dimension at constant injection rate and fixed proportion of injected fluids. The mobility of each phase is assumed to be a convex function of its own saturation and inversely proportional to the phase viscosity. The mathematical model consists of

2000 *Mathematics Subject Classification.* Primary: 35L65, 35L67; Secondary: 58J45, 76S05.

Key words and phrases. Conservation laws, Riemann problem, umbilic point, WAG injection, petroleum engineering, flow in porous media, Corey permeability model.

This work was supported in part by: CAPES/FAPERJ Postdoc Grant E-26/102.474/2010, DOE under Grant DE-FE0004832, Grant of the U. Wyoming WYDEQ49811FRTD, CNPq under Grants 472923/2010-2, 301564/2009-4, and FAPERJ under Grants E-26/102.965/2011, E-26/111.416/2010, E-26/110.337/2010.

two conservation laws representing Darcy's law combined with mass conservation for two of the phases. The flow problem depends on two viscosity ratios and the precise choice of mobilities. (The overall picture of solutions given in [2] is essentially unchanged in the more general class of models treated in this work, see [5, 6].)

A Corey-type model loses strict hyperbolicity at an umbilic point. Models without umbilic points have been considered for three-phase flow; see [7]. They yield simple solutions for the injection problem. However, they are unrealistic because immiscibility of the three phases seems to be related to loss of strict hyperbolicity [3, 14, 16], *i.e.*, either umbilic points or elliptic regions are present. Models with umbilic points have complicated solutions, but are still well behaved mathematically; see [9, 12] for a review of their properties.

This work is organized as follows. In Sec. 2 the convex permeability models are introduced; in Sec. 2.1.1 we give a brief review of rarefaction fans, shock waves and properties of quadratic Corey models. Section 3 describes certain structures in state space; in Sec. 3.1 we identify features of the umbilic point and in Sec. 3.2 we describe curves with a certain equal shock speed property to the boundaries, the intersection of which is the ESS point. Finally, the conclusions are in Sec. 4.

2. **Mathematical model.**

Consider the flow of a mixture of three fluid phases (which, for concreteness, are called water, gas and oil) in a thin, horizontal cylinder of porous rock. Let $s_{\mathrm{w}}(x,t)$, $s_{\mathrm{g}}(x,t)$ and $s_{\mathrm{o}}(x,t)$ denote the respective saturations at distance x along the cylinder, at time t. Because $s_{\mathrm{w}} + s_{\mathrm{g}} + s_{\mathrm{o}} = 1$ and $0 \leq s_{\mathrm{w}},\, s_{\mathrm{g}},\, s_{\mathrm{o}} \leq 1$, the state space of the fluid mixture is the saturation triangle Δ; see *e.g.* Fig. 1. In our analysis, we choose s_{w} and s_{g} as the two independent variables, thus $S := (s_{\mathrm{w}},\, s_{\mathrm{g}})$; the vertices of Δ are $\mathrm{W} = (1,\,0)$, $\mathrm{G} = (0,\,1)$ and $\mathrm{O} = (0,\,0)$.

2.1. **Conservation laws.**

Three-phase flow in 1D at constant injected rate is governed by the non-dimensionalized system $\partial S/\partial t + \partial F(S)/\partial x = 0$, or

$$\frac{\partial s_{\mathrm{w}}}{\partial t} + \frac{\partial f_{\mathrm{w}}(s_{\mathrm{w}},\, s_{\mathrm{g}})}{\partial x} = 0, \tag{1}$$

$$\frac{\partial s_{\mathrm{g}}}{\partial t} + \frac{\partial f_{\mathrm{g}}(s_{\mathrm{w}},\, s_{\mathrm{g}})}{\partial x} = 0, \tag{2}$$

representing conservation of water and gas. The flow functions $f_{\mathrm{w}}(s_{\mathrm{w}},\, s_{\mathrm{g}})$ and $f_{\mathrm{g}}(s_{\mathrm{w}},\, s_{\mathrm{g}})$ are determined by the relative permeabilities of the three phases.

Although each fluid phase becomes immobile below an residual saturation, for simplicity we assume that the relative permeabilities are strictly positive within the saturation triangle. (In Engineering language, s_{w}, s_{g}, and s_{o} are "reduced saturations".) From Darcy's law the fluxes are

$$f_\alpha(S) = \frac{\mathrm{m}_\alpha(S)}{\mathrm{m}(S)}, \quad \text{for} \quad \alpha = \mathrm{w, g, o}, \quad \text{where} \quad \mathrm{m} := \mathrm{m}_{\mathrm{w}} + \mathrm{m}_{\mathrm{g}} + \mathrm{m}_{\mathrm{o}} \tag{3}$$

is the total mobility; m_{w}, m_{g}, m_{o} represent the relative mobility of each phase. Each mobility is a ratio between relative permeability and viscosity of the fluid, it is described by the continuous function:

$$\mathrm{m}_\alpha(S) := \frac{k_\alpha(S)}{\mu_\alpha}, \qquad \alpha = \mathrm{w, g, o}, \tag{4}$$

where μ_α is the given constant viscosity of each phase α.

A Corey type model is defined by a set of mobilities $\mathrm{m}_\alpha(s_\alpha)$ that are nondecreasing continuous functions of their own saturation s_α. In this work we focus on convex Corey models, which obey the following restrictions.

Definition 2.1. A Corey model is said to be **convex** when the mobilities are $\mathcal{C}^1[0,\,1]\cap\mathcal{C}^2(0,\,1)$ functions satisfying:

1. $\mathrm{m}_\alpha(s_\alpha) > 0$ for $s_\alpha \in (0,\,1]$ and $\mathrm{m}_\alpha(0) = 0$,
2. $\mathrm{m}'_\alpha(s_\alpha) > 0$ for $s_\alpha \in (0,\,1]$ and $\mathrm{m}'_\alpha(0) = 0$,
3. $\mathrm{m}''_\alpha(s_\alpha) \geq 0$ for $s_\alpha \in (0,\,1)$,
4. no pair of the quantities $\mathrm{m}''_\mathrm{w}(s_\mathrm{w})$, $\mathrm{m}''_\mathrm{g}(s_\mathrm{g})$, $\mathrm{m}''_\mathrm{o}(s_\mathrm{o})$ vanish simultaneously for any point in the interior of the saturation triangle ($0 < s_\mathrm{w},\, s_\mathrm{g},\, s_\mathrm{o} < 1$).

Remark 1. In the presence of nonzero residual saturations, one can easily formulate an appropriate extension of Definition 2.1.

Remark 2. In this work, for the purpose of illustrating facts with figures, we use the following mobilities:

$$\mathrm{m}_\mathrm{w}(s_\mathrm{w}) = (s_\mathrm{w})^{3.2857}/1, \quad \mathrm{m}_\mathrm{g}(s_\mathrm{g}) = (s_\mathrm{g})^{2.65}/0.5, \quad \mathrm{m}_\mathrm{o}(s_\mathrm{o}) = (s_\mathrm{o})^{5.8357}/2,$$

based on a best fit for homogeneous porous media of the Corey-Brooks model, [4].

2.1.1. *Basic solutions.* Equations (1)–(2) have solutions that propagate as waves. The Jacobian matrix of the fluxes is the key for rarefaction curves. The characteristic speeds are the two eigenvalues of the Jacobian derivative matrix

$$\mathrm{J}(S) := \frac{\partial(f_\mathrm{w}(S),\, f_\mathrm{g}(S))}{\partial(s_\mathrm{w},\, s_\mathrm{g})} = \frac{\partial F(S)}{\partial S}, \tag{5}$$

provided that these eigenvalues are real, in which case the smaller one is called the slow-family characteristic speed $\lambda_\mathrm{s}(s_\mathrm{w},\, s_\mathrm{g})$ and the larger one is called the fast-family characteristic speed $\lambda_\mathrm{f}(s_\mathrm{w},\, s_\mathrm{g})$. For the Corey model, both eigenvalues are real and nonnegative for each state in the saturation triangle.

The *self-similarity* of solutions of a Riemann problem implies that if $u(x,\, t)$ is such a solution at a given time t, then $u(\alpha x,\, \alpha t)$ is also a solution for any $\alpha > 0$. Centered rarefaction and shock waves are based on self-similarity.

System (1)–(2) has continuous solutions called slow- and fast-family rarefaction waves. They arise by solving an ODE, namely,

$$\{\mathrm{J}(S) - \xi\,\mathrm{I}\}\vec{r}(S) = 0, \qquad \frac{dS}{d\xi} = \vec{r}(S),$$

where $S(\xi)$, for $\xi = x/t$, is the profile of the rarefaction provided ξ is monotonic increasing. Some integral curves appearing in the solution of Riemann problems are plotted in Fig. 1.

This system also admits solutions that have jump discontinuities. The Hugoniot locus of a point S^o, denoted as $\mathcal{H}(S^o)$, is given by all the points S that satisfy the Rankine-Hugoniot (RH) condition:

$$F(S) - F(S^o) = \sigma(S - S^o), \tag{6}$$

where $\sigma = \sigma(S^o,\, S)$ is the velocity of the discontinuity, and the fluxes $F(S)$ and saturations S are given as before. (Notice that S belongs to $\mathcal{H}(S^o)$ if and only if S^o belongs to $\mathcal{H}(S)$.) Admissibility of discontinuites for systems of conservation laws such as (1)–(2) is discussed in [2].

Notice that if the RH condition between states S^a and S^o holds with a certain speed σ, and it also holds for the same speed between states S^o and S^b, it is easy

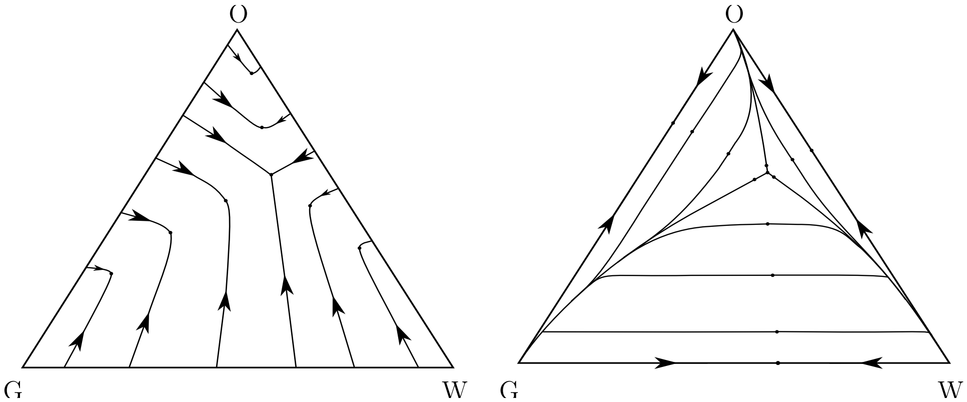


FIGURE 1. Integral curves; slow and fast families. The triple intersection is the umbilic point. Dots on integral curves are inflections, arrows point in the increasing eigenvalue direction. (The specific mobilities are in Remark 2.)

to see that the RH condition is satisfied between states S^a and S^b with the same speed. This is the essence of the triple-shock rule [9]. The definition of σ_{ij} as the shock speed $\sigma(S_i,\, S_j)$ will be useful.

Theorem 2.2 (Triple-shock rule)**.** *Let the states S_1, S_2 belong to $\mathcal{H}(S_0)$. If $\sigma_{01} = \sigma_{02}$ holds, then S_1 belongs to $\mathcal{H}(S_2)$ and the relations $\sigma_{01} = \sigma_{02} = \sigma_{12}$ hold.*

Proof. Define σ as $\sigma_{01} = \sigma_{02}$. Subtract versions of equation (6) written for $(S_0,\, S_1)$ and for $(S_0,\, S_2)$, obtaining $F(S_2) - F(S_1) = \sigma(S_2 - S_1)$, which indicates that S_1 belongs to $\mathcal{H}(S_2)$ and σ_{12} is equal to σ. □

The following variant of Theorem 2.2 has been used in several works appearing in this conference.

Lemma 2.3. *Let S_0, S_1, S_2 be non-collinear states such that S_1, S_2 belong to $\mathcal{H}(S_0)$ and S_1 belongs to $\mathcal{H}(S_2)$. Then $\sigma_{01} = \sigma_{02} = \sigma_{12}$ holds.*

Proof. Let us express the RH relations of the involved states; we have

$$\begin{aligned} F(S_1) - F(S_0) \;&=\; \sigma_{01}(S_1 - S_0), \quad F(S_2) - F(S_0) \;=\; \sigma_{02}(S_2 - S_0),\\ F(S_1) - F(S_2) \;&=\; \sigma_{12}(S_1 - S_2). \end{aligned} \tag{7}$$

By subtracting (7.b) and (7.c) from (7.a), we obtain

$$0 = \sigma_{01}(S_1 - S_0) - \sigma_{02}(S_2 - S_0) - \sigma_{12}(S_1 - S_2).$$

We subtract the trivial relation $0 = \sigma_{12}(S_1 - S_0) - \sigma_{12}(S_2 - S_0) - \sigma_{12}(S_1 - S_2)$ obtaining $0 = (\sigma_{01} - \sigma_{12})(S_1 - S_0) - (\sigma_{02} - \sigma_{12})(S_2 - S_0)$. Recalling that the sates are non-collinear, we notice that the latter relation holds if and only if $\sigma_{01} - \sigma_{12}$ and $\sigma_{02} - \sigma_{12}$ are zero, which proves the lemma. □

A system is called strictly hyperbolic if the characteristic speeds satisfy the inequality $\lambda_{\mathrm{s}}(S) < \lambda_{\mathrm{f}}(S)$ everywhere; they are well studied [10, 11]. In three-phase flow models there are points where the characteristic speeds coincide, which are called coincidence points. Furthermore, in Corey models there are isolated coincidence points where the Jacobian matrix is a multiple of the identity, *i.e.*, umbilic points.

The quadratic Corey model is defined by the permeabilities $k_\alpha(S) = s_\alpha^2$ for $\alpha = \mathrm{w, g, o}$. Such a model is well understood; in particular, the location and characteristics of umbilic points are well known. There is a unique umbilic point $U = (u_\mathrm{w},\, u_\mathrm{g})$ in the interior of Δ, with $u_\mathrm{o} = 1 - u_\mathrm{w} - u_\mathrm{g}$, the coordinates of which are

$$u_\alpha \;=\; \mu_\alpha/(\mu_\mathrm{w} + \mu_\mathrm{g} + \mu_\mathrm{o}), \qquad \text{for} \qquad \alpha = \mathrm{w, g, o}.$$

Such a point satisfies the following.

Property 2.4. *For the quadratic Corey model, the characteristic speeds are equal to* 2 *at the interior umbilic point.*

Three other umbilic points lie on the vertices of the saturation triangle.

Property 2.5. *For the quadratic Corey model, the shock speed from the interior umbilic point to vertices of the triangle are equal to 1.*

3. Structures in the saturation triangle for convex Corey models.

When two of the permeabilities in (4) cease to be scalar multiples of the same convex function, the umbilic point gives rise to two points: the first one is still an umbilic point, and Property 2.4 holds, and at the second one, only Property 2.5 holds. It is because of the shock speed equality that the latter point will be called Equal-Speed Shocks to vertices or ESS.

3.1. The umbilic point location.

Inmiscible three-phase flow models are typically non-strictly hyperbolic, except in the model in [7]. Lemma 3.1 follows from results in [14] for the case where the gravity force is not active. (In [8, 15] there are shorter proofs.)

Lemma 3.1. *Consider a convex Corey permeability model, see Definition 2.1. There is always a single point U in the interior of the saturation triangle satisfying*

$$\mathrm{m}'_\mathrm{w}(u_\mathrm{w}) = \mathrm{m}'_\mathrm{g}(u_\mathrm{g}) = \mathrm{m}'_\mathrm{o}(u_\mathrm{o}), \tag{8}$$

which is the unique umbilic point in the interior of the triangle. It has characteristic speed $\lambda(U) = \mathrm{m}'_\mathrm{w}(u_\mathrm{w})/\mathrm{m}(U)$.

An important feature of the models considered is the following: from properties (3) and (4) of Definition 2.1, one can see that the Hessian for the total mobility:

$$\begin{pmatrix} \mathrm{m}''_\mathrm{w} + \mathrm{m}''_\mathrm{o} & \mathrm{m}''_\mathrm{o} \\ \mathrm{m}''_\mathrm{o} & \mathrm{m}''_\mathrm{g} + \mathrm{m}''_\mathrm{o} \end{pmatrix} \tag{9}$$

is a positive definite matrix. Hence the motivation of the following result.

Corollary 1. *For a convex Corey type model, the total mobility has a single extremum in the interior of the triangle, which occurs at the umbilic point. The extremum is a minimum.*

Proof. Equating to zero the partial derivatives of m in (3.b) relatively to s_w and s_g implies $\mathrm{m}'_\mathrm{w} = \mathrm{m}'_\mathrm{o}$ as well as $\mathrm{m}'_\mathrm{g} = \mathrm{m}'_\mathrm{o}$; then Lemma 3.1 guarantees that this extremum occurs at the single umbilic point. Thus from the positive definiteness of (9) we obtained that this extremum is the minimum. □

Remark 3. Darcy's law says that the total flow rate of a fluid mixture is proportional to the pressure gradient; the proportionality coefficient is (minus) the total mobility. Corollary 1 implies that maximum pressure gradient is needed to displace the fluid mixture at saturations given by the umbilic point, for a given total flow rate. In other words the umbilic point gives the saturation proportion for which each of the three fluids hinders maximally the flow of the other two. (Total flow is minimal for a specific pressure gradient.)

We will call $\mathrm{m}'_{\mathrm{w}}(s_{\mathrm{w}})$ the sensitivity of the water mobility to water saturation. The first equality in (8), $\mathrm{m}'_{\mathrm{w}}(s_{\mathrm{w}}) = \mathrm{m}'_{\mathrm{g}}(s_{\mathrm{g}})$, defines the *equal water-gas sensitivity curve*, which can be parametrized either as a function of s_{w} or s_{g}; it contains U and O. Similarly we can define equal water-oil and gas-oil sensitivity curves. See the three dashed curves in Fig. 2. (In the absence of gravitational force these curves were called two-phase-like-flow sets in [14].)

Let us summarize properties of the equal sensitivity curves. First of all, recall that $\mathrm{m}'_{\mathrm{w}} = \mathrm{m}'_{\mathrm{g}}$ implies $\partial\mathrm{m}/\partial s_{\mathrm{w}} = \partial\mathrm{m}/\partial s_{\mathrm{g}}$, for brevity we call $\partial\mathrm{m}$ such a value, thus the Jacobian matrix at any point of the equal water-gas sensitivity curve is

$$\mathrm{J}(S) = \frac{1}{\mathrm{m}^2}\begin{pmatrix} \mathrm{m}'_{\mathrm{w}}\mathrm{m} - \mathrm{m}_{\mathrm{w}}\partial\mathrm{m} & -\mathrm{m}_{\mathrm{w}}\partial\mathrm{m} \\ -\mathrm{m}_{\mathrm{g}}\partial\mathrm{m} & \mathrm{m}'_{\mathrm{w}}\mathrm{m} - \mathrm{m}_{\mathrm{g}}\partial\mathrm{m} \end{pmatrix}.$$

Along the curve one eigenvalue is $\lambda = \mathrm{m}'_{\mathrm{w}}/\mathrm{m}$ with eigenvector $(1,\,-1)$ (in Cartesian coordinates), which is parallel to the side $s_{\mathrm{o}} = 0$.

Moreover, the total mobility is minimum on the equal sensitivity curve in the direction of such eigenvector. Indeed, $\nabla\mathrm{m}\cdot(1,\,-1) = \partial\mathrm{m}/\partial s_{\mathrm{w}} - \partial\mathrm{m}/\partial s_{\mathrm{g}}$ is zero on the sensitivity curve, which turns out to be at a minimum because the Hessian in (9) is positive definite. (Analogous statements hold for other sensitivity curves.)

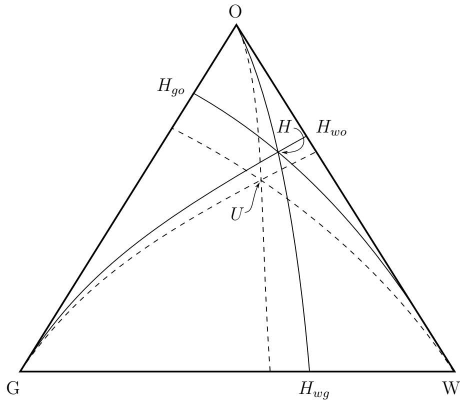


FIGURE 2. Location of umbilic and ESS points. Solid curves are Hugoniot loci from pure saturations. The umbilic location is given from similar dashed curves.

Remark 4. For non-convex Corey models we have the following facts. A converse to Lemma 3.1 holds: an umbilic point in the interior of the triangle satisfies (8). Instead of Corollary 1, every extremum of the total mobility is a coincidence point; such a point is umbilic provided that the second derivatives of two of the mobilities do not vanish simultaneously there. (The extrema are not necessarily unique and do not need to be minima.)

3.2. **The equal-speed shocks curves.** Let us consider the vertex $\mathrm{O} = (0,\,0)$, and look for points $S = (s_{\mathrm{w}},\,s_{\mathrm{g}})$ in Δ satisfying RH relation (6):

$$f_{\mathrm{w}}(S) = \sigma s_{\mathrm{w}}, \qquad f_{\mathrm{g}}(S) = \sigma s_{\mathrm{g}}; \tag{10}$$

where we used the fact that water and gas saturations for pure oil are zero, water and gas permeabilities are also zero. For the same reason the sides $s_{\mathrm{w}} = 0$ and $s_{\mathrm{g}} = 0$ are part of the Hugoniot locus of O. A third solution appears equating σ in Eqs. (10) leading to

$$\sigma(S,\, \mathrm{O}) \;=\; \frac{f_{\mathrm{w}}(S)}{s_{\mathrm{w}}} \;=\; \frac{f_{\mathrm{g}}(S)}{s_{\mathrm{g}}}; \tag{11}$$

points S satisfying the last equality in (11) form a curve inside Δ.

We denote by $\mathcal{H}_i(\mathrm{O})$ the locus in Δ that satisfies Eq. (11), *i.e.*, the "interior Hugoniot locus" from O; the Hugoniot locus of the vertex O is given by $\mathcal{H}_i(\mathrm{O})$ and the sides WO and GO. Since for any state S on $\mathcal{H}_i(\mathrm{O})$, $\mathcal{H}(S)$ intersects both sides WO and GO, see [5], from Lemma 2.3 we have the following

Claim 3.2. *All points in the internal Hugoniot locus* $\mathcal{H}_i(\mathrm{O})$ *satisfy the triple-shock rule between* O *and points on the boundary* WO*; they also satisfy the triple-shock rule between* O *and points on the boundary* GO*.*

We define $\mathcal{H}_i(\mathrm{O})$, from equality (11), as the *equal water-gas shock speed curve* (as we will show presently), which can be parametrized either as a function of s_{w} or s_{g}. Actually, since each $\mathrm{m}_\alpha(s_\alpha)$ is an increasing continuous function, its inverse is well defined and increasing. With aid of the constraint $s_{\mathrm{w}} + s_{\mathrm{g}} + s_{\mathrm{o}} = 1$, it is easy to see that points $(s_{\mathrm{w}},\, s_{\mathrm{g}})$ satisfying the second equality in relation (11) can be parametrized by s_{o}, *i.e.*, there exist smooth functions

$$H_{\mathrm{w}},\, H_{\mathrm{g}} : [0,\, 1] \to [0,\, 1] \qquad \text{s.t.} \qquad (H_{\mathrm{w}}(s_{\mathrm{o}}),\, H_{\mathrm{g}}(s_{\mathrm{o}})) \in \mathcal{H}_i(\mathrm{O}), \tag{12}$$

for all $s_{\mathrm{o}} \in [0,\, 1]$; notice that H'_{w} and H'_{g} are negative because when s_{o} increases $s_{\mathrm{w}} + s_{\mathrm{g}}$ decreases. Similarly we can define equal water-oil and gas-oil shock speed curves; $\mathcal{H}_i(\mathrm{G})$ and $\mathcal{H}_i(\mathrm{W})$.

The intersection of $\mathcal{H}_i(\mathrm{W})$, $\mathcal{H}_i(\mathrm{G})$ and $\mathcal{H}_i(\mathrm{O})$, is denoted by $H := (h_{\mathrm{w}},\, h_{\mathrm{g}})$, with $h_{\mathrm{o}} = 1 - h_{\mathrm{w}} - h_{\mathrm{g}}$, and satisfies

$$\sigma \;=\; \frac{f_{\mathrm{w}}(H)}{h_{\mathrm{w}}} \;=\; \frac{f_{\mathrm{g}}(H)}{h_{\mathrm{g}}} \;=\; \frac{f_{\mathrm{o}}(H)}{h_{\mathrm{o}}}. \tag{13}$$

This is the ESS point (Equal-Speed Shocks); the shock speeds from H to any vertex have the same value σ. Notice from relations (13) that H satisfies $\sigma = \Sigma_\alpha f_\alpha(H)/\Sigma_\alpha h_\alpha = 1$. Defining H_{wg}, H_{wo}, H_{go} as the intersection of the internal Hugoniot $\mathcal{H}_i(\mathrm{O})$, $\mathcal{H}_i(\mathrm{G})$, $\mathcal{H}_i(\mathrm{W})$ with the sides WG, WO, GO respectively (see Fig. 2), we notice that the triple-shock rule (see Theorem 2.2) holds with speed one for seven points, namely,

$$\sigma(A,\, B) \;=\; 1, \qquad \text{with} \qquad A,\, B \in \{H,\, \mathrm{W},\, \mathrm{G},\, \mathrm{O},\, H_{\mathrm{wg}},\, H_{\mathrm{wo}},\, H_{\mathrm{go}}\},$$

since each point belongs to the Hugoniot locus of the three vertices.

4. Concluding remark.

The internal Hugoniot loci of the vertices give rise to the ESS point, while the equal sensitivity curves give rise to the umbilic point.

As in [15] one can follow the ordering of increasing directions of fast rarefaction curves near the boundary, see Fig. 1, and notice that there is an orientation reversal, thus a quadratic expansion of the fluxes about the umbilic point shows that in our case it must be classified as Type I or II, see Fig. 3 and [13].

Acknowledgments. The authors are grateful to Prof. J. Bruining (TU-Delft) for many enlightening discussions.

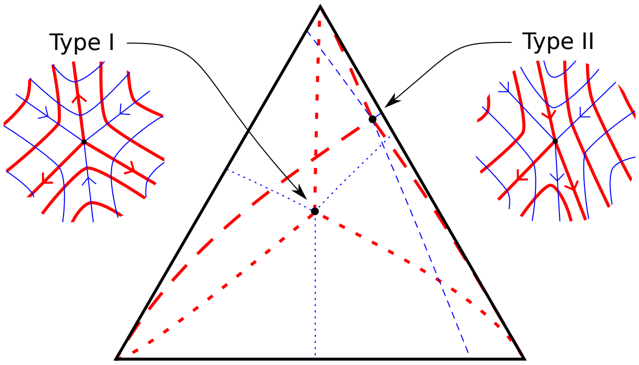


FIGURE 3. In the saturation triangle there are two possible umbilic point types for Corey permeability models with different viscosities. We represent the two possibilities. The rarefaction behavior around the umbilic type is sketched in the small insets. (Lighter curves represent slow family, darker curves represent fast family.)

REFERENCES

[1] F. Asakura (2012) "Stone-Marchesin model equations of three-phase flow in oil reservoir simulation", *Proceedings of HYP2012* (Padova, Italy).

[2] A.V. Azevedo, A. de Souza, F. Furtado, D. Marchesin, B. Plohr (2010) "The solution by the wave curve method of three-phase flow in virgin reservoirs", *TiPM* **83**: 99–125.

[3] A.V. Azevedo, D. Marchesin, B. Plohr and K. Zumbrun (2002) "Capillary instability in models for three-phase flow", *Z. Angew. Math. Phys.* **53**: 713–746.

[4] J. Bruining (2007) *Muiltiphase Flow in Porous Media.* TU-Delft, Lecture notes.

[5] P. Castañeda, F. Furtado and D. Marchesin (2013) "The convex permeability three-phase flow in reservoirs". *IMPA Preprint Série E* – **2258**: 1–34.

[6] P. Castañeda, E. Abreu, F. Furtado and D. Marchesin "The Riemann problem for convex permeability three-phase flow in virgin reservoirs". (*In preparation.*)

[7] R. Juanes and T. Patzek (2004) "Relative permeabilities for strictly hyperbolic models of three-phase flow in porous media", *Transp. Porous Media* **57**: 125–152.

[8] M.E. Gomes (1987) *Singular Riemann Problem for a Fourth-order Model for Multi-Phase Flow*, DSc Thesis, PUC-RJ. (In Portuguese.)

[9] E. Isaacson, D. Marchesin, B. Plohr and B. Temple (1992) "Multiphase flow models with singular Riemann problems", *Mat. Apl. Comput.* **11**: 147–166.

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

[11] T.-P. Liu (1975) "The Riemann problem for general systems of conservation laws", *J. Differential Equations* **18**: 218–234.

[12] D. Marchesin and B. Plohr (2001) "Wave structure in WAG recovery", *SPEJ* **6**: 209–219.

[13] V. Matos, P. Castañeda and D. Marchesin (2012) "Classification of the umbilic point in immiscible three-phase flow in porous media", *Proceedings of HYP2012* (Padova, Italy).

[14] H. Medeiros (1992) "Stable hyperbolic singularities for three-phase flow models in oil reservoir simulation", *Acta Appl. Math.* **28**: 135–159

[15] M. Shearer (1988) "Loss of strict hyperbolicity for the Buckley-Leverett equations of three phase flow in a porous medium", *Numerical simulation in oil recovery* IMA **11**: 263–283.

[16] M. Shearer and J. Trangenstein (1989) "Loss of real characteristics for models of three-phase flow in a porous medium", *Transp. Porous Media* **4**: 499–525.

E-mail address: castaneda@impa.br
E-mail address: marchesi@impa.br
E-mail address: furtado@uwyo.edu

LINEARLY IMPLICIT SCHEMES FOR CONVECTION-DIFFUSION EQUATIONS

Fausto Cavalli

University of Milan, Via Saldini 50
20133 Milan, Italy

Abstract. We present a family of schemes for the approximation of one dimensional convection-diffusion equations. It is based on a linearization technique that allows to treat explicitly the hyperbolic term and linearly implicitly the parabolic one. This avoids the parabolic stability constraint $\Delta t \leq ch^2$ of explicit schemes, and does not require any non-linear solver for the implicit problem. We present several numerical simulations to show the effectiveness of the proposed schemes and to investigate their stability, convergence and accuracy. In particular, since the proposed schemes provide to be accurate for both smooth and non-smooth solutions, they turn out to be attractive for adaptivity.

1. **Introduction.** In this paper we investigate the numerical behaviour of a family of schemes for the one dimensional non-linear convection-diffusion equation

$$\begin{aligned} &\partial_t u + \partial_x f(u) = \partial_{xx} p(u),\ (x,t) \in [a,b] \times [0,T], \\ &u(x,0) = u_0(x), \qquad x \in [a,b], \end{aligned} \tag{1}$$

where $p(u)$ is a non-decreasing regular function with Lipschitz constant L_p and $p(0) = 0$. For easy, we will just consider homogeneous Dirichlet boundary conditions, but other conditions can be taken into account as well.

Equation (1) is particularly challenging due to the presence of both the hyperbolic term $f(u)$ and the nonlinear parabolic term $p(u)$, in particular when the diffusion $p(u)$ is degenerate, i.e. when the derivative $p'(u)$ vanishes for some values. In this paper we will just consider the case of $p'(u) = 0$ for isolated values of u, as in the porous media equation with $p(u) = u^m$, $m > 1$, for which $p'(0) = 0$. The strongly degenerate case ($p'(u) = 0$ on whole intervals) is not considered for now.

When dealing with the approximation of equation (1), the hyperbolic term $f(u)$ is usually treated explicitly in time, while the spatial approximation is performed using the non-linear techniques developed for conservation laws, such as slope-limiters or ENO/WENO reconstructions (see for example [10]). The parabolic term can be handled explicitly too, and the non-linear approximation techniques used for the convection term provided to be very effective also in this case, as shown in [7, 11, 6]. The main drawback of fully explicit schemes lies in the constraint $\Delta t \leq ch^2$ that must be imposed on the time step Δt with respect to the grid size h to guarantee the stability of the parabolic term. On the other hand, non-linear implicit approximations of the parabolic term require non-linear iterative solvers.

2000 *Mathematics Subject Classification.* Primary: 65M06; Secondary: 35K65.

Key words and phrases. Convection-diffusion, degenerate parabolic problems, linearly implicit schemes, finite differences.

To overcome these problems we propose a linearization of the parabolic term, similar to that introduced in [4], in order to obtain numerical schemes that avoid non-linear implicit problems, and then we apply an implicit-explicit (IMEX) Runge-Kutta method to the convection-diffusion equation. Several numerical schemes for parabolic equations rest upon a similar linearization technique, as [12, 9, 13, 14], but they are usually first order approximations. Conversely, a high order approach turns out to be particularly attractive for adaptivity, since both the resolution of the space-time discretization and the order of the scheme can be modified according to the local regularity of the solution.

2. **Numerical schemes.**

Let us introduce $t^n = n\Delta t$, where Δt is a uniform time step, and, for the space discretization, let us consider the uniform grid $x_i = a+h/2+(i-1)h,\ i = 1, ..., N$ on $[a, b]$ where $h = (b-a)/N$. Finally, let us introduce u_i^n, which is the approximation of $u(x, t)$ in $x = x_i$ at time $t = t^n$ and set $u_h^n = (u_i^n)^T_{i=1,\dots,N}$.

The family of schemes we are going to introduce is based on an implicit-explicit (IMEX) Runge-Kutta time integration. In particular, we consider the IMEX(s,$s+1$,r) schemes proposed in [1]. These schemes can be represented through a pair of Butcher tableau's which describe the s-stages diagonally implicit scheme and the $(s+1)$-stages explicit scheme. Moreover, the coefficients $\tilde{a}_{i,j}, \tilde{b}_i$ of the implicit scheme and the coefficients $a_{i,j}, b_i$ of the explicit scheme are chosen so that the resulting scheme converges with order r.

For the space approximation we use finite difference methods. We detail the method only for the internal grid points, as the boundary conditions are treated in a standard way. For further details about the implementation of boundary conditions, we refer to [4] for the implicit parabolic term and to [8] for the convection term. For the discretization of ∂_{xx}, let us introduce the linear operator $\mathcal{L}_h : \mathbb{R}^N \to \mathbb{R}^N$, which is

$$(\mathcal{L}_h u_h)_i = \frac{u_{i+1} - 2u_i + u_{i-1}}{h^2}, \tag{2}$$

for a second order approximation, and

$$(\mathcal{L}_h u_h)_i = \frac{-u_{i+2} + 16u_{i+1} - 30u_i + 16u_{i-1} - u_{i-2}}{12h^2}, \tag{3}$$

for a fourth order one. For the convection flux, let us consider the operator $\mathcal{D}_h : \mathbb{R}^N \to \mathbb{R}^N$, defined by

$$(\mathcal{D}_h u_h)_i = \frac{\hat{F}_{i+1/2}(u_h) - \hat{F}_{i-1/2}(u_h)}{h},$$

where $\hat{F}$ is a numerical flux. In our simulations we used Lax-Friedrichs flux

$$\hat{F}_{i+1/2}(u_h) = \frac{f(u^+_{i+1/2}) + f(u^-_{i+1/2})}{2} - a\frac{u^+_{i+1/2} + u^-_{i+1/2}}{2}, \tag{4}$$

where $u^{\pm}_{i+1/2}$ are reconstructions of suitable order at the cell boundaries and $a = \max_u |f'(u)|$. Finally let us introduce

$$\xi(u^n) = p'(u^n) + \alpha^n, \tag{5}$$

where α^n is a constant in space, positive, correction term.

The family of schemes we present generalizes the first order scheme introduced in [2] for purely diffusive equations

$$\begin{aligned} q_h^n &= \frac{p(u_h^n)}{\xi}, \\ q_h^{n+1} &= q_h^n + \Delta t \mathcal{L}_h q_h^{n+1}, \\ u_h^{n+1} &= u_h^n + q_h^{n+1} - q_h^n, \end{aligned} \tag{6}$$

where ξ is a constant such that $\xi \geq \max_u p'(u)$. Our high order generalization of scheme (6) for the convection-diffusion problem (1) is based on the following steps: first we perform a linearization of the diffusion term introducing the variable q_h^n

$$q_h^n = \frac{p(u_h^n)}{\xi(u_h^n)}. \tag{7a}$$

Then, defining the values $q_h^{(0)} = q_h^n$ and $u_h^{(0)} = u_h^n$, we perform each stage of the IMEX scheme,

$$q_h^{(i)} = q_h^n + \Delta t \sum_{j=1}^{i} a_{i,j} \mathcal{L}_h(\xi(u_h^n) q_h^{(j)}) - \Delta t \sum_{j=0}^{i-1} \tilde{a}_{i+1,j+1} \mathcal{D}_h(u_h^{(j)}),\ i = 1, \dots, s \tag{7b}$$

where we treated in implicit the *linear* diffusion equation in $q_h^{(i)}$ and in explicit the convection term. At the end of each stage, we can reconstruct $u_h^{(i)}$ through

$$u_h^{(i)} = u_h^n + q_h^{(i)} - q_h^n. \tag{7c}$$

Finally, the reconstruction stage of the IMEX method allows to obtain

$$q_h^{n+1} = q_h^n + \Delta t \sum_{i=1}^{s} b_i \mathcal{L}_h(\xi(u_h^n) q_h^{(i)}) - \Delta t \sum_{i=0}^{s} \tilde{b}_{i+1} \mathcal{D}_h(u_h^{(i)}), \tag{7d}$$

and then to find the updated solution at time t^{n+1}

$$u_h^{n+1} = u_h^n + q_h^{n+1} - q_h^n. \tag{7e}$$

Let us make some considerations about (7). In (6) the linearization is performed using a constant ξ which approximates $p'(u_h^n)$ on its whole domain and stability is guaranteed under the condition $\xi \geq L_p$. Moreover, the evolution of q is performed using backward Euler method. Since ξ only provides a global approximation of $p'(u)$, we have that scheme introduced (6) is only first order accurate, even if we use a higher order scheme for the evolution of q^n, and it is not very accurate near discontinuities.

The scheme introduced in [2] was enhanced in [9, 13], where local versions of ξ were considered to improve the accuracy of the method. The ideal choice of ξ would be $\xi(u^n) = p'(u^n)$, but unfortunately the resulting scheme would be in general unstable. A solution is to introduce, like in (5), a positive correction. We remark that in all the previous works the authors considered only first order schemes.

In the matter of the correction term α^n , let us underline that if we chose α^n too small, instabilities would arise in the numerical solution, while too large α^n would deteriorate the accuracy of the scheme. We are studying in [3] a way to obtain optimal values of α^n from the solution u_h^n, together with its influence on the stability of the solution. In this work we do not detail explicitly the strategy to choose α^n . We will only plot the values of α^n we chose for the scheme to show that they guaranteed stability and accuracy. We remark that, to improve accuracy, also non-constant in space corrections $\alpha^n(x)$ can be considered.

The last remark on scheme (7) is about its convergence rate. It is easy to see that the time consistency error of the above schemes can be 2 only if $\alpha = O(\Delta t)$. It is worth noticing that, the family of schemes (7) does not allow approximations of order higher than 2, even if both the IMEX scheme and the spatial approximation have a higher order of accuracy. This is mainly due to the linearization of $p(u)$ through $\xi(u_h^n)$. Schemes of order higher than 2 are under consideration and some preliminary results shows how also the time integration need to be modified to obtain third order accuracy. Finally, for non-smooth solutions, it is necessary that α^n stay bounded as $\Delta t \to 0$, to avoid that increasingly large corrections destroy the accuracy of the scheme on fine grids.

3. **Numerical tests.** We consider three different schemes of the family (7). The first order scheme (identified in the following by L1), is obtained using IMEX(1,1,1) scheme for the time integration while the spatial approximation relies on constant reconstructions for the flux (4) and operator (2) for the discretization of ∂_{xx}. The second order scheme (L2) is obtained using IMEX(2,3,2) scheme, linear ENO reconstructions and again operator (2). The third scheme (L3) is obtained using IMEX(3,4,3) scheme, parabolic ENO reconstructions and operator (3). We remark that scheme L3 would give rise to a third order accurate scheme in absence of the linearization error, but actually it is only second order accurate. We will also consider for comparison the solutions of the explicit hyperbolic/non-linear implicit parabolic scheme obtained applying directly the IMEX(s,$s+1$,r) scheme to (1) without linearization. In this case, the final non-linear problem is solved by Newton iterative method and the schemes are identified by NL1,NL2 and NL3.

In the first simulations, we consider equation (1) with the hyperbolic Burger flux $f(u) = u^2$ and the porous media diffusion flux $p(u) = u^3$, and we test the scheme starting with the C^2 initial datum

$$u(x,0) = \cos^2\left(\frac{\pi}{2}x\right)\chi_{[-1,1]}, \quad x \in [-3/2, 3/2], \tag{8}$$

and with the discontinuous one

$$u(x,0) = 5\chi_{[-1/2,1/2]}. \tag{9}$$

Since analytic solutions are not available, in both cases the numerical approximations obtained with L1,L2 and L3 are compared with the solutions computed with NL3 on a very fine grid. We observe that the solution obtained evolving (8) stays regular for a small time, so in this case we can also study the convergence rate. In Table 1 we report the L^1 norm of the relative errors (E_1) and the estimated convergence rates (r) of the numerical solutions of problem (8) obtained at $t = 0.01$. As expected, scheme L1 is first order accurate, while schemes L2 and L3 are second order methods. We also note that L3 is more accurate than L2.

In Table 2 we compare the errors obtained with the three linear schemes for the initial datum (9) on a coarse and a fine grid. As we can see, we can take benefit from increasing the order of the approximation also in the case of a non-smooth solution.

As we remarked in Section 2, the constant α^n plays an important role in the stability and the accuracy of the schemes (7). In Figure 1 we plot the values of α^n we used in the previous simulations for both the smooth problem (8) and the non-smooth problem (9) for a grid with $N = 270$ cells and for different choices of

	L1		L2		L3	
N	E_1	r	E_1	r	E_1	r
10	1.86e-01		1.86e-01		1.86e-01	
30	2.25e-02	1.92	8.25e-03	2.84	6.14e-03	3.11
90	6.98e-03	1.07	6.61e-04	2.30	2.72e-04	2.84
270	2.04e-03	1.12	5.89e-05	2.20	1.15e-05	2.88
810	6.80e-04	1.00	5.90e-06	2.09	1.16e-06	2.09
2430	2.27e-04	0.99	6.32e-07	2.03	1.34e-07	1.96

TABLE 1. Comparison of the L^1 norms (E_1) of the relative errors and of the estimated convergence rates (r) for the solution of problem (9) on several grids with N cells. The expected convergence rates are achieved.

N	L1	L2	L3
90	9.22e-03	2.52e-03	1.94e-03
810	8.04e-04	8.32e-05	6.21e-05

TABLE 2. Comparison of the L^1 norms of the relative errors for the solution of problem (9) with schemes L1,L2 and L3 on two different grids.

Δt. We report only the results for L3, since those for L1 and L2 are very similar. As we can see, the values of α^n at each time step decreases as Δt decreases. Moreover, for the non-smooth problem we have that α^n stays bounded with respect to Δt, also at the earlier simulation times when the solution is less regular.

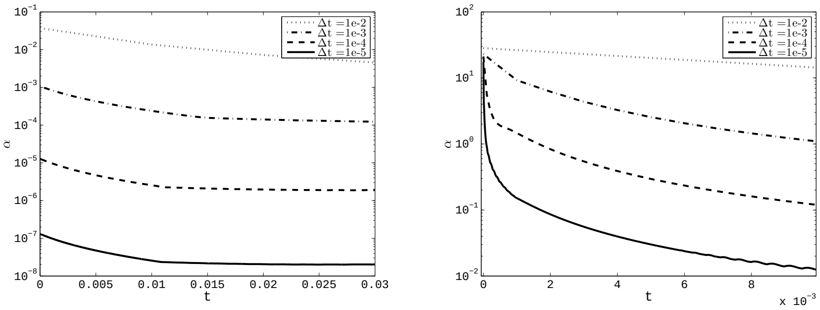


FIGURE 1. Evolution of α^n for the problem with the smooth initial datum (8) (left plot) and with the non-smooth initial datum (8) (right plot).

When the solution is smooth, it is also important that $\alpha^n = O(\Delta t)$ to grant second order accuracy, as remarked in Section 2. In Figure 2 we plot the behaviour of α^n at $t = 0.01, 0.02, 0.03$ for different choices of Δt and we can see that actually $\alpha^n = O(\Delta t^2)$.

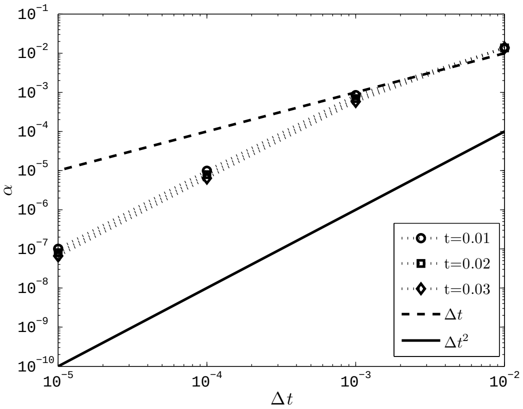


FIGURE 2. Evolution of α^n with respect to Δt at times $t = 0.01, 0.02, 0.03.$ in the simulation of problem (8)

We remark that we obtained similar results to those reported in Figures 1 and 2 when we considered $\Delta t = ch$ and studied α^n versus h for increasingly smaller values of h. Finally, we compare in Figure 3 the errors obtained with L1,L2 and L3 with those obtained with NL1,NL2 and NL3. As we can see, the linear and the non-linear scheme provide almost the same accuracy both in the case of regular and non-regular solutions. Only scheme NL3 is noticeably more accurate than L3 for the smooth problem (8) but this is quite predictable since NL3 is a true third order scheme, unlike L3 which is actually a second order scheme. Even if the regularity of the solution allows only a second order convergence rate, scheme NL3 can take advantage of its higher accuracy, especially on finer grids.

We remark that even if non-linear schemes are slightly more accurate, the approach we proposed can still be considered very effective since it is much less expensive, especially on finer grids, where the number of iterations required by the iterative method increases and the restrictions on Δt to guarantee the convergence can be quite constraining. If we wanted to compare two computationally equivalent approaches we would need to consider only one iteration per time step in the Newton method or to suitably decrease the time step Δt in the linear methods. In both the situations, schemes (7) provide better accuracy than the non-linear scheme. More details about the comparison of the two approaches can be found in [4] for diffusion equations, since the results for the convection-diffusion problem are similar. In the last simulation we consider the non-convex Buckley-Leverett flux

$$f(u) = \frac{u^2}{u^2 + (1-u)^2}, \tag{10a}$$

and the diffusion function

$$p(u) = 10^{-2}(2u^2 - \frac{4}{3}u^3), \tag{10b}$$

which is doubly degenerate, since $p'(u) = 0$ for $u = 0$ and $u = 1$. The initial datum is the step function

$$u(x,0) = \chi_{[1/2,3/4]},\ x \in [0,1].$$

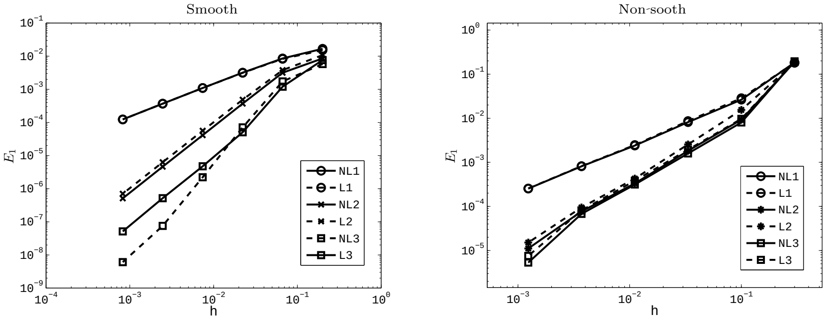


FIGURE 3. Comparison of the errors obtained with the linear schemes L1, L2 and L3 and with the non-linear ones NL1,NL2 and NL3 for the smooth problem (8) (left plot) and the non-smooth problem (9) (right plot).

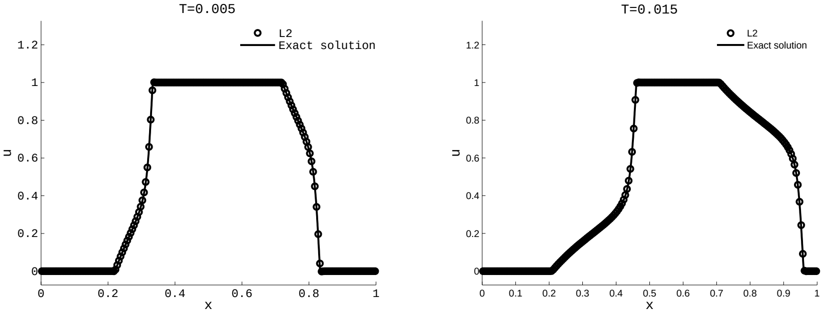


FIGURE 4. Numerical solution (circle) of problem (10) at time $t = 0.005$ (left plot) and at time $t = 0.015$ (right plot). The approximations are in accordance with the “exact” solution (solid line), computed with an entropic numerical scheme on a fine grid.

In particular, since the convective flux is non-convex, we want to check that the schemes we introduced be able to pick the correct entropic solution. In Figure 3 we plot the approximate solution at times $t = 0.005$ and $t = 0.015$ obtained with L3 and the “exact solution” obtained with a high order, explicit, entropic scheme on a very fine grid. Also in this case, the linear scheme is very accurate and approximates the correct entropic solution.

4. **Conclusion and perspectives.** We presented a family of schemes for the solution of convection-diffusion equations, based on the explicit discretization of the hyperbolic term and on the linearly implicit approximation of the parabolic term. The schemes we introduced are accurate for both smooth and non-smooth solutions and provide approximations which are comparable to those obtained with a non-linear scheme, being at the same time less expensive since each time step does not require iterative methods. Up to now, due to the linearization, we are able to achieve only second order accuracy, but we are looking for different time integration techniques to recover higher order of accuracy. In a work in preparation [3], we are

investigating the stability of the scheme with respect to the choice of the constant α^n. The final goal is to develop an adaptive scheme based on the family of schemes (7), to take advantage of the freedom of the spatial and temporal accuracy they provide. We also plan to investigate, as in [5], approximations with finite element methods, to treat also multidimensional problems on non-cartesian geometries.

REFERENCES

[1] U. Asher, S. Ruuth, and R.J. Spiteri, *Implicit-explicit Runge-Kutta methods for time dependent Partial Differential Equations*, Appl. Numer. Math., **25** (1997), 151–167.

[2] A.E. Berger, H. Brezis, and J.C.W Rogers, *A numerical method for solving the problem* $u_t - \Delta f(u) = 0$, RAIRO numerical analysis, **13** (1979), 297–312.

[3] F. Cavalli, *Stability of high order linearly implicit schemes for non-linear diffusion equation*, In preparation.

[4] [10.1007/s10440-012-9781-4] F. Cavalli. *Linearly implicit approximations of diffusive relaxation systems.* Acta Appl. Math., **125** (2013), 79–103

[5] F. Cavalli, G. Naldi, and I. Perugia, *Discontinuous Galerkin approximation of relaxation models for linear and nonlinear diffusion equations*, SIAM J. Sci. Comput., **34** (2012), A137–A160.

[6] F. Cavalli, G. Naldi, G. Puppo, and M. Semplice, *High-order relaxation schemes for non linear degenerate diffusion problems*, SIAM Journal on Numerical Analysis, **45(5)** (2007), 2098–2119.

[7] F. Cavalli, G. Naldi, G. Puppo, and M. Semplice, *A family of relaxation schemes for nonlinear convection diffusion problems*, Commun. Comput. Phys., **5(2–4)** (2009), 532–545.

[8] F. Cavalli, G. Naldi, G. Puppo, and M. Semplice, *Relaxed schemes based on diffusive relaxation for hyperbolic-parabolic problems: some new developments*, In "G. Puppo and G. Russo, editors, Numerical Methods for Balance Laws", volume 24, (2010) Aracne editrice (ITA).

[9] J. Kačur, A. Handlovičová, and M. Kačurová, *Solution of nonlinear diffusion problems by linear approximation schemes*, SIAM J. Numer. Anal., **30(6)** (1993), 1703–1722.

[10] R.J. LeVeque, "Numerical methods for conservation laws", 2^{nd} edition, lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, , 1992.

[11] Y. Liu, C.-W. Shu, and M. Zhang, *High order finite difference WENO schemes for nonlinear degenerate parabolic equations*, SIAM J. Sci. Comput., **33(2)** (2011), 939–965.

[12] E. Magenes, R. H. Nochetto, and C. Verdi, *Energy error estimates for a linear scheme to approximate nonlinear parabolic problems*, RAIRO Modél. Math. Anal. Numér., **21(4)** (1987), 655–678.

[13] I. S. Pop and W.-A. Yong, *A numerical approach to degenerate parabolic equations*, Numer. Math., **92(2)** (2002), 357–381.

[14] M. Slodička, *A robust and efficient linearization scheme for doubly nonlinear and degenerate parabolic problems arising in flow in porous media*, SIAM J. Sci. Comput., **23(5)** (2002), 1593–1614.

E-mail address: fausto.cavalli@unimi.it

THE INVISCID LIMIT FOR SLIP BOUNDARY CONDITIONS

Nikolai Chemetov

CMAF/Universidade de Lisboa, Av. Prof. Gama Pinto 2
1649-003 Lisboa, Portugal

Fernanda Cipriano

GFM and Dep. de Matemática FCT-UNL, Av. Prof. Gama Pinto 2
1649-003 Lisboa, Portugal

Abstract. We study the inviscid limit for the two dimensional Navier-Stokes equations with non-homogeneous Navier slip boundary condition. We show that the vanishing viscosity limit of Navier-Stokes's solutions verifies the Euler equations with the corresponding Navier slip boundary condition just on the inflow boundary. The convergence result is established with respect to the strong topology of the Sobolev spaces $W_p^1, p > 2$.

1. **Introduction.** We consider the Navier-Stokes equations for the viscous incompressible fluid in a bounded domain of $\mathbb{R}^2$ and investigate the convergence (up to the boundary) of their solutions as the viscosity goes to zero.

When the Navier-Stokes equations are supplemented with the usual Dirichlet boundary condition, the vanishing viscosity limit is a long-open problem. A discussion about the mathematical difficulties, related with Dirichlet boundary condition, can be found in the review articles [3], [12] and references therein. For a long time the Dirichlet boundary condition was considered as the natural one, however recent theoretical and experimental results ([8], [17], [27]) have pointed out, for instance, the relevance of the surface roughness on the slip behavior of the fluid particles on the surface wall. Slip type boundary conditions, which were firstly suggested by Navier (1823), have renewed interest in order to describe accurately physical processes.

The study of the vanishing viscous problem under homogeneous Navier's boundary conditions have been greatly developed in the last two decades. Lions in [19] considered a particular case of the Navier's boundary conditions, which is equivalent to the vorticity condition $\omega_\nu = 0$ on the boundary. In the article [9], the authors considered the Navier-Stokes equations with general homogeneous Navier slip boundary conditions and proved that the inviscid limit is a solution of the corresponding Euler equations. This study was performed in the functional space of L_∞-bounded vorticity. Later on, in [21], this result was generalized for the class of L_p-bounded vorticity with $p > 2$. A rate of the vanishing viscous convergence,

2000 *Mathematics Subject Classification.* Primary: 35M99, 76D09; Secondary: 76F40.

Key words and phrases. Navier-Stokes equations, Euler equations, Navier slip boundary conditions, vanishing viscosity, boundary layer.

The first author is supported by FCT and FEDER through the project POCTI / ISFL / 209 of Centro de Matemática e Aplicações Fundamentais da Universidade de Lisboa (CMAF / UL). The second author is supported by the projects FCT - PTDC/MAT/104173/2008 and EURATOM/IST.

in the class of almost L_∞-bounded vorticity was obtained in [18]. For the three dimensional developments we refer [4], [5], [31], where the $H^k, k \geq 3$, convergence results have been obtained for the domains with flat boundaries. For a general 3D domain a precise rate of the viscous convergence of velocities in L_2 and H^1-norms has been established in [16].

Throughout the twentieth century, extensive research has been carried out within the aircraft industry by the application of injection/suction devices to control turbulent boundary layers (see [1], [7], [23], [29]). To describe accurately injection-suction systems with slippage non-homogeneous Navier slip boundary conditions should be considered.

The purpose of this work is to investigate the problem of vanishing viscosity limit for the Navier-Stokes equations with non-homogeneous Navier's boundary conditions. To our knowledge, the vanishing viscous convergence results for the Navier-Stokes equations with non-homogeneous Navier boundary conditions have been obtained just in two articles. In the article [2], a particular case of Navier's boundary conditions (a prescribed vorticity on a permeable boundary) was studied. General non-homogeneous Navier boundary conditions was considered in [24], but in a very restrictive case, when the fluid domain is very small. However, it has not been proved that the inviscous limit fulfills the boundary condition on the injection zone. Both results are valid only in the W^1_∞-weak topology for velocities. See also an interesting conditional result in [30] for non-homogeneous Dirichlet's boundary condition.

Our article is organized as follows: in the section 2, we formulate the problem and state the main result Theorem 2.1. In the section 3, we establish the W^1_p-boundness for the velocity independently of the viscosity. This estimate allows to obtain the convergence of the solutions of the Navier-Stokes equations in W^1_p- weak topology. The section 4 contains the proof of Theorem 2.1. The key point of our approach is the application of the entropy method, which was introduced by S. Kruzkov and developed by F. Otto (see [22], [26]) for non linear hyperbolic equations. We extend the entropy method for the Navier-Stokes and Euler equations, in order to prove the strong convergence of ω_ν to ω in the L_p-topology and establish the trace result for ω.

2. **General setting.** The motion of an incompressible viscous fluid in a bounded domain $\Omega \subset \mathbb{R}^2$ is described by the Navier-Stokes equations

$$\mathbf{v}_t + \operatorname{div}(\mathbf{v} \otimes \mathbf{v}) - \bigtriangledown p = \nu \Delta \mathbf{v}, \quad \operatorname{div} \mathbf{v} = 0, \quad (t, \mathbf{x}) \in \ \Omega_T := (0, T) \times \Omega, \tag{1}$$

added by a given initial condition

$$\mathbf{v}(0, \mathbf{x}) = \mathbf{v}_0(\mathbf{x}) \quad \mathbf{x} \in \Omega, \quad \text{such that} \quad \operatorname{div} \mathbf{v}_0 = 0, \tag{2}$$

where $\mathbf{v} = \mathbf{v}(t, \mathbf{x})$ is the velocity, $p = p(t, \mathbf{x})$ is the pressure and ν is the viscosity of the fluid. Let us admit a flow of the fluid through the boundary $\Gamma \in C^2$ of the domain Ω, which can be described by the flux condition

$$\mathbf{v} \cdot \mathsf{n} = a \ \text{ on } \ \Gamma_T := (0, T) \times \Gamma \tag{3}$$

and by the Navier slip boundary condition

$$2D(\mathbf{v})\mathsf{n} \cdot \mathsf{s} + \alpha \mathbf{v} \cdot \mathsf{s} = b \tag{4}$$

on Γ_T. Here $D(\mathbf{v}) := \frac{1}{2}[\nabla \mathbf{v} + (\nabla \mathbf{v})^T]$ is the rate-of-strain tensor; n is the external normal to Γ and s is the tangent vector to Γ, such that (n, s) forms a standard orientation in $\mathbb{R}^2$. The quantity a of inflow and outflow fluid through Γ has to fulfill the natural condition

$$\int_\Gamma a(t,\mathbf{x})\ d\mathbf{x} = 0 \quad \text{for} \quad \forall t \in [0,T]. \tag{5}$$

The functions $\alpha = \alpha(t,\mathbf{x})$ and $b = b(t,\mathbf{x})$ prescribe physical properties of the boundary Γ.

We will show the vanishing viscous convergence up to the boundary Γ of solutions $\mathbf{v}_\nu$ of the Navier-Stokes system (1)-(5) (shortly designated by $\mathbf{NSS}_\nu$) to a solution of the Euler equations

$$\mathbf{v}_t + \operatorname{div}(\mathbf{v}\otimes\mathbf{v}) - \bigtriangledown p = 0, \quad \operatorname{div}\mathbf{v} = 0, \quad (t,\mathbf{x}) \in \Omega_T \tag{6}$$

with the initial-boundary conditions (2)-(3), and (4) just on the influx boundary $\Gamma_T^- := \{(t,\mathbf{x}) \in \Gamma_T :\ a(t,\mathbf{x}) < 0\}$.

In this work we consider the data in the following Banach's spaces:

$$\begin{aligned}
&a \in W_1^1(0,T;W_p^{1-\frac{1}{p}}(\Gamma)) \cap L_2(0,T;W_p^{2-\frac{1}{p}}(\Gamma)),\ \alpha \in L_\infty(\Gamma_T),\ \mathbf{v}_0 \in W_p^1(\Omega),\\
&b \in L_2(0,T;W_p^{1-\frac{1}{p}}(\Gamma)) \cap W_1^1(0,T;W_p^{-\frac{1}{p}}(\Gamma)) \ \text{ with } \ p \in (2,+\infty)
\end{aligned} \tag{7}$$

and prove the *strong convergence* of $\mathbf{v}_\nu$ to $\mathbf{v}$ in $W_p^1(\Omega)$.

Theorem 2.1. *Under the hypothesis (7), there exists a subsequence* $\{\mathbf{v}_\nu, \omega_\nu := \operatorname{rot}\mathbf{v}_\nu\}$ *of solutions to* $\mathbf{NSS}_\nu$*, such that*

$$\begin{aligned}
&\mathbf{v}_\nu \to \mathbf{v} \quad \textit{strongly in } L_r(0,T;W_p^1(\Omega)),\\
&\omega_\nu \to \omega \quad \textit{strongly in } L_r(0,T;L_p(\Omega)) \textit{ for any } r \in [1,\infty).
\end{aligned} \tag{8}$$

The limit pair $\{\mathbf{v},\omega\}$ *is a solution of the Euler equations (6)(see (27) too), satisfying (2)-(3) and the Navier slip boundary condition (4) on the influx boundary* Γ_T^-. *Moreover we have*

$$\begin{aligned}
&\int_{\Omega_T} \beta(\omega)\,\big(\psi_t + \mathbf{v}\cdot\nabla\psi\big)\ dt\,d\mathbf{x} + \int_\Omega \beta(\omega_0)\ \psi(0,\mathbf{x})\ d\mathbf{x}\\
&= \int_{\Gamma_T^-} a\beta(\omega_\Gamma(\mathbf{v}))\psi\ dt\,d\mathbf{x} \textit{ for any } \beta \in C(\mathbb{R}),
\end{aligned} \tag{9}$$

where $\psi \in C^{1,1}(\overline{\Omega}_T)$*, such that* $\psi = 0$ *at* $t = T$.

3. **Main estimates.** In the first paragraph of this section, we estimate the L_2-norm of the velocity $\mathbf{v}_\nu$ by the L_p-norm of the corresponding vorticity $\omega_\nu = rot\mathbf{v}_\nu$. In the next one, we obtain estimates for the solutions $\mathbf{v}_\nu$ and ω_ν independently of the viscosity.

Through the article, we denote by C all constants that are independent of the viscosity ν.

3.1. Estimate for the velocity by the corresponding vorticity.

The result of the following Lemma will be useful in the subsection 3.2 to establish a Gronwall type inequality for the vorticity.

Lemma 3.1. *Assume that the hypothesis (7) hold, then there exists a unique solution* $\mathbf{v}_\nu \in C([0,T]\,;\ W_2^1(\Omega))$ *for* $\mathbf{NSS}_\nu$*, satisfying the estimate*

$$\|\mathbf{v}_\nu(t,\cdot)\|^2_{L_2(\Omega)} \le C\left(\|\mathbf{v}_0\|^2_{L_2(\Omega)} + \int_0^t f(r)\|\omega_\nu(r,\cdot)\|^2_{L_p(\Omega)}\,dr + 1\right) \tag{10}$$

for any $t \in [0,T]$, where $f(t) \in L_1(0,T)$ depends only on the data.

Proof. Let h_ν, h_a be the solutions of the systems

$$\begin{cases} -\Delta h_\nu = \omega_\nu & \text{in } \Omega, \\ h_\nu = 0 & \text{on } \Gamma, \end{cases} \qquad \begin{cases} -\Delta h_a = 0 & \text{in } \Omega, \\ \frac{\partial h_a}{\partial \mathsf{n}} = a & \text{on } \Gamma \end{cases} \quad \text{a.e. on } (0,T). \tag{11}$$

Let us introduce the functions $\mathbf{u}_\nu := \nabla^\perp h_\nu$, $\mathbf{a} := \nabla h_a$. It is easy to check that $\mathbf{u}_\nu = \mathbf{v}_\nu - \mathbf{a}$ is a solution of the system

$$\begin{cases} \partial_t \mathbf{u}_\nu + \mathbf{a} \cdot \nabla \mathbf{u}_\nu - \bigtriangledown p = \nu \Delta \mathbf{u}_\nu + F_\nu, \ \operatorname{div} \mathbf{u}_\nu = 0 \ \text{ in } \Omega_T, \\ \mathbf{u}_\nu \cdot \mathsf{n} = 0, \ 2D(\mathbf{u}_\nu)\, \mathsf{n} \cdot \mathsf{s} + \alpha \mathbf{u}_\nu \cdot \mathsf{s} = \tilde{b} \text{ on } \Gamma_T, \\ \mathbf{u}_\nu(0,\mathbf{x}) = \mathbf{v}_0(\mathbf{x}) - \mathbf{a}(0,\mathbf{x}) \text{ in } \Omega \end{cases} \tag{12}$$

with $F_\nu := \nu \Delta \mathbf{a} - \partial_t \mathbf{a} - \mathbf{u}_\nu \cdot \nabla \mathbf{u}_\nu - \mathbf{a} \cdot \nabla \mathbf{a} - \mathbf{u}_\nu \cdot \nabla \mathbf{a}$ and $\tilde{b} := b - [2D(\mathbf{a})\, \mathsf{n} \cdot \mathsf{s} + \alpha \mathbf{a} \cdot \mathsf{s}]$.

Multiplying the first equation in (12) by $\mathbf{u}_\nu$, integrating over Ω and using the Calderon-Zygmund estimates ([25], Theorem 9.9, p. 230 in [14] and Theorem 1.8, p. 12 & Theorem 1.10, p. 15 in [15]) we obtain

$$\frac{1}{2}\frac{d}{dt} ||\mathbf{u}_\nu||^2_{L_2(\Omega)} + \nu \int_\Omega |D(\mathbf{u}_\nu)|^2 \, d\mathbf{x} \\ \le f(t)(||\mathbf{u}_\nu||^2_{L_2(\Omega)} + \|\omega_\nu\|^2_{L_p(\Omega)} + 1) + \nu \int_\Omega |D(\mathbf{a})|^2 \, d\mathbf{x}$$

with $f(t) \in L_1(0,T)$ depending only on the data a, b, α (independently on ν) due to (7). Applying Gronwall's Lemma, we deduce (10). □

3.2. **Uniform estimate for the vorticity.** Let us parametrize the boundary Γ using the arc length s and denote by k the curvature of Γ. Since $\Gamma \in C^2$ we know that k is a continuous function.

From Lemma 1 of [10] the Navier slip boundary condition (4) is equivalent to the following boundary condition for the vorticity

$$\omega_\nu = \omega_\Gamma(\mathbf{v}_\nu) := \gamma\, \mathbf{v}_\nu \cdot \mathsf{s} + g \ \text{ on } \ \Gamma_T \ \text{ with } \ \gamma := 2k - \alpha, \quad g := b - 2\frac{\partial a}{\partial s}. \tag{13}$$

Therefore $\mathbf{NSS}_\nu$ can be written in terms of vorticity as

$$\begin{cases} \partial_t \omega_\nu + \operatorname{div}(\omega_\nu \mathbf{v}_\nu) = \nu \Delta \omega_\nu, \ \operatorname{rot} \mathbf{v}_\nu = \omega_\nu, \ \ \operatorname{div} \mathbf{v}_\nu = 0 \ \text{ in } \ \Omega_T, \\ \mathbf{v}_\nu \cdot \mathsf{n} = a \ \text{ on } \Gamma_T \ \text{ and } \ \omega_\nu = \omega_\Gamma(\mathbf{v}_\nu) \ \text{ on } \Gamma_T, \\ \omega_\nu\big|_{t=0} = \omega_0 \ \text{ in } \ \Omega \ \text{ with } \ \omega_0 := \operatorname{rot} \mathbf{v}_0, \end{cases} \tag{14}$$

which we will continue to designate by $\mathbf{NSS}_\nu$. In the following lemma we deduce uniform estimates on the viscosity ν for the solutions ω_ν of $\mathbf{NSS}_\nu$.

Lemma 3.2. *Under the hypothesis (7), the estimates*

$$\|\omega_\nu\|_{L_\infty(0,T;L_p(\Omega))} \le C, \quad \|\mathbf{v}_\nu\|_{L_\infty(0,T;\, W^1_p(\Omega))} \le C, \tag{15}$$

$$\|\partial_t(\mathbf{v}_\nu - \mathbf{a})\|_{L_\infty(0,T;H^{-1}(\Omega))} \le C \tag{16}$$

hold. For the trace value of $\mathbf{v}_\nu - \mathbf{a}$ on Γ_T, we have

$$(\mathbf{v}_\nu - \mathbf{a}) \cdot \mathbf{s} \in \mathcal{P}[\Gamma_T] := L_2\big(0,T; H^{1/2}(\Gamma)\big) \cap H^{1/4}\big(0,T; L_2(\Gamma)\big),$$

such that

$$\|(\mathbf{v}_\nu - \mathbf{a}) \cdot \mathbf{s}\|_{\mathcal{P}[\Gamma_T]} \le C. \tag{17}$$

Proof. Let us extend $\omega_\Gamma(\mathbf{v}_\nu)$ into the domain Ω by the solution $B(t,\cdot)$ of the system

$$-\Delta B = 0 \text{ in } \Omega,\ B\big|_\Gamma = \omega_\Gamma(\mathbf{v}_\nu),\ \text{ for all } t \in (0,T)\,.$$

The function $z = \omega_\nu - B$ solves the system

$$\begin{cases} \partial_t z + \operatorname{div}(z\mathbf{v}_\nu) = \nu\Delta\, z + F \text{ in } \Omega_T, \\ z = 0 \text{ on } \Gamma_T,\ z|_{t=0} = \omega_0 - B|_{t=0} \text{ in } \Omega. \end{cases} \tag{18}$$

with $F := F_1 + F_2$, $F_1 = -\partial_t B - \mathbf{v}_\nu \cdot \nabla B$ and $F_2 = \nu\Delta B$. Multiplying the equation in (18) by $G := p|z|^{p-2}z$, integrating over Ω and using Lemma 3.1, finally we obtain

$$\|\omega_\nu\|^2_{L_\infty(0,t;L_p(\Omega))} \le C\left\{\int_0^t f(t)\|\omega_\nu(t,\cdot)\|^2_{L_p(\Omega)}\,dt + 1\right\} \tag{19}$$

for any $t \in [0,T]$. Applying Gronwall's Lemma we obtain the first estimate of (15). The second estimate of (15) is a direct consequence of the Calderon-Zygmund estimates.

In order to deduce (16) we take the time derivative of the 1st system in (11) and set a system for the function $\partial_t h_\nu$, then we deduce the following estimate

$$\|\partial_t h_\nu(t,\cdot)\|_{L_2(\Omega)} \le C\|G_\nu(t,\cdot)\|_{H^{-2}(\Omega)} \text{ a.a. } t \in (0,T) \tag{20}$$

with $G_\nu := \operatorname{div}(-\mathbf{v}_\nu\,\omega_\nu) + \nu\,\triangle\omega_\nu$. By (15) we get $\|G_\nu(t,\cdot)\|_{H^{-2}(\Omega)} \le C$, that implies (16). By Lemma 8, p. 700 of [11] we have (17). □

4. **Proof of Theorem 2.1.** The proof is based on the entropy method of F. Otto [26] (see also [22]).

Since $\Gamma \in C^2$, there exists a small $\delta_0 > 0$, such that the distance $d(\mathbf{x}) := \inf_{\mathbf{y}\in\Gamma}|\mathbf{x} - \mathbf{y}|$ is C^2-function in the set $U_{\delta_0}(\Gamma) := \{\mathbf{x} \in \Omega :\ d(\mathbf{x}) < \delta_0\}$. Let

$$l(\mathbf{x}) := \begin{cases} \ \ \min\{\delta_0,\ d(\mathbf{x})\},\ \text{if } x \in \Omega\ ; \\ -\min\{\delta_0,\ d(\mathbf{x})\},\ \text{otherwise}\ , \end{cases} \qquad L := \sup_{0<l(\mathbf{x})<\delta_0} |\Delta l(\mathbf{x})|. \tag{21}$$

By (15) and the embedding $W^1_p(\Omega) \hookrightarrow L_\infty(\Omega)$, we have $\|\mathbf{v}_\nu\|_{L_\infty(\Omega_T)} \le M$ for a constant M independent on ν. We define a cut-off function by

$$\xi_\nu(\mathbf{x}) := 1 - \exp(-\frac{M + \nu L}{\nu}\,l(\mathbf{x})), \quad \forall\nu > 0.$$

Let us take a convex function $\eta \in C^2(\mathbb{R})$ and a non negative function $\phi \in C^{2,1}(\overline{\Omega}_T)$ with $\phi|_{t=T} = 0$. If we multiply the differential equation of $\mathbf{NSS}_\nu$ by $\eta'(\omega_\nu)\phi\xi_\nu$ and integrate over Ω_T, we obtain the inequality

$$\begin{aligned} &\int_{\Omega_T} \eta(\omega_\nu)\ [\phi_t + \mathbf{v}_\nu\cdot\nabla\phi + \nu\,\Delta\phi]\ \xi_\nu + 2\,\nu\,\eta(\omega_\nu)\,(\nabla\phi\cdot\nabla\xi_\nu)\ dt\,d\mathbf{x} \\ &+ \int_{\Gamma_T} (M + \nu L)\eta(\omega_\Gamma(\mathbf{v}_\nu))\,\phi\,dt\,d\mathbf{x} + \int_\Omega \eta(\omega_0)\,\phi(0)\,\xi_\nu\,d\mathbf{x} \ge 0 \end{aligned} \tag{22}$$

with the help of the differential inequality (8.3), p. 129-131 of [22].

Let us denote by $|v|^+$ the positive part and $|v|^-$ the negative part of a function v. It is clear that inequality (22) is true also for $\eta(v) = |v - c|^+$ and $\eta(v) = |v - c|^-$ with arbitrary chosen $c \in \mathbb{R}$.

Estimates (15)-(17), Corollary 4 of [28] and the compact embedding $\mathcal{P}[\Gamma_T] \hookrightarrow L_2(\Gamma_T)$, guarantee the existence of an appropriate subsequence of $\{\omega_\nu, \mathbf{v}_\nu\}$ for $\nu \to 0^+$, satisfying

$$\begin{aligned} &|\omega_\nu - c|^+ \rightharpoonup z^1,\ |\omega_\nu - c|^- \rightharpoonup z^2,\ (\omega_\nu - c) \rightharpoonup (\omega - c) = z^1 - z^2 \\ &\text{weakly} - * \text{ in } L_\infty(0,T;L_p(\Omega)),\ \forall c \in \mathbb{R}, \\ &\mathbf{v}_\nu \to \mathbf{v} \ \text{ strongly in } L_2(\Omega_T) \text{ and weakly-* in } L_\infty(0,T;W_p^1(\Omega)), \\ &\mathbf{v}_\nu \cdot \mathbf{s} \to \mathbf{v} \cdot \mathbf{s} \text{ strongly in } L_2(\Gamma_T). \end{aligned} \tag{23}$$

Let us take $\eta(v) = |v - c|^+$ and $\eta(v) = |v - c|^-$ in (22), respectively, and pass to the limit as $\nu \to 0$. Using (23), Lebesgue's dominated convergence theorem and $\nu|\nabla \xi_\nu| \le M + L,\ \nu \nabla \xi_\nu \to 0$ in Ω, we obtain that the functions $z^1,\ z^2 \in L_\infty(0,T;L_p(\Omega))$ and $\mathbf{v} \in L_\infty(0,T;W_p^1(\Omega))$ satisfy the entropy inequalities

$$\begin{aligned} &\int_{\Omega_T} z^1 \left(\phi_t + \mathbf{v} \cdot \nabla \phi\right) \, dt\, d\mathbf{x} \\ &+ M \int_{\Gamma_T} |\omega_\Gamma(\mathbf{v}) - c|^+ \phi \; dt\, d\mathbf{x} + \int_\Omega |\omega_0 - c|^+ \phi(0,\mathbf{x}) \; d\mathbf{x} \ge 0 \end{aligned} \tag{24}$$

and

$$\begin{aligned} &\int_{\Omega_T} z^2 \left(\phi_t + \mathbf{v} \cdot \nabla \phi\right) \, dt\, d\mathbf{x} \\ &\quad + M \int_{\Gamma_T} |\omega_\Gamma(\mathbf{v}) - c|^- \phi \; dt\, d\mathbf{x} + \int_\Omega |\omega_0 - c|^- \phi(0,\mathbf{x}) \; d\mathbf{x} \ge 0, \end{aligned} \tag{25}$$

for any non negative function $\phi \in C^{1,1}(\overline{\Omega}_T)$ with $\phi|_{t=T} = 0$.

Formally, the entropy inequalities (24)-(25) for z^1 and z^2 can be expressed through the following systems

$$\begin{cases} \partial_t z^1 + (\mathbf{v} \cdot \nabla) z^1 \le 0 \text{ in } \Omega_T, \\ z^1 \le M|\omega_\Gamma(\mathbf{v}) - c|^+ \text{ on } \Gamma_T^-, \\ z^1 \le |\omega_0 - c|^+ \text{ at } t = 0 \end{cases} \quad \text{and} \quad \begin{cases} \partial_t z^2 + (\mathbf{v} \cdot \nabla) z^2 \le 0 \text{ in } \Omega_T, \\ z^2 \le M|\omega_\Gamma(\mathbf{v}) - c|^- \text{ on } \Gamma_T^-, \\ z^2 \le |\omega_0 - c|^- \text{ at } t = 0. \end{cases}$$

Since $z^1,\ z^2$ are non-negative functions, we can apply Diperna-Lions's [13] and Boyer's [6] methods, to verify that the function $z^1 z^2$ fulfills the system

$$\begin{cases} \partial_t (z^1 z^2) + (\mathbf{v} \cdot \nabla)(z^1 z^2) \le 0 \text{ in } \Omega_T, \\ z^1 z^2 = 0 \text{ on } \Gamma_T^-, \\ z^1 z^2 = 0 \text{ at } t = 0, \end{cases}$$

which implies $z^1\, z^2 = 0$ in Ω_T. Hence multiplying the identity $\omega - c = z^1 - z^2$ by z^1 and z^2, respectively, we obtain $|\omega - c|^+ = z^1$ and $|\omega - c|^- = z^2$. Taking the sum of (24) and (25) we obtain

$$\begin{aligned} &\int_{\Omega_T} |\omega - c| \left(\phi_t + \mathbf{v} \cdot \nabla \phi\right) \, dt\, d\mathbf{x} \\ &+ M \int_{\Gamma_T} |\omega_\Gamma(\mathbf{v}) - c| \phi \; dt\, d\mathbf{x} + \int_\Omega |\omega_0 - c| \; \phi(0,\mathbf{x}) \; d\mathbf{x} \ge 0, \end{aligned} \tag{26}$$

which is valid for any fixed constant $c \in \mathbb{R}$ and arbitrary non negative function $\phi \in C^1(\overline{\Omega}_T)$, such that $\phi = 0$ at $t = T$.

Due to (14) and (23) the pair $\{\mathbf{v},\omega\}$ satisfies the equality

$$\int_{\Omega_T} \omega\ [\psi_t + \mathbf{v}\cdot\nabla\psi]\ dt\,d\mathbf{x} + \int_{\Omega} \omega_0\,\psi(0,\mathbf{x})\,d\mathbf{x} = 0, \quad \forall\psi\in C^1(\overline{\Omega}_T) \tag{27}$$

with $\psi = 0$ on Γ_T and at $t = T$. Therefore, applying Theorem 3.1 of [6], the vorticity ω has a trace $\gamma\omega$ measurable on Γ_T, such that

$$\begin{aligned}\int_{\Omega_T} |\omega - c|\,\big(\phi_t + \mathbf{v}\cdot\nabla\phi\big)\ dt\,d\mathbf{x} + \int_{\Omega} |\omega_0 - c|\ \phi(0,\mathbf{x})\ d\mathbf{x}\\ = \int_{\Gamma_T} a|\gamma\omega - c|\phi\ dt\,d\mathbf{x}\end{aligned} \tag{28}$$

with ϕ as in (26). By (26) and (28) we conclude

$$a(t,\mathbf{x})\ |\gamma\omega(t,\mathbf{x}) - c| + M|\omega_\Gamma(\mathbf{v})(t,\mathbf{x}) - c|{\geq}0 \text{ for a.a. } (t,\mathbf{x})\in\Gamma_T.$$

Choosing $c := \omega_\Gamma(\mathbf{v})(t,\mathbf{x})$, we derive $\gamma\omega = \omega_\Gamma(\mathbf{v})$ on Γ_T^-. Hence we see that Theorem 3.1 of [6] implies (9).

From above we have

$$|\omega_\nu - c|^{\pm} \rightharpoonup |\omega - c|^{\pm} \text{ weakly } -* \text{ in } L_\infty(0,T;L_p(\Omega)), \quad \forall c\in\mathbb{R}.$$

Due to the formula

$$|v|^p = p(p-1)\left\{\int_0^\infty |v-c|^+ c^{p-2}\,dc + \int_{-\infty}^0 |v-c|^-\ |c|^{p-2}\,dc\right\},$$

which is valid for any $v\in\mathbb{R}$, we deduce

$$\int_{\Omega_T} |\omega_\nu|^p\ dt\,d\mathbf{x} \to \int_{\Omega_T} |\omega|^p\ dt\,d\mathbf{x}. \tag{29}$$

Therefore, applying Theorem 2.11, p. 57 of [20], we obtain $g_\nu(t) := ||\omega_\nu - \omega||^{p/2}_{L_p(\Omega)} \to 0$ and $|g_\nu(t)| \leq C$ for a.a. $t\in(0,T)$. By Lebesgue's dominated convergence theorem there exists a subsequence of ω_ν, denoted by the same index ν:

$$\omega_\nu \to \omega \text{ strongly in } L_r(0,T;L_p(\Omega)) \text{ for any } r\in[1,\infty).$$

Finally by Calderon-Zygmund´s estimate

$$\|\mathbf{v}_\nu(t,\cdot) - \mathbf{v}(t,\cdot)\|_{W_p^1(\Omega)} \leq C\|\omega_\nu(t,\cdot) - \omega(t,\cdot)\|_{L_p(\Omega)} \text{ for a.a. } t\in(0,T),$$

we have the strong convergence (8) for the velocity $\mathbf{v}$.

REFERENCES

[1] A. Abbas, J. de Vicente and E. Valero, *Aerodynamic technologies to improve aircraft performance*, Aerospace Science and Technology, Article in press (2012).

[2] G. V. Alekseev, *The solvability of an inhomogeneous boundary value problem for two-dimensional non-stationary equations of the dynamics of an ideal fluid*, (Russian) [Dinamika Splošn. Sredy Vyp.] **24**, Dinamika Zidk. so Svobod. Granicami], **169** (1976), 15–35.

[3] C. Bardos and E. S. Titi, *Euler equations for incompressible ideal fluids*, Russian Math. Surveys, **62** (2007), 409–451.

[4] H. Beirão da Veiga and F. Crispo, *Concerning the $W^{k,p}$-Inviscid Limit for 3-D Flows Under a Slip Boundary Condition*, J. Math. Fluid Mech., **13** (2011), 117–135.

[5] H. Beirão da Veiga and F. Crispo, *The 3-D Inviscid Limit Result Under Slip Boundary Conditions. A Negative Answer*, J. Math. Fluid Mech., **14** (2012), 55–59.

[6] F. Boyer, *Trace theorems and spatial continuity properties for the solutions of the transport equation*, Differential and integral equations, **18** (2005), 891–934.

[7] A. L. Braslow, "A History of Suction-Type Laminar-Flow Control with Emphasis on Flight Research", NASA History Division (1999).

[8] D. Bucur, E. Feireisl and S. Necasova, *Boundary Behavior of Viscous Fluids: influence of wall roughness and friction-driven Boundary Conditions*, Arch. Rational Mech. Anal., **197** (2010), 117–138.

[9] T. Clopeau, A. Mikelic and R. Robert, *On the vanishing viscosity limit for the 2D incompressible Navier-Stokes equations with the friction type boundary conditions*, Nonlinearity, **11** (1998), 1625–1636.

[10] N.V. Chemetov and S.N. Antontsev, *Euler equations with non-homogeneous Navier slip boundary condition*, Physica D: Nonlinear Phenomena, **237** (2008), 92–105.

[11] N.V. Chemetov, F. Cipriano and S. Gavrilyuk, *Shallow water model for lakes with friction and penetration*, Math. Methods Appl. Sci., **33** (2010), 687–703.

[12] P. Constantin, *Euler and Navier-Stokes equations*, Publ. Mat., **52** (2008), 235–265.

[13] R.J. DiPerna and P.-L. Lions, *Ordinary differential equations, transport theory and Sobolev spaces*, Invent. Math., **98** (1989), 511–547.

[14] D. Gilbarg and N.S. Trudinger, “Elliptic Partial Differential Equations”, Springer-Verlag, Berlin Heidelberg New-York (2001).

[15] V. Girault and P.-A. Raviart, “Finite Element Methods for Navier-Stokes equations”, Theory and Algorithms. Springer-Verlag, Berlin Heidelberg New-York (1986).

[16] D. Iftimie and F. Sueur, *Viscous boundary layers for the Navier-Stokes equations with the Navier slip conditions*, Arch. Ration. Mech. Anal., **199** (2011), 145–175.

[17] W. Jager and A. Mikelic, *On the roughness-induced effective boundary conditions for a viscous flow*, J. Differential Equations, **170** (2001), 96–122.

[18] J. Kelliher, *Navier-Stokes equations with Navier boundary conditions for a bounded domain in the plane*, SIAM J. Math. Anal., **38** (2006), 210–232.

[19] P.-L. Lions, “Mathematical Topics in Fluid Mechanics” Vol. 1, The Clarendon Press Oxford University Press, New York (1996).

[20] E.H. Lieb and M. Loss, “Analysis”, 2nd edition, Graduate Studies in Math., vol. 14, AMS, Providence RI (2001).

[21] M.C. Lopes Filho, H.J. Nussenzveig Lopes and G. Planas, *On the inviscid limit for 2d incompressible flow with Navier friction condition*, SIAM Math. Anal., **36** (2005), 1130–1141.

[22] J. Malek, J. Necas, M. Rokyta and M. Ruzicka, “Weak and Measure-Valued Solutions to Evolutionary PDEs”, Chapman & Hall, London (1996).

[23] L.A. Marshall, “Boundary-Layer Transition Results From the F-16XL-2 Supersonic Laminar Flow Control Experiment”, NASA/TM-1999-209013, Dryden Flight Research Center Edwards, California, 93523-0273, December, 1999.

[24] P. Mucha, *On the inviscid limit of the Navier–Stokes equations for flows with large flux*, Nonlinearity, **16** (2003), 1715–1732.

[25] J. Necas, “Direct Methods in the Theory of Elliptic Equations”, Springer-Verlag, Berlin Heidelberg New-York (2010).

[26] F. Otto, *Initial-boundary value problem for a scalar conservation law*, C.R. Acad. Sci. Paris Sér. I Math., **322** (1996), 729–734.

[27] N.V. Priezjev and S.M. Troian, *Influence of periodic wall roughness on the slip behavior at liquid/solid interfaces: molecular-scale simulations versus continuum predictions*, J. Fluid Mech., **554** (2006), 25–46.

[28] J. Simon, *Compact Sets in the space $L^p(0,T;B)$*, Annali di Matematica Pura ed Applicata, **146** (1986), 65–96.

[29] H. Schlichting and K. Gersten, “Boundary-Layer Theory”, Springer-Verlag, Berlin Heidelberg New-York (2003).

[30] R. Temam and X. Wang, *Boundary Layers Associated with Incompressible Navier–Stokes Equations: The Noncharacteristic Boundary Case*, J. Diff. Equations, **179** (2002), 647–686.

[31] Y. Xiao and Z. Xin, *On the vanishing viscosity limit for the 3D Navier-Stokes equations with a slip boundary condition*, Communications on Pure and Applied Mathematics, **60** (2007), 1027–1055.

E-mail address: `chemetov@ptmat.fc.ul.pt`
E-mail address: `cipriano@cii.fc.ul.pt`

A FINITE VOLUME EVOLUTION GALERKIN SCHEME FOR ACOUSTIC WAVES IN HETEROGENEOUS MEDIA

Koottungal Revi Arun

Institut für Geometrie und Praktische Mathematik
RWTH-Aachen, Templergraben 55
D-52056 Aachen, Germany

Guoxian Chen*

School of Mathematics and Statistics
Wuhan University, Wuhan, 430072, China
and
Institut für Geometrie und Praktische Mathematik
RWTH-Aachen, Templergraben 55
D-52056 Aachen, Germany

Sebastian Noelle

Institut für Geometrie und Praktische Mathematik
RWTH-Aachen, Templergraben 55
D-52056 Aachen, Germany

Abstract. In this paper, we present a numerical scheme for the propagation of acoustic waves in a heterogeneous medium in the context of the finite volume evolution Galerkin (FVEG) method (M. Lukáčová-Medviďová et al. *J. Comput. Phys.*, 183:533–562, 2002). As a mathematical model we consider a wave equation system with space dependent wave-speed and impedance, which is used to study the wave propagation in a complex media. A main building block of our scheme is a genuinely multidimensional evolution operator based on the bicharacteristic theory of hyperbolic systems under the assumption of space dependent Jacobian matrices. We employ a novel approximation of the evolution operator, resulting from quadratures, in the flux evaluation stage of a finite volume scheme. The results of several numerical case studies clearly demonstrate the efficiency and robustness of the new FVEG scheme.

1. **Introduction.** Hyperbolic conservation laws with spatially varying flux functions model acoustic or elastic waves in a heterogeneous medium [2]. In exploration seismology, e.g., one studies the propagation of small amplitude man-made waves in earth and their reflection off geological structures. The hope is to determine the geological structure (for example oil reservoirs) from measurements at the surface.

2000 *Mathematics Subject Classification.* Primary: 35L05, 35L40, 35L65, 35L67; Secondary: 65M08, 65M25, 76M12.

Key words and phrases. Heterogeneous Media, Acoustic Waves, Finite volume methods, Evolution Galerkin scheme, Hyperbolic conservation laws, bicharacteristics.

The first author is supported by the Alexander-von-Humboldt Foundation through a postdoctoral fellowship. The second and third authors are partially supported by the Deutsche Forschungsgemeinschaft (DFG) under grant number NO 361/3-2. The second author is also partially supported by the National Natural Science Foundation of China (No. 11001211, 11371023).

*Corresponding Author: gxchen.math@whu.edu.cn

A similar principle as in seismological exploration of the earth is used also in ultrasound exploration of human tissues. In all of these cases new phenomena can appear since reflections of waves at interfaces can lead to discontinuities even for linear equations.

The goal of the present work is to develop a numerical scheme for the propagation of acoustic waves in a heterogeneous medium in the context of the finite volume evolution Galerkin (FVEG) method. The FVEG method has been developed originally by Lukáčová and her coworkers, cf. e.g. [3, 4]. It is a predictor-corrector method combining the finite volume corrector step with the evolutionary predictor step. The corrector step approximates the fluxes by the midpoint rule in time and trapezoidal rule in space. At the space-time quadrature nodes, point values of the solution are predicted by a multidimensional approximate evolution operator. The latter is constructed using the theory of bicharacteristics under the assumption of spatially dependent Jacobian matrices. In the previous works of Lukáčová and others the evolution operators were derived for constant coefficient, locally linearised systems where bicharacteristics reduce to straight lines. An attempt to design a generalised FVEG scheme for linear hyperbolic systems with variable coefficients is done by Arun et al. in [1], where the methodology is demonstrated for a simple acoustic wave equation. The present work is a continuation along the lines of [1] to study wave propagation in complex media and to this end, we consider more general and practically relevant mathematical models.

Using a general version of the compatibility condition on a bicharacteristic curve [5], we derive an exact and (then use it to get) an approximate evolution operator. As shown in [1], in order to obtain a stable scheme, the coefficients of the heterogeneous medium must be approximated over a staggered grid that is centered at the integration points on cell interfaces. Our numerical experiments for wave propagation with continuous as well as discontinuous wave speeds, through smooth as well as non-smooth interfaces confirm robustness and reliability of the new FVEG scheme.

2. **Finite Volume Evolution Galerkin Method.**

In this section we design an FVEG scheme for the numerical simulation of acoustic waves in a heterogeneous medium. In contrast to [1], the mathematical model used here for the propagation of acoustic waves is obtained by linearising the isentropic Euler equations or the elasticity equations; see [2] for a derivation. The system of equations reads

$$\partial_t U + \partial_x F_1(U) + \partial_y F_2(U) = 0, \tag{1}$$

with the vector of unknowns U and the flux-vectors $F_1(U)$ and $F_2(U)$ given as

$$U = \begin{pmatrix} \phi \\ \rho u \\ \rho v \end{pmatrix}, \; F_1(U) = \begin{pmatrix} u \\ K\phi \\ 0 \end{pmatrix}, \; F_2(U) = \begin{pmatrix} v \\ 0 \\ K\phi \end{pmatrix}. \tag{2}$$

Here, ϕ can be thought as the amplitude of a pressure wave and u, v are respectively the velocities in the x and y directions. The parameters $K(x,y)$ and $\rho(x,y)$ are respectively the bulk modulus and density and hence, are material dependent.

Let X be the vector-valued space of solutions to (1) and let $E(\tau)\colon X \to X$ be the exact solution operator, i.e.

$$U(\cdot, t+\tau) = E(\tau)U(\cdot, t). \tag{3}$$

Let V_r be an approximation space of vector-valued piecewise polynomials of degree r and let us denote by U^n, the approximation to the exact solution $U(\cdot,t^n)$ in the space V_r. Since the exact solution is not always available, we suppose that an adequate approximate solution operator $E_\tau\colon V_r \to X$ is given. Let us also denote by $R\colon V_s \to V_r$, a suitable recovery operator, where $V_s \subset V_r$ is the space of vector-valued piecewise polynomials of degree s. Starting from U^n, the FVEG scheme can be recursively defined by

Definition 2.1.

$$U^{n+1} = U^n - \frac{1}{\Delta x}\int_0^{\Delta t} \delta_x F_1\left(U^{n+\frac{\tau}{\Delta t}}\right) d\tau - \frac{1}{\Delta y}\int_0^{\Delta t} \delta_y F_2\left(U^{n+\frac{\tau}{\Delta t}}\right) d\tau. \tag{4}$$

Here, δ_x and δ_y are finite difference operators, e.g. $\delta_x f(x) = f(x+h/2) - f(x-h/2)$ and $\delta_x F_1(U^{n+\tau/\Delta t})$ and $\delta_y F_2(U^{n+\tau/\Delta t})$ are respectively the flux differences in the x and y directions at time $t^n+\tau$. In order to evolve these fluxes, the approximate evolution operator is used, i.e.

$$U^{n+\frac{\tau}{\Delta t}} = \sum \left(\frac{1}{|\partial\Omega|}\int_{\partial\Omega} E_\tau R U^n d\sigma\right)\chi_{\partial\Omega}, \tag{5}$$

where χ is the characteristic function of the edge $\partial\Omega$ and summation is taken over all the computational cells.

In traditional predictor-corrector schemes like the two step Lax-Wendroff scheme, the predictor step is done by a multi-dimensional finite difference operator, for example the Lax-Friedrichs scheme. The FVEG scheme tries to replace the Lax-Friedrichs step by a more accurate evolution operator based on the theory of bicharacteristic curves, which is then approximated by quadrature; see [3] for details. The appealing element of the latter is that it systematically tries to take into account the infinitely many directions of wave propagation.

3. **Exact and Approximate Evolution Operators.**

Let us write the wave equation system in the primitive form:

$$\partial_t V + A_1 \partial_x V + A_2 \partial_y V = 0, \tag{6}$$

where

$$V = \begin{pmatrix} p \\ u \\ v \end{pmatrix}, \quad A_1 = \begin{pmatrix} 0 & K(x,y) & 0 \\ \frac{1}{\rho(x,y)} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 0 & K(x,y) \\ 0 & 0 & 0 \\ \frac{1}{\rho(x,y)} & 0 & 0 \end{pmatrix}. \tag{7}$$

We define the wavespeed c and the impedance Z via the relations $c := \sqrt{K/\rho}$ and $Z := \sqrt{K\rho}$.

We fix a point $P = (x, y, t^n + \tau)$ in space-time and consider the characteristic conoid of (6), passing through P and enveloped by the bicharacteristics given by

$$\frac{dx}{dt} = -c(x,y)\cos\theta, \ \frac{dy}{dt} = -c(x,y)\sin\theta, \ \frac{d\theta}{dt} = -\sin\theta\partial_x c + \cos\theta\partial_y c. \tag{8}$$

Here, $(\cos\theta, \sin\theta)$ is the unit normal to the wavefront, which is the section of the conoid by $t = \text{Const}$ hyperplanes. We solve the system of equations (8) with the initial values $x(\omega, t+\tau) = x,\ y(\omega, t+\tau) = y$ and $\theta(\omega, t+\tau) = \omega \in [0, 2\pi]$. Let Q and

$\tilde{Q}$ be respectively arbitrary points on the wavefronts at $t = t^n$ and $t = \tilde{t} \in (t^n, t^n+\tau)$. Proceeding as in [1], we can derive the exact evolution operators

$$\begin{aligned}
\begin{pmatrix} p \\ u \\ v \end{pmatrix}(P) = \frac{1}{2\pi}\int_0^{2\pi} (p - Z\cos\theta u - Z\sin\theta v)(Q)\begin{pmatrix} 1 \\ \cos\omega \\ \sin\omega \end{pmatrix} d\omega \\
- \frac{1}{2\pi}\int_0^{2\pi}\begin{pmatrix} 1 \\ \frac{-2\cos\omega}{Z(P)} \\ \frac{-2\sin\omega}{Z(P)} \end{pmatrix} d\omega \int_{t^n}^{t^n+\tau}\left\{u\frac{d}{dt}(Z\cos\theta) + v\frac{d}{dt}(Z\sin\theta)\right\}(\tilde{Q})d\tilde{t} \\
- \frac{1}{2\pi}\int_0^{2\pi}\begin{pmatrix} 1 \\ \frac{-2\cos\omega}{Z(P)} \\ \frac{-2\sin\omega}{Z(P)} \end{pmatrix} d\omega \int_{t^n}^{t^n+\tau}(ZS)(\tilde{Q})d\tilde{t},
\end{aligned} \tag{9}$$

where

$$S := c\left\{\partial_x u\sin^2\theta - (\partial_y u + \partial_x v)\sin\theta\cos\theta + \partial_y v\cos^2\theta\right\}. \tag{10}$$

We begin the approximation to the operator (9) by applying the rectangular quadrature rule in time. The approximation of the last integral (involving the term S) is done exactly as in [1] and hence we do not elaborate them here. However, the approximation of the first two terms in (9) are done differently as outlined below. Let

$$I := (Z\cos\theta u)(Q) + \int_{t^n}^{t^n+\tau} u\frac{d}{dt}(Z\cos\theta)(\tilde{Q})d\tilde{t} \tag{11}$$

Using Taylor development for $Z(Q)\cos\theta$ in the first summand of (11) and rectangle rule for the time integral in the second summand yields

$$\begin{aligned}
I &= u(Q)\left\{Z(P)\cos\omega + (t^n - (t^n+\tau))\frac{d}{dt}(Z\cos\theta)(P)\right\} + \mathcal{O}(\tau^2) \\
&\quad + \tau u\frac{d}{dt}(Z\cos\theta)(Q) + \mathcal{O}(\tau^2). \\
&= u(Q)\left\{Z(P)\cos\omega - \tau\frac{d}{dt}(Z\cos\theta)(P)\right\} + \tau u(Q)\frac{d}{dt}(Z\cos\theta)(P) + \mathcal{O}(\tau^2) \\
&= u(Q)Z(P)\cos\omega + \mathcal{O}(\tau^2).
\end{aligned} \tag{12}$$

The terms involving v are treated analogously. Using these approximations together with the approximation of the source term integrals as in [1] yields the approximate evolution operators, e.g. for pressure,

$$\begin{aligned}
p(P) =& \frac{1}{2\pi}\int_0^{2\pi}[p - Z(P)(u\cos\omega + v\sin\omega)]d\omega \\
&- \frac{1}{2\pi}\sum_j\Big[Z(\omega)(-u\sin\omega + v\cos\omega)\Big]_{\omega_j^-}^{\omega_j^+} \\
&- \frac{1}{2\pi}\int_0^{2\pi}\left[u\frac{(dZ(\omega)\sin\omega)}{d\omega} - v\frac{d(Z(\omega)\cos\omega)}{d\omega}\right]d\omega.
\end{aligned} \tag{13}$$

The expressions for u and v are analogous; see also [1, 3, 4] for more details on the use of the evolution operators in a finite volume framework, leading to the FVEG scheme.

4. **Numerical Case Studies.** In this section we demonstrate the performance of our scheme for smooth and non-smooth data. The scheme is implemented as follows: In order to avoid the overlap of the discontinuities along the integration paths in (13), both c and Z are stored at the centres of a staggered grid, whereas p, u and v are stored in the physical grid; see also [1]. We use a piecewise linear reconstruction for Z, p, u and v on their respective grids with the minmod limiter to limit overshoots and undershoots of the linearly recovered approximations. The rectangle rule is used for spatial integral of flux (5) on the edges and for all the numerical experiments performed, the CFL number is set to be 0.5.

4.1. **Order of Convergence.** Our first goal is to demonstrate the second order convergence of the scheme by computing the experimental order of convergence (EOC). To this end, we choose smooth coefficients and initial data

$$\begin{aligned}\rho(x,y) = K(x,y) &= 1 + \frac{1}{4}\left(\sin(4\pi x) + \cos(4\pi y)\right),\\ p(x,y,0) = \sin(2\pi x) + \cos(2\pi y),&\ u(x,y,0) = v(x,y,0) = 0.\end{aligned}$$

The computational domain $[0,1] \times [0,1]$ is successively divided into $10 \times 10, 20 \times 20, \ldots, 320 \times 320$ mesh cells and the final time is set to $t = 1.0$. The boundary conditions are periodic everywhere. Since an exact solution of this initial value problem is not available, the numerical solution obtained an $N \times N$ grid is compared to the one obtained on a $2N \times 2N$ grid. The errors in p, u and v and the corresponding EOCs obtained in the L^1 norm is shown in table 1. The table clearly shows the second order convergence of the scheme.

N	error of p	EOC	error of u	EOC	error of v	EOC
10	8.52E-02	-	7.54E-02	-	5.54E-02	-
20	3.96E-02	1.11	2.57E-02	1.55	1.55E-02	1.84
40	1.70E-02	1.22	5.93E-03	2.12	4.60E-03	1.75
80	3.53E-03	2.27	1.37E-03	2.11	1.24E-03	1.89
160	6.35E-04	2.47	3.05E-04	2.17	3.00E-04	2.05
320	1.32E-04	2.27	7.30E-05	2.06	7.37E-05	2.02

TABLE 1. Wave propagation in a medium with smoothly varying density and bulk modulus: EOCs for p, u and v measured in the L^1-norm.

4.2. **Wave Propagation in a Heterogeneous Layered Medium.** This test problem is motivated an anlogous study in [1] and the problem models the propagation of a pressure pulse through a heterogeneous layered medium with a single interface. The density and bulk modulus are initialised as

$$(\rho(x,y), K(x,y)) = \begin{cases}(1,1), & \text{if } x \leq 0.5,\\ (4,2), & \text{otherwise.}\end{cases}$$

The initial data read

$$\begin{aligned}p(x,y,0) &= \begin{cases}1 + 0.5(\cos(\pi r/0.1) - 1), & \text{if } r \leq 0.1,\\ 0, & \text{otherwise,}\end{cases}\\ u(x,y,0) &= 0 = v(x,y,0),\end{aligned}$$

where r denotes the distance $r = \sqrt{(x-0.25)^2 + (y-0.4)^2}$. The computational domain is $[-0.95, 1.05] \times [-0.8, 1.6]$ and the boundary conditions are absorbing via simple extrapolation of the variables on all sides. The contours of p, u and v at times $t = 0.2, 0.4, 0.6$ and 0.8 are plotted in Figure 1. In the figure we clearly notice a good resolution of the circular waves, which confirms the genuine multidimensional behaviour of the FVEG scheme. There are no spurious oscillations at the interface and the deformation of the wave due to the change in the medium is captured very well. Due to the jump in the impedances of the media, a part of the wave is reflected backwards as seen in the plots at $t = 0.4$ onwards.

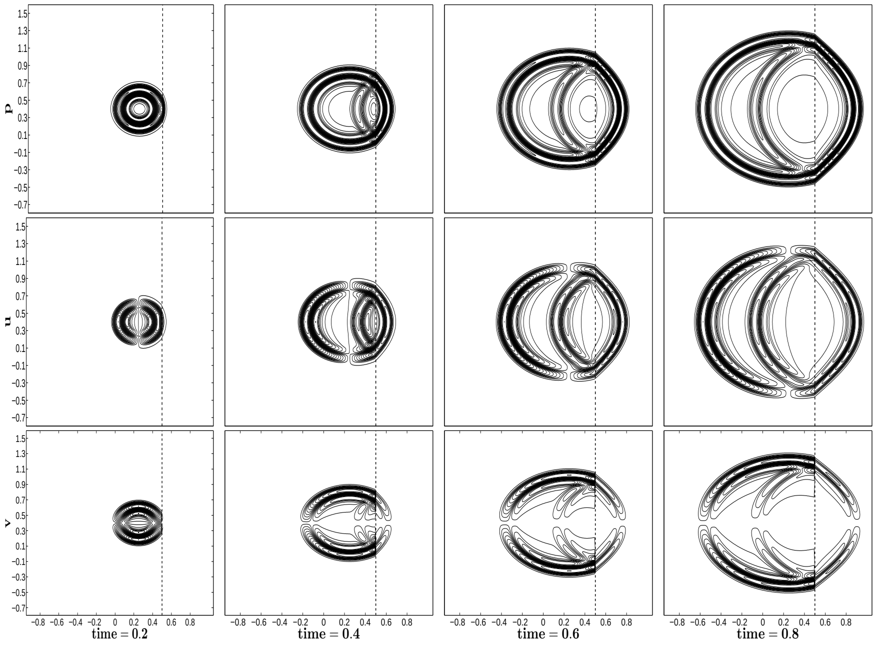


FIGURE 1. Waves passing through the interface of a layered medium.

4.3. **Waves passing through a wavy interface.** In this test we simulate the waves passing through a complex, wavy interface, which is not aligned to the grid. The initial values of p, u and v are same as in the previous problem. The material parameters K and ρ are initialised as

$$K(x,y) = 1, \;\; \rho(x,y) = \begin{cases} 1, & \text{if } x \le 0.5\cos(2\pi(y-0.4)) + 0.4, \\ 4, & \text{otherwise.} \end{cases}$$

The computational domain is $[-0.95, 1.2] \times [-0.675, 1.475]$ and the boundary conditions are absorbing everywhere. The isolines of the solutions at times $t = 0.2, 0.4, 0.6$ and 1.0 are depicted in Figure 2. As in the previous problem we observe both the reflection and transmission of the waves at the interface. However, due to the wavy geometry of the material interface, a complex flow pattern of the reflected waves can be observed at the interface.

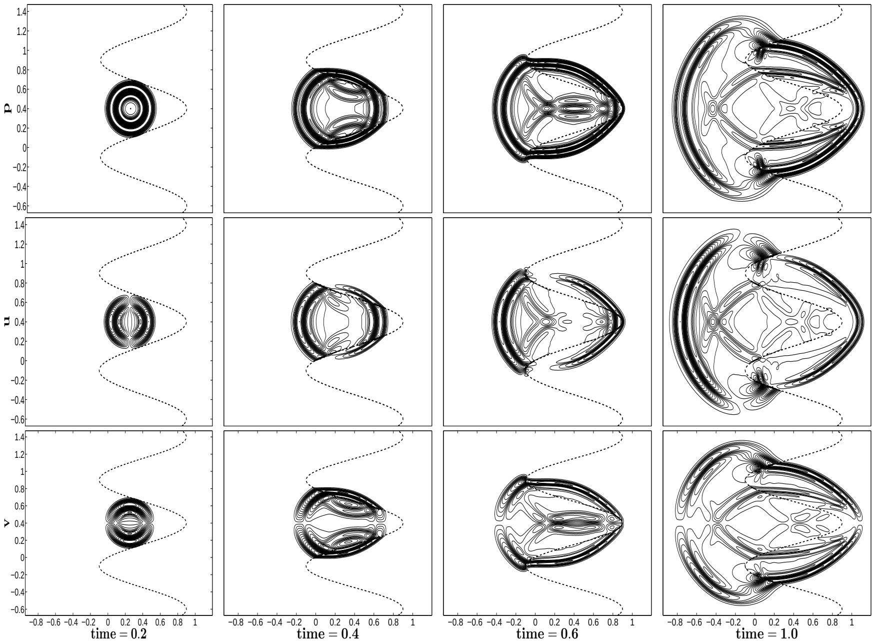


Figure 2. Waves passing through a wavy interface of a layered medium.

4.4. **Wave Propagation through a Nonsmooth Interface.** This test is taken from the reference [2]. The setup consists of a planar square wave pressure pulse passing through a heterogeneous medium with piecewise constant density and bulk modulus. The density and bulk modulus have the initial values

$$\rho(x,y)=1.0,\ K(x,y)=\begin{cases}0.25 & \text{if } x>0 \text{ and } y<0.55x,\\ 1.0 & \text{otherwise.}\end{cases}$$

The initial data read

$$v(x,y,0)=0,\ p(x,y,0)=u(x,y,0)=\begin{cases}1 & \text{if } -0.35<x<-0.2,\\ 0 & \text{otherwise.}\end{cases}$$

We apply periodic boundary conditions and the simulations are performed for $t=0.4, 0.6$ and 1.0. The isolines of the pressure obtained on a 100×100 mesh are plotted in Figure 3 and for the sake of comparison we also plot the pressure obtained on finer mesh of 400×400 cells. The results clearly show the reflection and transmission of waves at the interface. After passing through interface, a part of the waves get reflected off due to the ramp-like geometry of the interface. It has to be noted that both reflected and transmitted waves are oblique to the grid and the genuinely multidimensional FVEG scheme resolve these waves without any grid alignment effect or spurious oscillations.

REFERENCES

[1] K. R. Arun, M. Kraft, M. Lukáčová-Medviďová, and P. Prasad. Finite volume evolution Galerkin method for hyperbolic conservation laws with spatially varying flux functions. *J. Comput. Phys.*, 228:565–590, 2009.

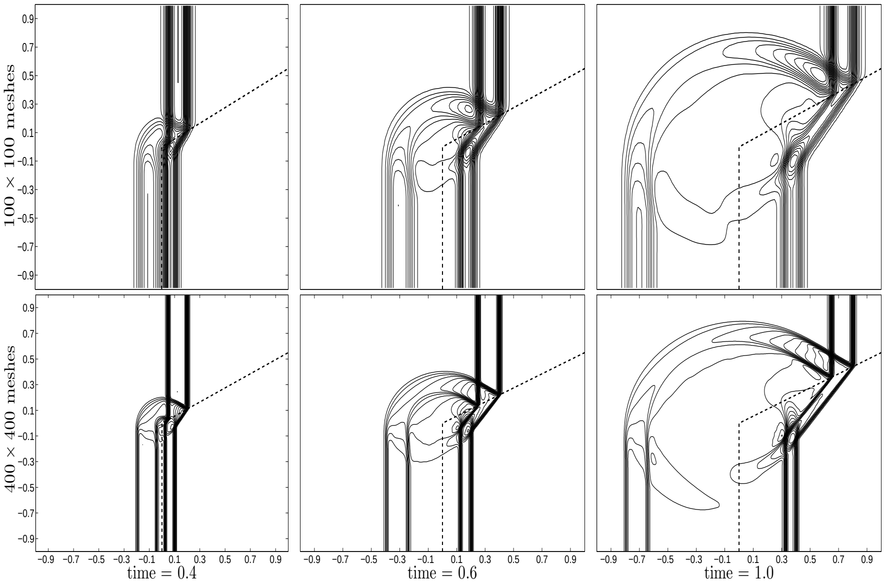


FIGURE 3. Waves passing through a non-smooth interface of a layered medium.

[2] R. J. LeVeque. *Finite Volume Methods for Hyperbolic Problems.* Cambridge University Press, 2002.

[3] M. Lukáčová-Medviďová, K. W. Morton, and G. Warnecke. Finite volume evolution Galerkin (FVEG) methods for hyperbolic problems. *SIAM J. Sci. Comput.*, 26:1–30, 2004.

[4] M. Lukáčová-Medviďová, J. Saibertová, and G. Warnecke. Finite volume evolution Galerkin methods for nonlinear hyperbolic systems. *J. Comput. Phys.*, 183:533–562, 2002.

[5] P. Prasad. *Nonlinear Hyperbolic Waves in Multi-dimensions.* Chapman and Hall/CRC, London, 2001.

E-mail address: arun@igpm.rwth-aachen.de
E-mail address: gxchen.math@whu.edu.cn
E-mail address: noelle@igpm.rwth-aachen.de

ON THE ASYMPTOTIC STABILIZATION OF A GENERALIZED HYPERELASTIC-ROD WAVE EQUATION

Fabio Ancona

Department of Mathematics, University of Padova
Via Trieste 63, 35121 Padova, Italy

Giuseppe Maria Coclite

Department of Mathematics, University of Bari
Via E. Orabona 4, 70125 Bari, Italy

Abstract. We discuss the problem of asymptotic stabilization of the hyperelastic-rod wave equation on the real line

$$\partial_t u - \partial^3_{txx} u + 3u\partial_x u = \gamma\left(2\partial_x u\, \partial^2_{xx} u + u\, \partial^3_{xxx} u\right), \quad t > 0,\ x \in \mathbb{R}.$$

We consider the equation with an additional forcing term of the form $f : H^1(\mathbb{R}) \to H^{-1}(\mathbb{R})$, $f[u] = -\lambda(u - \partial^2_{xx} u)$, for some $\lambda > 0$. We resume the results of [1] on the existence of a semigroup of global weak dissipative solutions of the corresponding closed-loop system defined for every initial data $u_0 \in H^1(\mathbb{R})$. Any such solution decays exponentially to 0 as $t \to \infty$.

1. **Introduction.** In this note we present some recent results obtained in [1] on the problem of asymptotic stabilization of the *hyperelastic-rod wave equation* on the real line

$$\partial_t u - \partial^3_{txx} u + 3u\partial_x u = \gamma\big(2\partial_x u \partial^2_{xx} u + u\partial^3_{xxx} u\big), \quad t > 0, x \in \mathbb{R}. \tag{1}$$

Here $u(t, x)$ represents a (small amplitude) radial deformation in a cylindrical compressible hyperelastic rod and $\gamma > 0$ is some given constant depending on the material and on the prestress of the rod (see Dai [8, 7, 9]). When $\gamma = 1$, the equation (1) can be also seen as a model of propagation of water waves in the shallow water regime with a flat bottom, the classical *Camassa-Holm equation* [5, 15], where $u(t, x)$ represents the fluid velocity.

The asymptotic stabilizability of the Camassa–Holm equation through a stationary feedback law was recently established in [10] by means of a forcing term acting as a control, within the space of H^2 solutions on a torus, and in [16] by means of a boundary feedback, for H^2 solutions on a bounded interval. On the other hand, it is well known that there exist solitary wave solutions of (1), called *peakons*, that have discontinuous first derivatives, as well as solutions of (1) whose slope blows up in finite time. The existence of solutions of the Camassa–Holm equation and of the hyperelastic-rod wave equation that possess only H^1 regularity was shown in various papers of the last decade (see [2, 3, 6, 11]). We have thus addressed

2000 *Mathematics Subject Classification.* Primary: 35G25, 35L05; Secondary: 35B40.

Key words and phrases. Hyperelastic-rod wave equation, Camassa–Holm equation, weak solutions, asymptotic stabilization.

in [1] the problem of constructing a feedback law for (1) that makes the corresponding closed-loop dynamic asymptotically stable at the equilibrium state for solutions with initial data

$$u(0,\cdot)=u_0\in H^1(\mathbb{R}). \tag{2}$$

As a first step in this direction, we have considered the equation (1) with an additional forcing term acting on the whole line of the form

$$f[u]\doteq-\lambda(u-\partial^2_{xx}u), \tag{3}$$

for some $\lambda>0$. We are thus concerned with the Cauchy problem for the nonlinear equation

$$\partial_t u-\partial^3_{txx}u+3u\partial_x u=\gamma\big(2\partial_x u\partial^2_{xx}u+u\partial^3_{xxx}u\big)-\lambda(u-\partial^2_{xx}u),\quad t>0, x\in\mathbb{R}. \tag{4}$$

Rewriting (4) as

$$(1-\partial^2_{xx})\partial_t u+\gamma(1-\partial^2_{xx})(u\partial_x u)+\partial_x\left(\frac{3-\gamma}{2}u^2+\frac{\gamma}{2}(\partial_x u)^2\right)=-\lambda(1-\partial^2_{xx})u,$$

and applying the operator $(1-\partial^2_{xx})^{-1}$, we obtain the formally equivalent elliptic-hyperbolic system

$$\partial_t u+\gamma u\partial_x u+\partial_x P=-\lambda u, \tag{5}$$

$$-\partial^2_{xx}P+P=\frac{3-\gamma}{2}u^2+\frac{\gamma}{2}\left(\partial_x u\right)^2. \tag{6}$$

Since $e^{-|x|}/2$ is the Green's function of the Helmholtz operator $-\partial^2_{xx}+1$, we see that a Lipschitz continuous map $t\to u(t,\cdot)$ with values in $H^1(\mathbb{R})$ will be a solution of (5)-(6) if it satisfies the equality

$$\frac{d}{dt}u=-\gamma u\partial_x u-\partial_x P[u]-\lambda u,\qquad \text{for a.e. } t\geq 0, \tag{7}$$

with the source term $P[u]$ defined as a convolution:

$$P[u]\doteq\frac{e^{-|x|}}{2}*\left(\frac{3-\gamma}{2}u^2+\frac{\gamma}{2}(\partial_x u)^2\right). \tag{8}$$

Observe that, if u is a smooth solution of (5)-(6), differentiating (5) w.r.t. x, then multiplying the equation (5) by u and its x-differentiation by $\partial_x u$, we get

$$\partial_t\left(\frac{u^2+(\partial_x u)^2}{2}\right)+\partial_x\left(\frac{\gamma}{2}u(\partial_x u)^2-\frac{1-\gamma}{2}u^3+uP\right)=-\lambda\left(u^2+(\partial_x u)^2\right). \tag{9}$$

Hence, the total energy

$$E(t):=\|u(t,\cdot)\|^2_{H^1(\mathbb{R})}=\int_{\mathbb{R}}\left(u(t,x)^2+(\partial_x u(t,x))^2\right)dx \tag{10}$$

for smooth solutions satisfies the ordinary differential equation

$$\frac{d}{dt}E(t)=-2\lambda E(t)$$

and we have

$$E(t)=E(0)e^{-2\lambda t},\qquad t\geq 0. \tag{11}$$

On the other hand, as observed before, the hyperelastic-rod wave equation (1) possesses solutions that produce wave breaking in finite time, in the sense that the functions remain uniformly bounded while the spatial derivatives becomes unbounded. Moreover, when solutions experience the presence of peakons (moving to the right) and antipeakons (moving to the left) that collide annihilating each other, it immediately raises the question about the behaviour of the solution after the collision time. Two different scenarios can be put forward (cfr. [12]-[13] and references therein). The first one assumes a total annihilation at the collision time, which leads to the null solution for all times after the annihilation has taken place. Instead, the second scenario assumes that a switching phenomenon occurs: the waves pass through each other, each one continuing unscathed as a solitary wave. For the hyperelastic-rod wave equation with forcing term (4), in view of (11), it is natural to consider dissipative solutions, which correspond to the first scenario where a partial or even total loss of energy can occur during wave breaking. Moreover, we shall require that our solutions satisfy an Oleynik type inequality that should characterize such solutions within the class of dissipative solutions. We thus employ the following

Definition 1.1. Given an initial datum $u_0 \in H^1(\mathbb{R})$, we say that an Hölder continuous function $u = u(t,x)$ defined on $[0,\infty) \times \mathbb{R}$ and such that $u(t,\cdot) \in H^1(\mathbb{R})$ at every $t \in [0,\infty)$, is a weak solution of the Cauchy problem (5)-(6),(2), on $[0,\infty)$ if the map $t \to u(t,\cdot)$ is Lipschitz continuous from $[0,\infty)$ into $L^2(\mathbb{R})$, the initial condition (2) holds and (7) is satisfied as equality between functions in $L^2(\mathbb{R})$. Moreover, we say that u is dissipative if its energy $E(t)$ in (10) is a non-increasing function of time and if it satisfies the inequality

$$\partial_x u(t,x) \leq C\left(1 + \frac{1}{t}\right) \qquad \forall\, t > 0,\ \text{a.e. } x \in \mathbb{R}, \tag{12}$$

for some constant $C > 0$. Here and in (7) $\partial_x u$ is understood as the distributional derivative of u.

Following the approach for the analysis of the Camassa-Holm equation developed in [2]-[3], we have construct in [1] a semigroup of dissipative weak solutions of (5)-(6) on the space $H^1(\mathbb{R})$, whose energy decays exponentially in time. More precisely, our main result in [1] is the following:

Theorem 1.2. *There exists a semigroup $S : [0,\infty) \times H^1(\mathbb{R}) \to H^1(\mathbb{R})$ that enjoys the following properties.*

i. For every $u_0 \in H^1(\mathbb{R})$, the function $u(t,x) \doteq S_t(u_0)(x)$ is a dissipative weak solution of (5)-(6),(2) in the sense of Definition 1.1. Moreover, the constant $C > 0$ in (12) depends only on $\|u_0\|_{H^1(\mathbb{R})}$.

ii. The total energy introduced in (10) decays exponentially, namely

$$E(t) \leq E(0)e^{-2\lambda t}, \qquad t \geq 0. \tag{13}$$

iii. For every given sequence of initial data $\{u_{0,n}\}_n \subset H^1(\mathbb{R})$ and $u_0 \in H^1(\mathbb{R})$, one has

$$u_{0,n} \to u_0 \textit{ in } H^1(\mathbb{R}) \implies S(u_{0,n}) \to S(u_0) \textit{ in } L^\infty_{loc}((0,\infty) \times \mathbb{R}). \tag{14}$$

Notice that the semigroup S of dissipative solutions provided by Theorem 1.2 is not continuous as a map with values in H^1. In fact, even a single trajectory $t \mapsto S_t(u_0)$ may fail to be continuous with respect to the the H^1-norm. Indeed, as observed in [3, Section 7] for the Camassa-Holm equation, one can construct a dissipative solution of (5)-(6) with a pair of peakons and antipeakons of the same strength that collide and completely annihilate each other, thus producing a discontinuity in the H^1-norm at the time of annihilation.

In the following sections we will we will sketch the main arguments envolved in the proof of Theorem 1.2. For all details we refer to [1].

2. **Semilinear System.** Let u be a smooth solution of (1). Following [1, 3] we introduce an energy variable $\xi \in \mathbb{R}$. This will play the role of a Lagrangian variable, remaining constant along characteristics. The map

$$y \in \mathbb{R} \longmapsto \int_0^y \left(1 + (\partial_x u_0)^2\right) dx$$

is continuous, increasing, and goes to ∞ and $-\infty$ as $y \to \infty$ and $y \to -\infty$, respectively. So we can define implicitly the function $y_0 = y_0(\xi)$ by the relation

$$\int_0^{y_0(\xi)} \left(1 + (\partial_x u_0)^2\right) dx = \xi, \qquad \xi \in \mathbb{R}. \tag{15}$$

Let $t \mapsto y(t, \xi)$ be the characteristic curve starting at $y_0(\xi)$, namely

$$\partial_t y(t, \xi) = \gamma u(t, y(t, \xi)), \qquad y(0, \xi) = y_0(\xi). \tag{16}$$

Throughout the following, we use the notation

$$u(t, \xi) := u(t, y(t, \xi)), \qquad P(t, \xi) := P(t, y(t, \xi)),$$

and define the variables $v = v(t, \xi)$ and $q = q(t, \xi)$ as

$$v := 2 \arctan(\partial_x u), \qquad q := (1 + (\partial_x u)^2)\partial_\xi y. \tag{17}$$

We have

$$q \geq 0. \tag{18}$$

Recall that

$$\begin{aligned} P(t, x) =& \frac{1}{2} \int_{\mathbb{R}} e^{-|x-y|} \left(\frac{3-\gamma}{2} u(t, y)^2 + \frac{\gamma}{2} \left(\partial_x u(t, y)\right)^2\right) dy, \\ \partial_x P(t, x) =& \frac{1}{2} \int_{\mathbb{R}} e^{-|x-y|} \operatorname{sign}(y - x) \left(\frac{3-\gamma}{2} u(t, y)^2 + \frac{\gamma}{2} \left(\partial_x u(t, y)\right)^2\right) dy. \end{aligned} \tag{19}$$

The system satisfied by $u = u(t, \xi)$, $v = v(t, \xi)$, $q = q(t, \xi)$ is

$$\begin{cases} \partial_t u = -\partial_x P - \lambda u, \\ \partial_t v = \left(\frac{3-\gamma}{2} u^2 - P\right)(1 + \cos(v)) - \gamma \sin^2\left(\frac{v}{2}\right) - \lambda \sin(v), \\ \partial_t q = \left(\frac{3-\gamma}{2} u^2 - P + \frac{\gamma}{2}\right) \sin(v) q - 2\lambda \sin^2\left(\frac{v}{2}\right) q, \\ u(0, \xi) = u_0(y_0(\xi)), \quad v(0, \xi) = 2 \arctan(\partial_x u_0(y_0(\xi))), \quad q(0, \xi) = 1, \end{cases} \tag{20}$$

where

$$\begin{aligned} P(t, \xi) =& \frac{1}{2} \int_{\mathbb{R}} e^{-\left|\int_\xi^{\xi'} \cos^2\left(\frac{v(t,s)}{2}\right) q(t,s) ds\right|} \times \\ & \times \left(\frac{3-\gamma}{2} u(t, \xi')^2 \cos^2\left(\frac{v(t, \xi')}{2}\right) + \frac{\gamma}{2} \sin^2\left(\frac{v(t, \xi')}{2}\right)\right) q(t, \xi') d\xi', \\ \partial_x P(t, \xi) =& \frac{1}{2} \int_{\mathbb{R}} e^{-\left|\int_\xi^{\xi'} \cos^2\left(\frac{v(t,s)}{2}\right) q(t,s) ds\right|} \times \\ & \times \operatorname{sign}(xi - \xi') \left(\frac{3-\gamma}{2} u(t, \xi')^2 \cos^2\left(\frac{v(t, \xi')}{2}\right) + \frac{\gamma}{2} \sin^2\left(\frac{v(t, \xi')}{2}\right)\right) q(t, \xi') d\xi'. \end{aligned}$$

In order to obtain global dissipative solutions, a modification of the system (20) is needed. In essence, we require the following. Assume that, along a given characteristic $t \mapsto y(t, \xi)$, the wave breaks at a first time $t = \tau(\xi)$. Arguing as for the Burgers equation and reminding that $\partial_x u$ satisfies (12), the wave break means

$\partial_x u(t,\xi) \to -\infty$, as $t \to \tau(\xi)^-$. For all $t \geq \tau(\xi)$ we then set $v(t,\xi) \equiv -\pi$ and remove the values of $u(t,\xi)$, $v(t,\xi)$, $q(t,\xi)$ from the computation of P and $\partial_x P$. More precisely, we replace (20) by

$$\begin{cases} \partial_t u = -\partial_x P - \lambda u, \\ \partial_t v = \begin{cases} \left(\frac{3-\gamma}{2}u^2 - P\right)(1+\cos(v)) - \gamma \sin^2\left(\frac{v}{2}\right) - \lambda \sin(v), & \text{if } v > -\pi, \\ 0, & \text{if } v \leq -\pi, \end{cases} \\ \partial_t q = \begin{cases} \left(\frac{3-\gamma}{2}u^2 - P + \frac{\gamma}{2}\right)\sin(v)q - 2\lambda \sin^2\left(\frac{v}{2}\right)q, & \text{if } v > -\pi, \\ 0, & \text{if } v \leq -\pi, \end{cases} \\ u(0,\xi) = u_0(y_0(\xi)), \quad v(0,\xi) = 2\arctan(\partial_x u_0(y_0(\xi))), \quad q(0,\xi) = 1, \end{cases} \tag{21}$$

and

$$P(t,\xi) = \frac{1}{2}\int_{\{v(t,\xi')>-\pi\}} e^{-\left|\int_{\{\xi' \leq \xi,\, v(t,\xi')>-\pi\}} \cos^2\left(\frac{v(t,s)}{2}\right)q(t,s)ds\right|} \times$$
$$\times \left(\frac{3-\gamma}{2}u(t,\xi')^2 \cos^2\left(\frac{v(t,\xi')}{2}\right) + \frac{\gamma}{2}\sin^2\left(\frac{v(t,\xi')}{2}\right)\right) q(t,\xi')d\xi',$$
$$\partial_x P(t,\xi) = \frac{1}{2}\int_{\{v(t,\xi')>-\pi\}} e^{-\left|\int_{\{\xi' \leq \xi,\, v(t,\xi')>-\pi\}} \cos^2\left(\frac{v(t,s)}{2}\right)q(t,s)ds\right|} \times$$
$$\times\, \mathrm{sign}(\xi-\xi')\left(\frac{3-\gamma}{2}u(t,\xi')^2 \cos^2\left(\frac{v(t,\xi')}{2}\right) + \frac{\gamma}{2}\sin^2\left(\frac{v(t,\xi')}{2}\right)\right) q(t,\xi')d\xi'.$$

The local existence and uniqueness of the solution is obtained following the main lines of [4], see [1, Theorem 3.1]. A global bound on the total energy

$$E(t) = \int_{\{v(t,\xi)>-\pi\}} \left(u^2(t,\xi)\cos^2\left(\frac{v(t,\xi)}{2}\right) + \sin^2\left(\frac{v(t,\xi)}{2}\right)\right) q(t,\xi)d\xi \tag{22}$$

guarantees that the local solutions of the semilinear system (21) can be globally extended for all times $t \geq 0$, see [1, Section 4].

2.1. **Stability for the semilinear system.** The following statement holds [1, Section 5]

Let $\{u_{0,n}\}_n \subset H^1(\mathbb{R})$ and $u_0 \in H^1(\mathbb{R})$. If

$$u_{0,n} \longrightarrow u_0 \qquad \text{in } H^1(\mathbb{R}), \tag{23}$$

then

$$u_n \longrightarrow u \qquad \text{in } L^\infty((0,T)\times\mathbb{R}) \text{ for every } T > 0, \tag{24}$$

where u_n and u are the solutions of the semilinear dissipative system (21) in correspondence of $u_{0,n}$ and u_0, respectively.

Let (u,v,q) and $(\widetilde{u},\widetilde{v},\widetilde{q})$ be any two solutions of the semilinear dissipative system (21) and $T > 0$.

For every

$$\xi \in \{\xi \in \mathbb{R}; v(T,\xi) = -\pi\} \cup \{\xi \in \mathbb{R}; \widetilde{v}(T,\xi) = -\pi\},$$

let $\tau(\xi)$ be the first time at which one of the two solutions reaches the value $-\pi$, namely

$$\tau(\xi) = \inf\{t \in [0,T]; \min\{v(t,\xi), \widetilde{v}(t,\xi)\} = -\pi\}.$$

Since the map $\tau(\cdot)$ is measurable, we can construct a measure-preserving, measurable map $\alpha \longmapsto \xi(\alpha)$ from $[0,\alpha^*]$ onto Λ such that

$$\alpha \leq \alpha' \Longleftrightarrow \tau(\xi(\alpha')) \leq \tau(\xi(\alpha)).$$

The inverse mapping $\xi \longmapsto \alpha(\xi)$ from Λ into $[0, \alpha^*]$ is still measure-preserving.

We introduce the distance functional

$$\begin{aligned} j(t) =& J((u(t,\cdot), v(t,\cdot), q(t,\cdot)), (\widetilde{u}(t,\cdot), \widetilde{v}(t,\cdot), \widetilde{q}(t,\cdot))) \\ =& \|u(t,\cdot) - \widetilde{u}(t,\cdot)\|_{L^\infty(\mathbb{R})} + \|v(t,\cdot) - \widetilde{v}(t,\cdot)\|_{L^2(\mathbb{R})} + \|q(t,\cdot) - \widetilde{q}(t,\cdot)\|_{L^2(\mathbb{R})} \\ &+ K_0 \int_0^{\alpha^*} e^{K\alpha} |v(t, \xi(\alpha)) - \widetilde{v}(t, \xi(\alpha))| d\alpha. \end{aligned}$$

Choosing suitably the positive constants K and K_0 we get

$$\frac{d}{dt} j(t) \leq M j(t),$$

for some constant $M > 0$. Therefore

$$j(t) \leq e^{Mt} j(0), \qquad 0 \leq t \leq T. \tag{25}$$

Consider a sequence $\{u_{0,n}\}_n \subset H^1(\mathbb{R})$ and $u_0 \in H^1(\mathbb{R})$ satisfying (23). The boundedness of $\{\|u_{0,n}\|_{H^1(\mathbb{R})}\}_n$, the Sobolev embedding $H^1(\mathbb{R}) \subset L^\infty(\mathbb{R})$ (see [14, Theorem 8.5]), and (25) imply (24). Then our claim is proved.

3. Global Dissipative Solutions in the Original Variables.

Let (u, v, q) be the solution of the semilinear system (21). Define

$$y(t, \xi) = y_0(\xi) + \int_0^t u(\tau, \xi) d\tau.$$

For each fixed ξ, the function $t \longmapsto y(t, \xi)$ solves

$$\partial_t y(t, \xi) = u(t, \xi), \qquad y(0, \xi) = y_0(\xi).$$

We set

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

Clearly

$$u(0, x) = u_0(x) \quad x \in \mathbb{R}.$$

Due to the energy estimate on $\|u(t,\cdot)\|_{H^1(\mathbb{R})}$ and the fact that the image of the singular set where $v = -\pi$ has measure zero (in the x-variable), i.e.,

$$\text{meas}(\{y(t, \xi); v(t, \xi) = -\pi\}) = 0$$

we have that u is continuous, $t \longmapsto u(t, \cdot) \in L^2(\mathbb{R})$ is Lipschitz continuous, and

$$\frac{d}{dt} u = -\gamma u \partial_x u - \partial_x P - \lambda u.$$

Therefore u is a weak solution of the hyperelastic rod wave equation in the sense of Definition 1.1.

3.1. The Semigroup.

Given an initial data $u_0 \in H^1(\mathbb{R})$, we denote by $u(t, x) = S_t(u_0)(x)$ the corresponding global solution of the hyperelastic-rod wave equation. We have to prove

$$S_t(S_\tau(u_0)) = S_{t+\tau}(u_0), \qquad t, \tau > 0.$$

Let (u, v, q) be the solution of the problem in the auxiliary variables. Call $\widetilde{u} = S_\tau(u_0)$. We choose ξ_0 such that $y(\tau, \xi_0) = 0$ and consider the new energy variable $\sigma = \sigma(\xi)$ as a solution of

$$\frac{d}{d\xi} \sigma(\xi) = \begin{cases} q(\tau, \xi) & \text{if } v(\tau, \xi) > -\pi, \\ 0 & \text{if } v(\tau, \xi) = -\pi, \end{cases} \qquad \sigma(\xi_0) = 0.$$

We have

$$\int_0^{y(\tau,\xi)} \left(1+(\partial_x u(\tau,x))^2\right) dx = \sigma(\xi).$$

The map $\xi = \xi(\sigma)$

$$\xi(\overline{\sigma}) = \sup\{s; \sigma(s) \le \overline{\sigma}\}$$

provides almost everywhere an inverse of $\sigma(\cdot)$.

Define

$$\widetilde{u}(t,\sigma) = u(\tau+t,\xi(\sigma)), \quad \widetilde{v}(t,\sigma) = v(\tau+t,\xi(\sigma)), \quad \widetilde{q}(t,\sigma) = \frac{q(\tau+t,\xi(\sigma))}{q(\tau,\xi(\sigma))}.$$

Since $(\widetilde{u},\widetilde{v},\widetilde{q})$ solves the same equations of (u,v,q), S is a semigroup.

3.2. **The decay as $t \longrightarrow \infty$.** Given an initial data $u_0 \in H^1(\mathbb{R})$, let $u = u(t,x)$ be the corresponding global solution of the hyperelastic-rod wave equation and (u,v,q) be the solution of the problem in the auxiliary variables. Consider

$$\begin{aligned} E(t) =& \|u(t,\cdot)\|^2_{H^1(\mathbb{R})} = \int_{\mathbb{R}} \left(u(t,x)^2 + (\partial_x u(t,x))^2\right) dx \\ =& \int_{\{v(t,\xi)>-\pi\}} \left(u^2(t,\xi)\cos^2\left(\frac{v(t,\xi)}{2}\right) + \sin^2\left(\frac{v(t,\xi)}{2}\right)\right) q(t,\xi)d\xi. \end{aligned}$$

We have

$$\frac{d}{dt}E(t) \le -2\lambda E(t).$$

Therefore

$$E(t) \le e^{-2\lambda t}E(0), \qquad t \ge 0.$$

3.3. **The Oleinik type estimate.** Since

$$\partial_t v\Big|_{v=\pi} = \left(\frac{3-\gamma}{2}u^2 - P\right)(1+\cos(v)) - \gamma\sin^2\left(\frac{v}{2}\right) - \lambda\sin(v)\Big|_{v=\pi} = -\gamma,$$

we can choose $\delta > 0$ so that

$$\partial_t v(t,\xi) \le -\frac{\gamma}{2}, \qquad v \in [\pi-\delta,\pi).$$

As a consequence

$$v(t,\xi) < \min\left\{\pi-\delta, \pi - \frac{t\gamma}{2}\right\}, \qquad v \in [\pi-\delta,\pi).$$

Hence (12) follows from the identity

$$\partial_x u = \frac{\sin(v)}{1+\cos(v)}.$$

4. **Further Problems.** It is clear that [1] is only the first attempt on the controllability and stabilizability of weak solutions of the hyperelastic-rod wave equation. Indeed the results of [10, 16] apply to H^2 solutions.

In [1] we damp the waves of an hyperelastc-rod adding an external forcing term $f[u]$ acting on all the rod. It is very interesting to consider the case of an external forcing term acting only on a region w of the rod or only on one of the end points of it.

Here we consider the asymptotic stabilization, in the sense that we use our feedback law to damp all the waves of the rod and in the limit we do not want any wave. Of course one can be interested in a particular profile at infinity (say a peakon or any stationary wave). In this case one should find a feedback law damping all the waves except the desired one. Clearly the asymptotic profile can be H^1 as the initial condition or smoother and vice-versa. In other words we can ask for a feedback law that regularize all the singularities generated by the initial condition or on the contrary a feedback laws that is able to generate say a peakon even if the initial profile is not doing that.

REFERENCES

[1] F. Ancona and G. M. Coclite, *Asymptotic Stabilization of weak solutions to a generalized hyperelastic-rod wave equation: dissipative semigroup*, preprint.

[2] A. Bressan and A. Constantin, *Global conservative solutions of the Camassa-Holm equation*, Arch. Ration. Mech. Anal., **183** (2007), 215–239.

[3] A. Bressan and A. Constantin, *Global dissipative solutions of the Camassa-Holm equation*, Analysis and Applications, **5** (2007), 1–27.

[4] A. Bressan and W. Shen. *Unique solutions of directionally continuous ordinary differential equations in Banach spaces*, *Analysis and Applications*, **4** (2006), 247–262.

[5] R. Camassa and D. D. Holm, *An integrable shallow water equation with peaked solitons*, *Phys. Rev. Lett.*, **71** (1993), 1661–1664.

[6] G. M. Coclite, H. Holden, and K. H. Karlsen. Global weak solutions to a generalized hyperelastic-rod wave equation. *SIAM J. Math. Anal.*, **37** (2005), 1044–1069.

[7] H.-H. Dai, *Exact travelling-wave solutions of an integrable equation arising in hyperelastic rods*, Wave Motion, **28** (1998), 367–381.

[8] H.-H. Dai, *Model equations for nonlinear dispersive waves in a compressible Mooney–Rivlin rod*, Acta Mech., **127** (1998), 193–207.

[9] H.-H. Dai and Y. Huo, *Solitary shock waves and other travelling waves in a general compressible hyperelastic rod*, R. Soc. Lond. Proc. Ser. A, **456** (2000), 331–363.

[10] O. Glass, *Controllability and asymptotic stabilization of the Camassa-Holm equation*, J. Differential Equations, **245** (2008), 1584–1615.

[11] H. Holden and X. Raynaud. Global conservative solutions of the generalized hyperelastic-rod wave equation. *J. Differential Equations*, **233** (2007), 448–484.

[12] H. Holden and X. Raynaud. Global conservative multipeakons solutions of the Camassa-Holm equation. *J. Hyperbolic Differ. Equ.* , **4** (2007), 39–64.

[13] H. Holden and X. Raynaud. Global dissipative multipeakons solutions of the Camassa-Holm equation. *Comm. Partial Differential Equations*, **33** (2008), 2040–2063.

[14] E. H. Lieb and M. Loss, “Analysis”, American Mathematical Society, Providence, RI, 2001.

[15] R. S. Johnson, *Camassa–Holm, Korteweg–de Vries and related models for water waves*, J. Fluid Mech., **455** (2002), 63–82.

[16] V. Perrollaz, *Initial boundary value problem and asymptotic stabilization of the Camassa-Holm equation on an interval*, J. Funct. Anal., **259** (2010), 2333–2365.

E-mail address: ancona@math.unipd.it
E-mail address: coclitegm@dm.uniba.it

BOUNDARY TREATMENT IN GHOST POINT FINITE DIFFERENCE METHODS FOR COMPRESSIBLE GAS DYNAMICS IN DOMAIN WITH MOVING BOUNDARIES

Armando Coco

Dipartimento di Scienze della Terra e Geoambientali
Università degli Studi di Bari Aldo Moro
Via Orabona, 4
70125, Bari, Italy

Giovanni Russo

Dipartimento di Matematica e Informatica
Università degli Studi di Catania
Viale Andrea Doria, 6
95125, Catania, Italy

Abstract. We propose a set of boundary conditions to be employed in the context of the Euler equation for compressible gas dynamics with fixed or moving smooth impenetrable obstacles in two space dimensions. The conditions are prescribed on the primitive variables. The equations are discretized on a regular Cartesian grid, in which the domain is described by a level set function. The evolution of interior points is determined by solving the Euler equations, discretized by a finite difference shock capturing scheme. Ghost points out of the domain, near the boundary, are used to close the system. The value of the field variables at the boundary are obtained by discretization of the boundary conditions. For fixed domain, in addition to classical conditions of impenetrability and pressure gradient balancing centrifugal force, we adopt two additional conditions, one relating pressure and density gradient, and one on the normal component of the tangential velocity. The last condition is derived by imposing that the flow near the boundary is locally irrotational. Equivalence between this condition and the condition of constant enthalpy is shown. Generalization to moving domains is derived. An iterative technique to discretize the boundary conditions on the ghost points, which is based on an approach recently applied in the context of elliptic problems [5], is presented. Some numerical results for moving boundaries are illustrated, that show the effect of the additional terms on the pressure gradient which are not present in the case of fixed boundary.

1. **Introduction.** We are interested in developing a simple, accurate, and robust numerical method for the boundary treatment in compressible fluid dynamics in presence of solid obstacles. In the two-dimensional (2-D) case, the governing equations are the compressible Euler equations, written in conservation form:

$$\frac{\partial w}{\partial t} + \frac{\partial f(w)}{\partial x} + \frac{\partial g(w)}{\partial y} = 0 \tag{1}$$

2000 *Mathematics Subject Classification.* Primary: 35L65, 65M06; Secondary: 35L50, 76N99.

Key words and phrases. Boundary condition, ghost-point, finite difference, compressible gas dynamics, moving boundaries.

where $w = (\rho, \rho u, \rho v, E)^T$, $f = (\rho u, \rho u^2 + p, \rho uv, u(E+p))^T$, $g = (\rho v, \rho uv, \rho v^2 + p, v(E+p))^T$.

Here ρ is the fluid density, u and v are the velocities, E is the total energy, and p is the pressure. System (1) is closed using the equation of state (EOS), which, for ideal gases, reads:

$$E = \frac{p}{\gamma - 1} + \frac{\rho}{2}(u^2 + v^2), \quad \gamma = \text{const.} \tag{2}$$

We also introduce the notation $c_s^2 := (\partial p/\partial \rho)_s = \gamma p/\rho$ for the square of the sound speed, which will be used throughout the paper (s denotes the entropy).

The computational domain $\Omega(t) \subseteq \mathcal{R}$ is identified by a region in which a time dependent level-set function $\phi(\mathbf{x}, t)$ is negative. The rectangular region $\mathcal{R}$ in which the $\Omega(t)$ is immersed is discretized by a regular square grid. In the whole paper we assume that Ω is the region in $\mathcal{R}$ external to a fixed or moving obstacle $D(t)$ (see Fig. 1).

Two sets of nodes are identified in $\mathcal{R}$ at each time t: internal nodes $\mathbf{x} \in \Omega(t)$, and *ghost* nodes, i.e. nodes in $\mathcal{R}$, which are external to Ω, but are *close* to the boundary (within one or few grid points from an internal node). For fixed domains, i.e. if Ω does not depend on time, the sets of internal and ghost points do not change, otherwise it has to be updated at every time step. Conservative finite difference will be used as space discretization. In one time step, from t^n to t^{n+1}, the evolution of the system is performed as follows: for points that will be internal at time t^{n+1} the field variables are evolved by integrating the semidiscrete system in time, while the values of all the ghost points which are required to close the system of equations are computed by making use of boundary conditions. This approach has been adopted in the forthcoming paper [2].

For Euler equation, each node contains four quantities in two space dimensions, say density, pressure and the two components of the velocity, therefore four equations are needed for each ghost point. Of course, because of the hyperbolic nature of the problem, the conditions cannot be applied independently, and have to be compatible with the equations.

We assume that the boundary conditions on the obstacle $D(t)$ are the classical no slip conditions of inviscid Euler equation on a wall, so one boundary condition states that the normal velocity of the gas on ∂D is equal to the normal velocity of the boundary. The second condition is obtained balancing centrifugal force on the gas with pressure gradient. The third condition is obtained from adiabaticity, and relates variations in pressure and density, and the last condition, imposed on the transversal velocity, is a condition on the enthalpy, commonly adopted in gas dynamics [6].

Because the conditions on one ghost point are related to the conditions on neighbor ghost points, they are not independent, rather they constitute a system that has to be solved quickly in order to proceed with the integration of the equations on internal points. High order extrapolation will be able to define the equations for the ghost points to high order accuracy in space.

A recently developed relaxation procedure, successfully applied in the numerical solution of elliptic problems [5, 4], is applied to the solution of the system of equations to compute the field at ghost nodes. A detailed exposition on the finite difference discretization for inside equations and numerical tests showing the second order accuracy will be provided in [2]. Other works recently developed about boundary methods in compressible fluid-dynamics are [6, 8, 1].

In presence of moving boundary the condition on the pressure is extended with additional terms (see Sec. 3.3). A numerical test presented in Sec. 4 validates such extension.

2. **Finite difference scheme.** Let us suppose that the domain Ω is defined by a level set function ϕ, namely as the set $\{\vec{x} \in \mathcal{R} \colon \phi(\vec{x}) < 0\}$, while the obstacle is identified by $\{\vec{x} \in \mathcal{R} \colon \phi(\vec{x}) > 0\}$ and the boundary by $\{\vec{x} \in \mathcal{R} \colon \phi(\vec{x}) = 0\}$.

By the level-set function we identify different points (see Fig. 1):

- internal points $\vec{X}_{jk} \in \Omega$, $(j, k \in \mathcal{I})$. These are the points where we solve the problem, and for which we write the differential equation.
- Ghost points $\vec{X}_{jk}$, $(j, k \in \mathcal{G})$. These are points external to Ω, that are *near* the boundary. In particular, for a second order method, they are points within one or two grid cells (in either direction x or y) from the boundary.
- Inactive points
  By inactive points we denote points external to the domain, which are not ghost.

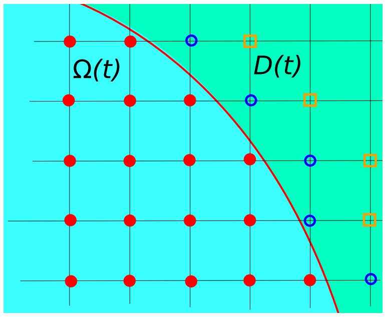


FIGURE 1. Grid points of the setup: the red full circular points are inner points, while the blue empty circular and the yellow empty squared points are respectively the first and second layer of ghost points.

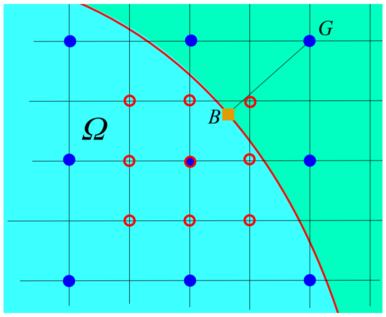


FIGURE 2. $St_G^{(I)}$ is the stencil composed by the red empty circular grid points, while $St_G^{(II)}$ is composed by the blue full circular grid points.

Within ghost points, we distinguish between first layer $\mathcal{L}_1$ and second layer, $\mathcal{L}_2 : \mathcal{L}_1 \cup \mathcal{L}_2 = \mathcal{G}$. The first layer of points is within one grid cell from the boundary (in either direction). The second layer is made of points within two grid points from the boundary, which are not in the first layer.

The Finite difference approximation on system (1) is:

$$\frac{d\,w_{jk}}{dt} + \frac{\widehat{f}_{j+\frac{1}{2},k} - \widehat{f}_{j-\frac{1}{2},k}}{\Delta x} + \frac{\widehat{g}_{j,k+\frac{1}{2}} - \widehat{g}_{j,k-\frac{1}{2}}}{\Delta y}, \quad (j,k) \in \mathcal{I}.$$

These equations require the computation of the fluxes, which are at interface between internal cells, or between internal cells and first layer cells. They are computed as follows:

$$\widehat{f}_{j+\frac{1}{2},k} = \widehat{f}^{+}_{j+\frac{1}{2},k} + \widehat{f}^{-}_{j+\frac{1}{2},k} \tag{3}$$

$$\widehat{f}^{+}_{j+\frac{1}{2},k} = \widehat{F}^{+}_{j,k}\left(x_{j+\frac{1}{2}}, y_k\right), \qquad \widehat{f}^{-}_{j+\frac{1}{2},k} = \widehat{F}^{-}_{j+1,k}\left(x_{j+\frac{1}{2}}, y_k\right) \tag{4}$$

$$\widehat{g}_{j,k+\frac{1}{2}} = \widehat{g}^{+}_{j,k+\frac{1}{2}} + \widehat{g}^{-}_{j,k+\frac{1}{2}} \tag{5}$$

$$\widehat{g}^{+}_{j,k+\frac{1}{2}} = \widehat{G}^{+}_{j,k}\left(x_j, y_{k+\frac{1}{2}}\right), \qquad \widehat{g}^{-}_{j,k+\frac{1}{2}} = \widehat{G}^{-}_{j,k+1}\left(x_j, y_{k+\frac{1}{2}}\right) \tag{6}$$

The four flux functions $\widehat{F}^{\pm}$, $\widehat{G}^{\pm}$ have to be reconstructed in cells $(j,k) \in \mathcal{I} \cup \mathcal{L}_1$ from the pointwise values of the fluxes $f^{\pm}$ and $g^{\pm}$, which, in turn, determine a split of the flux functions:

$$f(w) = f^{+}(w) + f^{-}(w), \quad g(w) = g^{+}(w) + g^{-}(w)$$

We use local Lax-Friedrichs splitting. Once the values of the fluxes are computed at each grid node,

$$f^{\pm}_{j,k} = f^{\pm}(w_{j,k}), \;\; g^{\pm}_{j,k} = g^{\pm}(w_{j,k}), \quad (j,k) \in \mathcal{I} \cup \mathcal{G}.$$

then $\widehat{F}^{\pm}$, $\widehat{G}^{\pm}$ are reconstructed using classical WENO reconstruction from cell averages to pointwise values [7].

The definition of field values at ghost cells requires the solution of equations that arise from the discretization of the boundary conditions.

3. Boundary treatment of Euler equations in 2D.

3.1. Fixed boundary.

3.1. Fixed boundary. We denote by $\vec{n}$ and $\vec{\tau}$ respectively the normal and the tangential unit vector to the boundary, while κ is the signed curvature (see Fig. 3). We assume that the unit normal points outside the fluid domain.

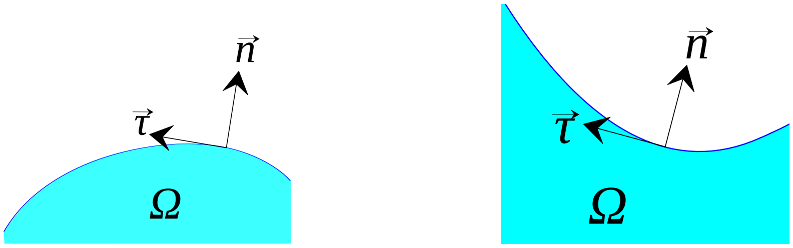


FIGURE 3. Locally convex boundary $\kappa < 0$ (left) and locally concave boundary $\kappa > 0$ (right).

3.1.1. *Condition on the normal component of the velocity.* Let us denote by $u_n = \vec{u} \cdot \vec{n}$ and $u_\tau = \vec{u} \cdot \vec{\tau}$ respectively the normal and tangential velocity. The condition on the normal velocity is:

$$u_n = 0 \text{ on } \partial\Omega, \tag{7}$$

3.1.2. *Condition on the pressure.* The equation of motion for a fluid particle (balance of momentum) for Euler equations reads:

$$\rho \frac{D\vec{u}}{Dt} + \nabla p = 0 \tag{8}$$

where $D/Dt = \partial/\partial t + \vec{u} \cdot \nabla$ denotes the Lagrangian derivative.
Condition (7) implies that, along the boundary of the domain, the velocity vector is $\vec{u} = u_\tau \vec{\tau}$ where u_τ denotes the tangential component. It is therefore:

$$\frac{D\vec{u}}{Dt} = \frac{Du_\tau}{Dt}\vec{\tau} + u_\tau \frac{D\vec{\tau}}{Dt} = a_\tau \vec{\tau} + u_\tau^2\, \kappa\, \vec{n} \tag{9}$$

where κ denotes the curvature. It is $|\kappa| = 1/R$, where R is the local radius of curvature of the boundary. With the notation in the Fig. 3, the sign of κ is negative for locally convex regions, and positive for locally concave regions, and a_τ denotes the tangential acceleration of the fluid. By projecting Eq. (8) on the normal direction, and making use of (9), one obtain the boundary condition on the pressure:

$$\frac{\partial p}{\partial n} = -\rho\, u_\tau^2\, \kappa. \tag{10}$$

3.1.3. *Condition on the tangential component of the velocity.* We impose the following condition for u_τ

$$\frac{\partial u_\tau}{\partial n} = u_\tau\, \kappa. \tag{11}$$

Such condition can be obtained in the following two manners.

- By imposing that the vorticity is zero. This means imposing that $\oint_\Gamma \vec{u} \cdot d\vec{l} = 0$ for each closed circuit Γ. We choose the circuit in Fig. 4, which is composed by two concentric arcs and two segments aligned with the common center of the arcs. The arc with the smaller radius and the boundary are tangent and have the same curvature.

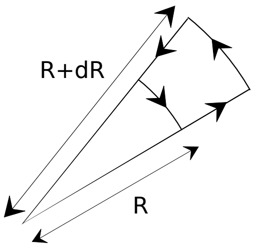


FIGURE 4. We impose the vorticity is zero.

Supposing that the u_τ is constants along the arcs, we obtain:

$$\oint_\Gamma \vec{u} \cdot d\vec{l} = (R + \Delta R)\, u_\tau|_{R+\Delta R} - R\, u_\tau|_R = 0. \tag{12}$$

Using Taylor expansion, and going to the limit $\Delta R \to 0$, we obtain (11) (observing that $|\kappa| = 1/R$).

- By imposing that the normal derivative of the total enthalpy (density) is zero on the boundary (this is commonly adopted in gas dynamics [6]), because enthalpy is conserved for smooth flows, and it is assumed that the far field is a constant flow. Let us recall that the enthalpy is $H = \frac{1}{2}u^2 + e + \frac{p}{\rho}$, where $e = e(\rho, p) = \frac{p}{(\gamma - 1)\rho}$ is the internal energy. The condition reads:
$$0 = \frac{\partial H}{\partial n} = \vec{u} \cdot \frac{\partial \vec{u}}{\partial n} + \left(\frac{1}{\gamma - 1} + 1\right)\left(\frac{\partial p}{\partial n}\frac{1}{\rho} - \frac{p}{\rho^2}\frac{\partial \rho}{\partial n}\right).$$
Using the boundary conditions (10) and (14) and the fact that $\vec{u} = v_\tau \tau$ on the boundary, we obtain:
$$0 = \frac{\partial u_\tau}{\partial n}u_\tau + \frac{\gamma}{\gamma - 1}\frac{\partial p}{\partial n}\left(\frac{1}{\rho} - \frac{p}{c_s^2\rho^2}\right) = \frac{\partial u_\tau}{\partial n}u_\tau + \frac{\gamma}{\gamma - 1}\kappa\, u_\tau^2\left(1 - \frac{p}{c_s^2\rho}\right) \qquad (13)$$
where c_s^2 is the square of the speed sound. For a polytropic gas we have $c_s^2 = \gamma p/\rho$. Therefore:
$$0 = \frac{\partial u_\tau}{\partial n}u_\tau + \kappa\, u_\tau^2$$
which implies the (11).

3.1.4. *Condition on the density.* Finally, the condition on the density is given by the requirement that the boundary is adiabatic, i.e. that locally the entropy is flat (just as in the case of the enthalpy, one uses the fact that entropy is an invariant of the flow, for smooth flows, and assumes that the far field is constant). This means $\partial s/\partial n = 0$, which implies
$$\frac{\partial p}{\partial n} = c_s^2\frac{\partial \rho}{\partial n}. \qquad (14)$$

3.2. **Discretization of the boundary conditions.** We use the boundary conditions (7),(11),(10),(14) to extrapolate $\vec{u}$, p, ρ in ghost points. From now on we assume that the level set function ϕ is actually a signed distance function from the boundary, therefore $|\nabla\phi| = 1$. We also assume that grid spacing in x and y are the same, i.e. $\Delta x = \Delta y = h$.

The linear system coming from the discretization of the boundary conditions is obtained by writing a linear equation for each unknown of the system, i.e. for each $u(G)$, $v(G)$, $p(G)$, $\rho(G)$, where $G \in \mathcal{G}$ is a ghost point. In detail: let G be a ghost point. We compute the projection point B on the interface, making use of the signed distance function ϕ, that is $B \equiv (x_B, y_B) = G - \phi(G)\vec{n}_G = G - \phi(G)\nabla\phi(G)$. Let us define two 3×3 stencils: $St_G^{(I)}$ and $St_G^{(II)}$. Each of this stencil can be divided in four (overlapped) 2×2 sub-stencils. $St_G^{(I)}$ is a stencil in upwind direction containing G, possibly enlarged in such a way that the boundary point B is contained in the 2×2 sub-stencil containing G (see Fig. 2). In formulas:
$$St_G^{(I)} = \left\{(x_G + s_x\, k_1\, h, y_G + s_y\, k_2\, h) \colon (k_1, k_2) \in \{0,1,2\}^2\right\},$$
where $s_x = \lceil |x_B - x_G|/h \rceil \operatorname{sgn}(x_B - x_G)$, $\quad s_y = \lceil |y_B - y_G|/h \rceil \operatorname{sgn}(y_B - y_G)$. Stencil $St_G^{(II)}$ is not enlarged and it contains the boundary point B and in upwind direction. It cannot contain the grid point G. See Fig. 2. Note that $St_G^{(II)} \equiv St_G^{(II)}$ if G belongs to the first layer of ghost points, i.e. $G \in \mathcal{L}_1$. Let us denote by $\mathcal{L}[\omega; St]$ the biquadratic interpolant of ω in the stencil St.

We transform the system of the boundary conditions (7), (11), (10), (14) into a time dependent problem with a fictitious time τ:

$$\begin{aligned}
\frac{\partial u_n}{\partial \tau} + u_n &= 0 && (15)\\
\frac{\partial u_\tau}{\partial \tau} + \mu_1 \frac{\partial u_\tau}{\partial n} &= \mu_1 u_\tau \, \kappa && (16)\\
\frac{\partial p}{\partial \tau} + \mu_2 \frac{\partial p}{\partial n} &= -\mu_2 \rho \, \kappa \, u^2 && (17)\\
\frac{\partial \rho}{\partial \tau} + \mu_3 \frac{\partial \rho}{\partial n} &= \mu_3 \frac{1}{c_s^2} \frac{\partial p}{\partial n} && (18)
\end{aligned}$$

where μ_i, $i = 1, 2, 3$ are suitable constants. The iterative scheme is obtained by discretizing the previous problem in space and time and looking for the steady state solution (the fictitious time τ represents an iterative parameter). The partial derivatives with respect to τ are computed to first order in space (i.e. the quantities are evaluated at the ghost points G rather than in B), since they vanish as τ goes to infinity, while all the other terms are discretized to second order (in the boundary point B). The iterative scheme becomes:

$$\begin{aligned}
u_n^{G\,(m+1)} &= u_n^{G\,(m)} - \Delta\tau \mathcal{L}[u_n^{(m)}; St_G^{(I)}](B) && (19)\\
u_\tau^{G\,(m+1)} &= u_\tau^{G\,(m)} \\
&- \mu_1 \Delta\tau \left(\frac{\partial \mathcal{L}[u_\tau^{(m)}; St_G^{(I)}](B)}{\partial n} - \mathcal{L}[u_\tau^{(m)}; St_G^{(I)}](B)\, \kappa(B) \right) && (20)\\
p_G^{(m+1)} &= p_G^{(m)} \\
&- \mu_2 \Delta\tau \left(\frac{\partial \mathcal{L}[p^{(m)}; St_G^{(I)}](B)}{\partial n} \right. && (21)\\
&+ \left. \kappa(B)\, \mathcal{L}[\rho^{(m)}; St_G^{(II)}](B) \left(\mathcal{L}[u_\tau^{(m)}; St_G^{(II)}](B) \right)^2 \right) && (22)\\
\rho_G^{(m+1)} &= \rho_G^{(m)} - \mu_3 \Delta\tau \left(\frac{\partial \mathcal{L}[\rho^{(m)}; St_G^{(I)}](B)}{\partial n} \right. && (23)\\
&- \left. \frac{1}{c_s^2(B)} \frac{\partial \mathcal{L}[p^{(m)}; St_G^{(II)}](B)}{\partial n} \right) && (24)
\end{aligned}$$

where, in order to simplify the notation, we denoted $c_s^2(B)$ the quantity $\gamma p^{(m)}/\rho^{(m)}$ reconstructed by $\mathcal{L}$ in B. The time step $\Delta\tau$ and the constants μ_i, $i = 1, 2, 3$ are chosen in order to satisfy the CFL conditions for (19), (20), (22), (24), i.e. $\Delta\tau < 1, \quad \mu_i \Delta\tau < h, \quad i = 1, 2, 3$. The iterations scheme (19) - (24) is performed until the residual falls below a fixed tolerance.

Such iterative technique for the boundary conditions is successfully employed in the context of elliptic equations, where it is merged with an iteration scheme for the inner equations and it is embedded in a multigrid framework [5]. However, in this case we do not iterate on inside grid points, since we already have an updated value in such points, while we just want to extrapolate the quantity $q = \rho, u, p$ in ghost points according to the boundary conditions. If we perform the iteration scheme (19) - (24) without a multigrid solver, it results in a slow convergence. For this reason, it is convenient to apply a block-relaxation, since the equations for the ghost points are not always fully coupled, unlike the elliptic case (that involves also the inner equations).

For the special case of a *rigid ball* inside a fluid, the ghost equations are even fully decoupled. In fact, it can be easily proved that the stencils $St_G^{(I,II)}$ involve only ghost points closer to the interface (besides inside points). Then, if we order the ghost points according to the distance from the interface and we perform the iteration from the closer to the farther point in a Gauss-Seidel fashion, we get convergence in one sweep. Without ordering the ghost points, we however obtain the convergence in few iterations. The Gauss-Seidel fashion can be obtained from the iterative scheme (19)-(24) with a suitable choice of the time step $\Delta\tau$ and the constant μ_i.

3.3. **Treatment of moving boundary.** Let us assume that the normal boundary velocity is given by a function $V(\vec{x},t)$. Such function can be easily computed, for example, in the case in which the distance function is known analytically, as is the case of the motion of a rigid body: if $\phi(\vec{x},t)$ is a signed distance function, then one has

$$\frac{\partial\phi}{\partial t} + V\,|\nabla\phi| = 0 \Rightarrow V(\vec{x},t) = -\frac{\partial\phi}{\partial t}.$$

For example, if we have a disk of radius R whose center $\vec{c} \equiv (x_0, y_0)$ moves according to some law $\vec{U_0}(t) = (u_0(t), v_0(t))$, then one has

$$V(\vec{x},t) = \frac{\vec{c}-\vec{x}}{|\vec{c}-\vec{x}|}\cdot\vec{U_0} = \vec{n}\cdot\vec{U_0}. \tag{25}$$

In the case of a moving boundary, the boundary conditions on $\partial\Omega$ for the velocity (i.e. (7) and (11)) become:

$$u_n = V \tag{26}$$

$$\frac{\partial u_\tau}{\partial n} = u_\tau\,\kappa. \tag{27}$$

The condition for the pressure is obtained as follows. The velocity of the fluid on the boundary is given by:

$$\vec{u} = u_\tau\vec{\tau} + V\vec{n}. \tag{28}$$

From the equation of motion, one has:

$$\rho\frac{D\vec{u}}{Dt} + \nabla p = 0.$$

Projecting this relation along the normal to the line, we obtain:

$$-\frac{\partial p}{\partial n} = \rho\frac{D\vec{u}}{Dt}\cdot\vec{n}. \tag{29}$$

Differentiating (28) along the trajectory, taking the scalar product with $\vec{n}$, considering that $\vec{\tau}\cdot\vec{n} = 0$ and $\frac{D\vec{n}}{Dt}\cdot\vec{n} = 0$, from (29) we obtain:

$$-\frac{1}{\rho}\frac{\partial p}{\partial n} = u_\tau\vec{n}\cdot\frac{D\vec{\tau}}{Dt} + \frac{DV}{Dt}$$

Furthermore,

$$\frac{D\vec{\tau}}{Dt} = \frac{\partial\vec{\tau}}{\partial t} + \vec{u}\cdot\nabla\vec{\tau} = \frac{\partial\vec{\tau}}{\partial t} + u_\tau\frac{\partial\vec{\tau}}{\partial\tau} = \frac{\partial\vec{\tau}}{\partial t} + u_\tau\,\kappa\,\vec{n}$$

which gives

$$-\frac{1}{\rho}\frac{\partial p}{\partial n} = \vec{n}\cdot\frac{\partial\vec{\tau}}{\partial t}u_\tau + u_\tau^2\,\kappa + \frac{DV}{Dt}. \tag{30}$$

In Cartesian coordinates, we obtain the expression

$$\vec{n}\cdot\frac{\partial\vec{\tau}}{\partial t} = -\phi_x\phi_{yt} + \phi_y\phi_{xt} \tag{31}$$

where the subscripts denote derivatives. $DV/Dt = \partial V/\partial t + \vec{u}\cdot\nabla V$ should be easily obtained from the analytical expression of V.

In principle, Eq. (30), together with the conditions on the velocity (26) and (27) and the usual adiabatic condition (14), should be sufficient to provide second order boundary conditions.

4. **Numerical test.** We show that, in case of moving boundary, the condition on the pressure (30) is better than (10). The numerical test is taken from [3, Example 4]. We consider a simple rigid movement of a ball with center (x_c, y_c) with respect to the following equations

$$x_c(t) = 0.5, \qquad y_c(0) = 0.5, \qquad \dot{y}_c(t) = 0.01\cos(10\pi t).$$

The initial conditions of the characteristic variables are:

$$(\rho(x,y,0), u(x,y,0), v(x,y,0), p(x,y,0)) = (1,0,0,10).$$

We compute the pressure at time $t = 0.025$. The pressure p is computed in three directions (red, blue and green) through the ball, as illustrated in Fig. 5. The numerical results obtained by the conditions (10) and (30) are illustrated in Fig. 6. We observe that, going inside the ball, the extrapolation obtained by (30) converges more quickly as the grid is refined, than the one obtained by (10). In all figures, star points refer to the test with 25 × 25, dot points refer to the test with 50 × 50, circle points refer to the test with 100 × 100.

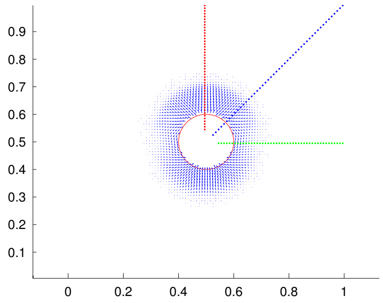


FIGURE 5. Velocity field and the three directions in which we compute p: vertical (red), oblique (blue) and horizontal (green) direction.

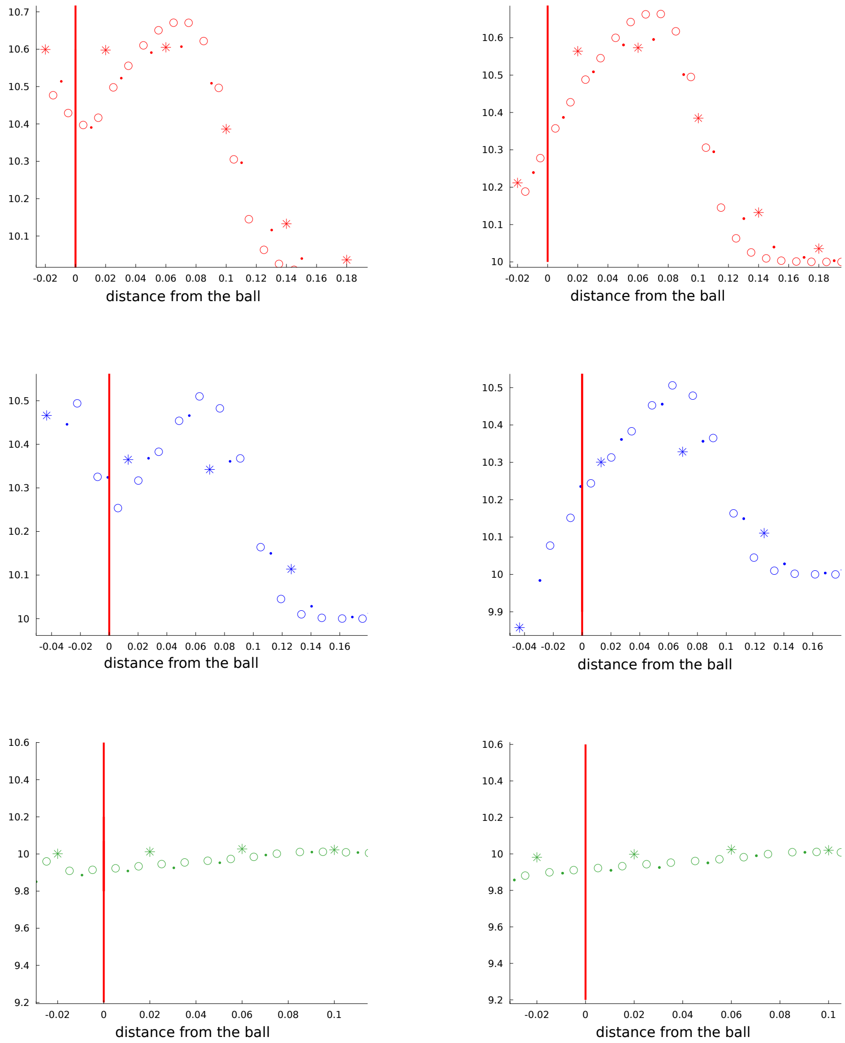


FIGURE 6. p computed in the vertical (red), oblique (blue) and horizontal (green) directions of Fig. 5 using the condition (10) (left) and (30) (right).

REFERENCES

[1] V. CAYOL AND H. C. CORNET, *Effects of topography on the interpretation of the deformation field of prominent volcanoes - Application to Etna*, Geophys. Res. Lett., 25 (1998), pp. 1979–1982.

[2] A. CHERTOCK, A. COCO, A. KURGANOV, AND G. RUSSO, *A Second-Order Finite-Difference Method for Compressible Fluids in Domains with Moving Boundaries*. In preparation.

[3] A. CHERTOCK AND A. KURGANOV, *A simple Eulerian finite-volume method for compressible fluids in domains with moving boundaries*, Commun. Math. Sci., 6 (2008), pp. 531–556.

[4] A. COCO AND G. RUSSO, *Second order multigrid methods for elliptic problems with discontinuous coefficients on an arbitrary interface, I: one dimensional problems*, Numerical Mathematics: Theory, Methods and Applications, 5 (2012), pp. 19–42.

[5] A. Coco and G. Russo, *Finite-Difference Ghost-Point Multigrid Methods on Cartesian Grids for Elliptic Problems in Arbitrary Domains*, Journal of Computational Physics, 241 (2013), pp. 464–501.

[6] A. Dadone and B. Grossman, *Ghost-cell method for analysis of inviscid three-dimensional flows on Cartesian-grids*, Computers and Fluids, 36 (2007), pp. 1513–1528.

[7] C. Shu, *Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws*, in Advanced numerical approximation of nonlinear hyperbolic equations (Cetraro, 1997), vol. 1697 of Lecture Notes in Math., Springer, Berlin, 1998, pp. 325–432.

[8] S. Tan and C.-W. Shu, *A high order moving boundary treatment for compressible inviscid flows*, Journal of Computational Physics, 230 (2011), pp. 6023–6036.

E-mail address: coco@dmi.unict.it
E-mail address: russo@dmi.unict.it

CONSERVATION LAWS IN THE MODELING OF MOVING CROWDS

Rinaldo M. Colombo
INDAM Unit, Brescia University, Brescia, Italy

Mauro Garavello
Department of Mathematics and Applications
Milano – Bicocca University, Milano, Italy

Magali Lécureux-Mercier
Technion, Israel Institute of Technology, Haifa, Israel

Nikolay Pogodaev
Russian Academy of Sciences, Irkutsk, Russia

Abstract. Models for crowd dynamics are presented and compared. Well posedness results allow to exhibit the existence of optimal controls in various situations. A new approach not based on partial differential equations is also briefly considered.

1. **Introduction.** From a macroscopic viewpoint, a moving crowd can be described through its density $\rho = \rho(t, x)$, a function of time $t \in \mathbb{R}^+$ and space $x \in \mathbb{R}^2$ attaining values in $[0, 1]$. In standard situations, the number of pedestrians is conserved, so that $\int_{\mathbb{R}^2} \rho(t, x)\, \mathrm{d}x$ is independent of t. Hence, it is natural to use the *conservation law*

$$\partial_t \rho + \mathrm{div}_x \left(\rho\, V\right) = 0\,. \tag{1}$$

Any model of this kind depends on the *speed law* that defines the velocity V of the crowd as a function of t, x, ρ, ... A simple version of (1) is obtained assigning

$$V = v(\rho)\, \vec v(x) \quad \text{with} \quad \begin{array}{l} v \in \mathbf{C}^2([0,1]; \mathbb{R}^+) \text{ non increasing and } v(1) = 0\,, \\ \vec v \in \mathbf{C}^2(\mathbb{R}^2; \mathbb{S}^1)\,. \end{array} \tag{2}$$

In this case, Kružkov Theorem [24, Theorem 1] applies and ensures that the Cauchy problem for (1)–(2) has a unique solution in $\mathbf{C}^{0,1}\left(\mathbb{R}^+; \mathbf{L}^1(\mathbb{R}^2; [0,1])\right)$ which depends Lipschitz continuously from the data and, by [12, Theorem 2.6], also from v and $\vec v$.

According to (2), at time t the pedestrian at x moves along a prescribed trajectory, an integral curve of $\vec v$, with a speed $v(\rho)$ that depends on ρ evaluated at point x and time t. On the contrary, Section 2 is devoted to (1) with the speed of the individual at x depending on an *average* of the density ρ in a *neighborhood* of x. The resulting model has a rich analytical structure, the solutions being also *differentiable* with respect to the data and to the speed law.

In Section 3 the direction chosen by the pedestrian at x depends from an *average* of the density *gradient* $\nabla \rho$ around x, while his/her speed depends from ρ evaluated at x. The resulting solutions display qualitative properties usually seen in context

Key words and phrases. Conservation laws, nonlocal conservation laws, crowd dynamics.

where individuals have a proper volume such as the *Braess paradox* [3] and the formation of queues [23].

If the various individuals have different destinations then it is possible to subdivide the crowd under consideration into different, say n, populations with densities $\rho_1, \ldots, \rho_n$, each having a different destination. The resulting model

$$\partial_t \rho_i + \operatorname{div}_x (\rho_i \, V_i) = 0 \qquad i = 1, \ldots, n \tag{3}$$

consists of a system of conservation laws that, when $n = 1$, reduces to (1). The results in both Section 2 and Section 3 can be extended to this more general setting.

Finally, Section 4 approaches the problem of driving a crowd with a few moving individuals. First, a model based on (1) is recalled and then an approach based on differential inclusions is presented. The latter approach, developed following [5, 6], neglects the crowd internal dynamics and allows for a simpler analytical framework.

We refer for instance to [2] for an account of the fast development of the recent macroscopic modeling of crowd dynamics. Moreover, measure valued conservation laws were considered in [18, 26]; the results in [25] deal with constrained velocity models; various 1D attempts are found in [1, 15, 16, 20, 21]. Throughout, for the basic results in the theory of conservation laws we refer to [4, 19].

2. **Nonlocal speed choice.**

Consider (1) with the nonlocal speed law

$$V(\rho) = v\left(\rho(t) * \eta\right) \, \vec{v} \, . \tag{4}$$

Here, the speed v at time t of the pedestrian at x depends on the averaged density $\left(\rho(t) * \eta\right)(x) = \int_{\mathbb{R}^2} \rho(t, x-y)\, \eta(y) \, dy$. The direction of the velocity is given by the (fixed) vector $\vec{v}(x)$.

For simplicity, we state the results below in $\mathbb{R}^2$. However, the case where the region available to the crowd is constrained by, say, walls or doors can be easily recovered in the present framework, along the technique used in [7, 8]

As is typical whenever Kružkov techniques apply, space dimension 2 plays no role and the results below can be extended to $\mathbb{R}^n$.

Existence and uniqueness of a solution to the Cauchy problem for (1)–(4) follow from the next result.

Theorem 2.1. [9, Proposition 4.1], [10, Theorem 2.2] *Let* $v \in (\mathbf{C}^2 \cap \mathbf{W}^{2,\infty})(\mathbb{R}; \mathbb{R})$, $\vec{v} \in (\mathbf{C}^2 \cap \mathbf{W}^{2,1})(\mathbb{R}^2; \mathbb{R}^2)$, $\eta \in (\mathbf{C}^2 \cap \mathbf{W}^{2,\infty})(\mathbb{R}^2; \mathbb{R})$. *Assume* $\rho_o \in (\mathbf{L}^1 \cap \mathbf{L}^\infty \cap \mathbf{BV})(\mathbb{R}^2; \mathbb{R}^+)$. *Then,* (1)–(4) *with initial condition* ρ_o *admits a unique weak entropy solution* $\rho \in \mathbf{C}^0\left(\mathbb{R}^+; \mathbf{L}^1(\mathbb{R}^2; \mathbb{R}^+)\right)$. *Furthermore, we have the estimate* $\left\|\rho(t)\right\|_{\mathbf{L}^\infty} \leq \left\|\rho_o\right\|_{\mathbf{L}^\infty} e^{Ct}$, *where the constant* C *depends on* v, $\vec{v}$ *and* η.

The definition of weak entropy solutions is based on Kružkov notion [24, Definition 1], see also [9, 10]. The proof relies on a contraction argument based on the key estimates provided by [12, Theorem 2.6].

Another contraction argument, based on tools from optimal transport theory, allows to extend the above result to the measure valued setting in [17]. (Below, $\mathcal{M}^+(\mathbb{R}^+)$ is the set of positive Radon measures on $\mathbb{R}^2$).

Theorem 2.2. [17, Theorem 1.1] *Assume* $v \in (\mathbf{L}^\infty \cap \mathbf{Lip})(\mathbb{R}; \mathbb{R})$, $\vec{v} \in (\mathbf{L}^\infty \cap \mathbf{Lip})(\mathbb{R}^2; \mathbb{R}^2)$, $\eta \in (\mathbf{L}^\infty \cap \mathbf{Lip})(\mathbb{R}^2; \mathbb{R}^+)$. *Let* $\rho_o \in \mathcal{M}^+(\mathbb{R}^2)$. *Then, there exists a unique weak measure valued solution* $\rho \in \mathbf{L}^\infty(\mathbb{R}^+; \mathcal{M}^+(\mathbb{R}^2))$ *to* (1)–(4) *with initial condition* ρ_o. *If furthermore* $\rho_o \in \mathbf{L}^1(\mathbb{R}^2; \mathbb{R}^+)$, *then* $\rho \in \mathbf{C}^0\left(\mathbb{R}^+; \mathbf{L}^1(\mathbb{R}^2; \mathbb{R}^+)\right)$.

In general, in (1)–(4) no *a priori* uniform $\mathbf{L}^\infty$ bound on the density is possible. Indeed, assume that the density is 1 all along the trajectory of the pedestrian at x. The averaged density around x may well be less than 1, forcing the pedestrian to proceed and, hence, leading to a increase in the density. This behavior can be related to the rise of panic, see [15, 16]. In the literature, values of ρ of up to 10 individuals per square meter were measured, see for instance [22].

Aiming at preventing the insurgence of these phenomena, it is natural to consider control problems where functionals of the density of the type

$$J_T(\rho_o) = \int_0^T \int_\Omega f\left(\rho(t,x)\right) \,\mathrm{d}x\,\mathrm{d}t \quad \text{where } \rho \text{ solves (1)–(4) with datum } \rho_o \tag{5}$$

have to be minimized. Here, Ω is the region where the density needs to be controlled and f is a $\mathbf{C}^1$ function weighing 0 on acceptable densities and quickly increasing when ρ approaches dangerous values. Necessary conditions for the minima of (5) are available once the differentiability of the solution to (1)–(4) with respect to the initial datum is proved. This motivates the following result.

Theorem 2.3. [9, Theorem 4.2] [10, Theorem 2.2] *Let $\rho_o \in (\mathbf{W}^{2,\infty}\cap\mathbf{W}^{2,1})(\mathbb{R}^2;\mathbb{R}^+)$, $r_o \in (\mathbf{W}^{1,1}\cap\mathbf{L}^\infty)(\mathbb{R}^2;\mathbb{R})$. Assume $v \in (\mathbf{C}^4\cap\mathbf{W}^{2,\infty})(\mathbb{R};\mathbb{R})$, $\vec\nu \in (\mathbf{C}^3\cap\mathbf{W}^{2,1})(\mathbb{R}^2;\mathbb{R}^2)$, $\eta \in (\mathbf{C}^3 \cap \mathbf{W}^{2,\infty})(\mathbb{R}^2;\mathbb{R}^+)$. Then, there exists a unique weak entropy solution $r \in \mathbf{C}^0\left(\mathbb{R}^+;\mathbf{L}^1\left(\mathbb{R}^2;\mathbb{R}\right)\right)$ to the Cauchy problem*

$$\partial_t r + \operatorname{div}\left(r\, v(\rho*\eta)\,\vec\nu(x)\right) = -\operatorname{div}\left(\rho\, v'(\rho*\eta)\,\vec\nu(x)\right), \qquad r(0) = r_o\,. \tag{6}$$

Furthermore, for all $\rho_o \in (\mathbf{W}^{2,1}\cap\mathbf{W}^{2,\infty})(\mathbb{R}^2;\mathbb{R}^+)$ and $r_o \in (\mathbf{W}^{1,1}\cap\mathbf{L}^\infty)(\mathbb{R}^2;\mathbb{R})$, call ρ_h the solution to (1)–(4) with initial datum $\rho_o + hr_o$. Then, for all $t\in\mathbb{R}^+$,

$$\lim_{h\to 0}\left\| \frac{\rho_h(t)-\rho(t)}{h} - r(t)\right\|_{\mathbf{L}^1} = 0 \tag{7}$$

i.e., the solution ρ to (1)–(4) is Gâteaux differentiable in ρ_o along any direction r_o.

To prove this theorem, first the well posedness of (6) is obtained and then the limit (7) is computed. In both steps, the estimates in [12] play a key role. At present, no analog to Theorem 2.3 is available in the setting of Theorem 2.2. Indeed, a good definition of Gâteaux differentiability on the set of probability measures equipped with the Wasserstein distance of order 1 is, to our knowledge, not available.

3. Nonlocal Route choice.

Consider (1) with the nonlocal speed law

$$V(\rho) = v(\rho)\,\left(\vec\nu(x) + \mathcal{I}(\rho)\right)\,. \tag{8}$$

Here, the individual in x at time t moves at the speed $v\left(\rho(t,x)\right)$ that depends on the density $\rho(t,x)$ evaluated at the same time t and x. The vector $\vec\nu(x)\in\mathbb{R}^2$ is the preferred direction of the pedestrian at x, while $\mathcal{I}(\rho)(x)$ describes how the pedestrian at x deviates from the preferred direction, given that the crowd distribution is ρ. Thus, the individual at time t in x is assumed to move in the direction of the vector $\vec\nu(x) + \Big(\mathcal{I}\left(\rho(t)\right)\Big)(x)$. The basic well posedness result for (1)–(8) is the following.

Theorem 3.1. [8, Theorem 2.1, Theorem 2.2] *Let the following conditions hold:*

- **(v):** *$v \in \mathbf{C}^2(\mathbb{R};\mathbb{R})$ is non increasing, $v(0) = V$ and $v(R) = 0$ for fixed $V, R > 0$.*
- **($\vec\nu$):** *$\vec\nu \in (\mathbf{C}^2\cap\mathbf{W}^{1,\infty})(\mathbb{R}^2;\mathbb{R}^2)$ is such that $\operatorname{div}\vec\nu \in (\mathbf{W}^{1,1}\cap\mathbf{W}^{1,\infty})(\mathbb{R}^2;\mathbb{R})$.*
- **(I):** *$\mathcal{I} \in \mathbf{C}^0\left(\mathbf{L}^1(\mathbb{R}^2;[0,R]);\mathbf{C}^2(\mathbb{R}^2;\mathbb{R}^2)\right)$ satisfies the estimates:*
  - **(I.1)** *There exists an increasing $C_I \in \mathbf{L}^\infty_{loc}(\mathbb{R}^+;\mathbb{R}^+)$ such that, for all $r \in \mathbf{L}^1(\mathbb{R}^2;[0,R])$, $\left\|\mathcal{I}(r)\right\|_{\mathbf{W}^{1,\infty}} \le C_I(\|r\|_{\mathbf{L}^1})$ and $\left\|\operatorname{div}\mathcal{I}(r)\right\|_{\mathbf{L}^1} \le C_I(\|r\|_{\mathbf{L}^1})$.*

(I.2) *There exists an increasing* $C_I \in \mathbf{L}^\infty_{loc}(\mathbb{R}^+;\mathbb{R}^+)$ *such that, for all* $r \in \mathbf{L}^1(\mathbb{R}^2;[0,R])$, $\left\|\nabla \operatorname{div} \mathcal{I}(r)\right\|_{\mathbf{L}^1} \leq C_I(\|r\|_{\mathbf{L}^1})$.

(I.3) *There exists a constant* K_I *such that for all* $r_1, r_2 \in \mathbf{L}^1(\mathbb{R}^2;[0,R])$,

$$\begin{aligned} \left\|\mathcal{I}(r_1) - \mathcal{I}(r_2)\right\|_{\mathbf{L}^\infty} &\leq K_I \cdot \|r_1 - r_2\|_{\mathbf{L}^1}\,, \\ \left\|\mathcal{I}(r_1) - \mathcal{I}(r_2)\right\|_{\mathbf{L}^1} + \left\|\operatorname{div}\left(\mathcal{I}(r_1) - \mathcal{I}(r_2)\right)\right\|_{\mathbf{L}^1} &\leq K_I \cdot \|r_1 - r_2\|_{\mathbf{L}^1}\,. \end{aligned}$$

Choose any $\rho_o \in (\mathbf{L}^1 \cap \mathbf{BV})(\mathbb{R}^2;[0,R])$. *Then, there exists a unique weak entropy solution* $\rho \in \mathbf{C}^0\left(\mathbb{R}^+;\mathbf{L}^1(\mathbb{R}^2;[0,R])\right)$ *to* (1)–(8). *Moreover,* ρ *satisfies the bounds*

$$\begin{aligned} \left\|\rho(t)\right\|_{\mathbf{L}^1} &= \|\rho_o\|_{\mathbf{L}^1}\,, \text{ for a.e. } t \in \mathbb{R}^+\,, \\ \operatorname{TV}(\rho(t)) &\leq \operatorname{TV}(\rho_o)\, e^{kt} + \frac{\pi}{4} t e^{kt} N \|q\|_{\mathbf{L}^\infty([0,R])} \left(\|\nabla \operatorname{div} \vec{\nu}\|_{\mathbf{L}^1} + C_I(\|\rho_o\|_{\mathbf{L}^1})\right)\,, \end{aligned}$$

where $k = (2N+1)\left\|q'\right\|_{\mathbf{L}^\infty([0,R])} \left(\|\nabla\vec{\nu}\|_{\mathbf{L}^\infty} + C_I(\|\rho_o\|_{\mathbf{L}^1})\right)$. *If also the speed law*

$$V'(\rho) = v'(\rho)\,\left(\vec{\nu}'(x) + \mathcal{I}'(\rho)\right) \tag{9}$$

satisfies the same assumptions, then the solution ρ *to* (1)–(8) *and* ρ' *to* (1)–(9), *with data* $\rho_o, \rho_o' \in (\mathbf{L}^1 \cap \mathbf{BV})(\mathbb{R}^2;[0,R])$, *satisfy*

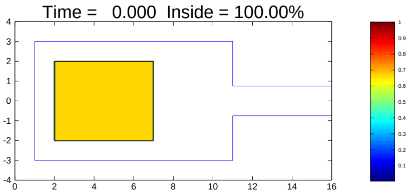
$$\begin{aligned} \left\|\rho_1(t) - \rho_2(t)\right\|_{\mathbf{L}^1} \leq\; & (1 + C(t))\left\|\rho_{0,1} - \rho_{0,2}\right\|_{\mathbf{L}^1} + C(t)\left(\|q_1 - q_2\|_{\mathbf{W}^{1,\infty}} + d(\mathcal{I}_1, \mathcal{I}_2)\right) \\ & + C(t)\left(\|\vec{\nu}_1 - \vec{\nu}_2\|_{\mathbf{L}^\infty} + \left\|\operatorname{div}(\vec{\nu}_1 - \vec{\nu}_2)\right\|_{\mathbf{L}^1}\right) \end{aligned}$$

where

$$d(\mathcal{I}_1,\mathcal{I}_2) = \sup\left\{\left\|\mathcal{I}_1(\rho) - \mathcal{I}_2(\rho)\right\|_{\mathbf{L}^\infty} + \left\|\operatorname{div}\left(\mathcal{I}_1(\rho) - \mathcal{I}_2(\rho)\right)\right\|_{\mathbf{L}^1} : \rho \in \mathbf{L}^1(\mathbb{R}^2;[0,R])\right\}.$$

The map $C \in \mathbf{C}^0(\mathbb{R}^+;\mathbb{R}^+)$ *vanishes at* $t = 0$ *and depends on* $\operatorname{TV}(\rho_{0,1})$, $\left\|\rho_{0,1}\right\|_{\mathbf{L}^1}$, $\|\vec{\nu}_1\|_{\mathbf{L}^\infty}$, $\|\operatorname{div}\vec{\nu}_1\|_{\mathbf{W}^{1,1}}$, $\|q_1\|_{\mathbf{W}^{1,\infty}}$, $\|q_2\|_{\mathbf{W}^{1,\infty}}$.

In operation research, Braess paradox states that adding extra capacity to a network can, in some cases, reduce the overall performance of the network, see [3]. A relevant problem in the design of escape routes is the planning of suitable devices that reduce the exit time. The model (1)–(8) allows to show that the careful introduction of suitable obstacles in suitable locations does indeed reduce the exit time. In fact, these obstacles reduce congested areas at the sides of the door jambs.



$$\begin{aligned} v(\rho) &= 6(1-\rho)\,, \\ \eta(x) &= \left[1 - \left(\tfrac{x_1}{r}\right)^2\right]^3 \left[1 - \left(\tfrac{x_2}{r}\right)^2\right]^3 \\ &\quad \times \chi_{[-r,r]^2}(x), \\ \rho_o(x) &= 0.75\,\chi_{[2,7]\times[-2,2]}(x)\,, \\ r &= 0.6\,, \qquad \varepsilon = 0.4\,. \end{aligned} \tag{10}$$

FIGURE 1. Initial datum and room geometry, without obstacles.

We consider a room with an exit, as in Figure 1. The vector $\vec{\nu} = \vec{\nu}(x)$ is the unit vector tangent at x to the geodesic connecting x to the exit and $\mathcal{I}(\rho) = -\varepsilon\,\left(\nabla(\rho * \eta)\right) \Big/ \sqrt{1 + \left\|\nabla(\rho * \eta)\right\|^2}$, see (10).

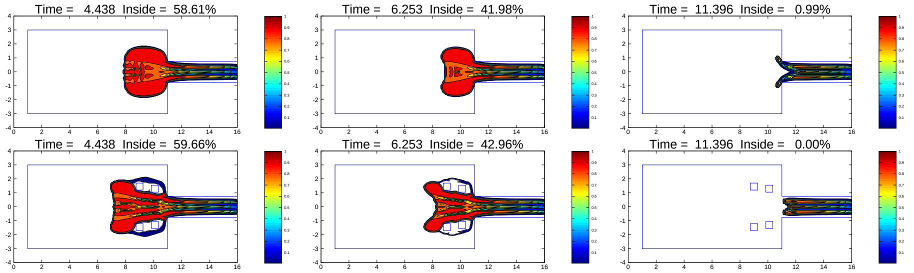


FIGURE 2. Solution to (1)–(8)–(10) with $\varepsilon = 0.2$, at times $t = 4.438, 6.253, 11.396$. On the first line, no obstacle is present. On the second line, 4 columns direct the crowd flow. The exit time in the latter case is *shorter* than in the former one, see [7].

The careful positioning of obstacles as in the second line of Figure 2 diminishes the size of the congested region and, with the chosen initial datum, gives an exit time lower than that with the room free from any obstacle, see Figure 2.

4. **Individuals driving a population.** We finally introduce a model describing the situation in which a discrete set of isolated individuals interacts with a continuum crowd. Examples can be a (group of) predator(s) running after their preys, shepherd dogs driving a herd of sheep, or a leader attracting a group of followers to a given region. Let $\rho \in \mathbb{R}^+$ be the population density and $p \equiv (p_1, \ldots, p_k) \in \mathbb{R}^{2k}$ be the positions of the k individuals. Following [11], the interaction is described by

$$\left\{ \begin{array}{l} \partial_t \rho + \operatorname{div}\left(\rho\, V\big(t, x, \rho(t,x), p(t)\big)\right) = 0\,, \\ \dot p = \varphi\left(t, p(t), \rho(t)\right) . \end{array} \right. \tag{11}$$

Here, φ is typically nonlocal, meaning that the individuals p react to averages of quantities depending on ρ. The well posedness of (11) is proved in [11, Theorem 2.2], by means of Kružkov theory, the estimates in [12] and tools from the stability of ordinary differential equations.

As a first illustrating example, assume that the vector $p \in \mathbb{R}^2$ is the position of a leader (e.g. a magic piper) and ρ is the density of the followers (e.g. rats). We are thus lead to consider (11) with

$$\begin{array}{rcl} V(t,x,\rho,p) & = & v(\rho)\,(p - x)\, e^{-\|p-x\|} \\ \varphi(t,p,\rho) & = & \big(1 + (\rho * \eta)(p)\big)\; \vec\psi(t)\,. \end{array} \tag{12}$$

The function v essentially describes the speed of the followers and is, as usual, a smooth decreasing function vanishing at, say, $\rho = 1$. The follower located at x moves along $p(t) - x$ toward the leader, with a speed exponentially decreasing with the distance $\|p - x\|$ between leader and follower. The speed of the leader increases with the averaged density $\rho * \eta$, computed at the leader's position. Indeed, we expect the leader to wait for the followers to join him when the followers' density around him is small. The direction $\vec\psi$ of the leader is chosen *a priori*. See Figure 3 for a numerical integration of (11)–(12) and [11] for further details.

As a further example, consider n shepherd dogs, located in $p_i(t) \in \mathbb{R}^2$ for $i \in \{1, \ldots, n\}$ and a group of sheep of density ρ. The dogs have to confine the sheep

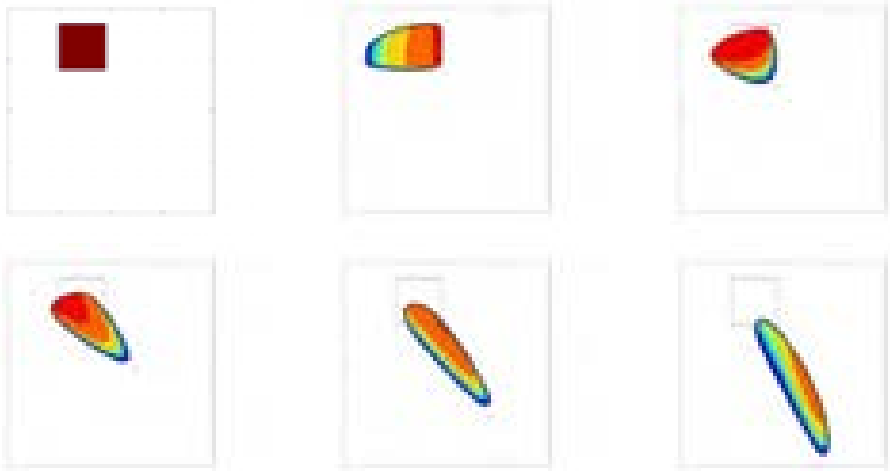


FIGURE 3. Solution of (11)–(12) in the square $[-1,1] \times [-1,1]$ at times 0.000, 0.171, 0.543, 0.945, 1.447 and 1.930, from [11].

within a given area. We are thus lead to consider (11) with

$$\begin{array}{rcl} V(t,x,\rho,p) & = & v(\rho)\,\vec{\nu}(x) + \sum_{i=1}^{n} (x-p_i) e^{-\|p_i - x\|} \\ \varphi(t,p,\rho) & = & \frac{(\rho * \nabla \eta)^{\perp}(p_i)}{\sqrt{1+\left\|\rho*\nabla\eta(p_i)\right\|^2}}\,, \qquad \text{for } i \in \{1, \ldots n\}\,. \end{array} \tag{13}$$

As above, the speed of the sheep is given by the decreasing function $v(\rho)$ that vanishes in $\rho = 1$. The direction of a sheep located at x is a sum of two terms. The first one is the sheep's preferred direction $\vec{\nu}(x)$; the second one is the vector $\sum_{i=1}^{n} (x - p_i(t)) e^{-\|p_i - x\|}$ representing the repulsive effects of the dogs on the sheep. Each dog runs around the flock along the direction perpendicular to the gradient of the sheep average density.

5. **A different approach.** Following [6, 14, 13], we present another framework to describe the population–individuals interactions. Initially, the population occupies the compact set $K_o \subset \mathbb{R}^2$. If there are no individuals, the member at x of the population is free to wander in $\mathbb{R}^2$, according to the differential inclusion

$$\dot{x} \in B(0,c)\,, \qquad x(0) \in K_o\,, \tag{14}$$

c being the maximal wandering speed and $B(0,c)$ the closed ball in $\mathbb{R}^2$ centered at 0 with radius c. Hence, the population fills the reachable set of (14). Introduce now n individuals sited at $\xi \equiv (\xi_1, \xi_2, \ldots, \xi_n) \in \mathbb{R}^{2n}$. Then, the interaction between the individuals and each population member leads to the modified differential inclusion

$$\dot{x} \in v\left(x, \xi(t)\right) + B(0,c), \quad x(0) \in K_o, \tag{15}$$

where the vector field $v \in \mathbf{C}^{0,1}(\mathbb{R}^2 \times \mathbb{R}^{2n}; \mathbb{R}^2)$ is the drift speed due to the attractive or repulsive effect that each agent has on each member of the population. Thus, given the individuals' trajectory $\xi \in \mathbf{C}^{0,1}_{loc}(\mathbb{R}^+; \mathbb{R}^{2n})$, the reachable set $\mathcal{R}_\xi(K_o, t)$ of (15) at time t is the set occupied by the population at time t under the effect of the agents. With the present assumptions, $\mathcal{R}_\xi(K_o, t)$ is non-empty and compact.

If only one agent is present $(n = 1)$ and v is spherically symmetric, i.e.,

$$v(x,\xi) = \psi(|x-\xi|)(x-\xi) \quad \text{for a suitable } \psi\colon \mathbb{R} \to \mathbb{R}. \tag{16}$$

the next result exhibits a trajectory ξ confining the population in a given set K.

Theorem 5.1. [14, Theorem 2.8] *Let $c > 0$. Fix a bounded $\psi \in \mathbf{C}^{1,1}_{loc}(\mathbb{R}^n;\mathbb{R})$ and define v as in (16). Assume that there exist positive R_*^-, R_*^+ and R such that*

$$\frac{1}{\pi}\int_0^\pi \psi\left(\sqrt{R^2+R_*^2-2R_*\,R\cos\theta}\right)(R_*-R\cos\theta)\,\mathrm{d}\theta < -c$$

for all $R_* \in [R_*^-, R_*^+]$. Then, there exists a $\xi \in \mathbf{C}^{0,1}_{loc}\left(\mathbb{R}^+;\partial B(0,R)\right)$ such that, calling K_o the region initially occupied by the population,

$$\text{if} \quad K_o \subseteq B(0,R_*^-) \quad \text{then} \quad \mathcal{R}_\xi(t,K_o) \subseteq B(0,R_*^+) \quad \text{for all} \quad t \geq 0\,.$$

Note that the confining strategy $t \mapsto \xi(t)$ above is constructed explicitly, see [13, Theorem 2.5]. A negative result is also available. Before stating it, recall that for a measurable function $\varphi\colon \mathbb{R}^+ \to \mathbb{R}$, its *non–decreasing rearrangement* is the function $\varphi_*\colon \mathbb{R}^+ \to \mathbb{R}$, which is non–decreasing and satisfies $\mathcal{L}^1\left(\varphi_*^{-1}(]-\infty,a])\right) = \mathcal{L}^1\left(\varphi^{-1}(]-\infty,a])\right)$ for all $a \in \mathbb{R}$.

Theorem 5.2. [14, Theorem 2.7] *Let $c > 0$. Fix a bounded $\psi \in \mathbf{C}^{1,1}_{loc}(\mathbb{R};\mathbb{R})$ and define v as in (16). Let φ_* be the non–decreasing rearrangement of the function*

$$\varphi(s) = \psi'\left(\sqrt[2]{\frac{s}{\pi}}\right)\sqrt[2]{\frac{s}{\pi}} + 2\,\psi\left(\sqrt[2]{\frac{s}{\pi}}\right).$$

If the initial set K_o is such that

$$2\,c\sqrt{\pi\sigma} + \int_0^\sigma \varphi_*(s)\,\mathrm{d}s > 0 \qquad \text{for all} \qquad \sigma \geq \mathcal{L}^2(K_o)$$

then, for every $\xi \in \mathbf{C}^{0,1}_{loc}(\mathbb{R}^+;\mathbb{R}^{kn})$, the measure $\mathcal{L}^2\left(\mathcal{R}_\xi(t,K_o)\right)$ of the reachable set $\mathcal{R}_\xi(t,K_o)$ of (15) increases unboundedly in time, so that no confinement is possible.

We refer to the cited references for the statement of these results in arbitrary space dimension. Theorem 5.2 holds also in the case of several individuals, each acting as in (16) (see [14]).

REFERENCES

[1] D. Amadori and M. Di Francesco. The one-dimensional Hughes model for pedestrian flow: Riemann–type solutions. *Acta Mathematica Scientia*, 32(1):259–196, 2011.

[2] N. Bellomo and C. Dogbe. On the modeling of traffic and crowds: a survey of models, speculations, and perspectives. *SIAM Rev.*, 53(3):409–463, 2011.

[3] D. Braess. Über ein Paradoxon aus der Verkehrsplanung. *Unternehmensforschung*, 12:258–268, 1968.

[4] A. Bressan. *Hyperbolic systems of conservation laws*, volume 20 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2000. The one-dimensional Cauchy problem.

[5] A. Bressan. Differential inclusions and the control of forest fires. *J. Differential Equations*, 243(2):179–207, 2007.

[6] A. Bressan and D. Zhang. Control problems for a class of set valued evolutions. *Set-Valued and Variational Analysis*, pages 1–21, 2012. 10.1007/s11228-012-0204-5.

[7] R. M. Colombo, M. Garavello, and M. Lécureux-Mercier. Non-local crowd dynamics. *C. R. Math. Acad. Sci. Paris*, 349(13-14):769–772, 2011.

[8] R. M. Colombo, M. Garavello, and M. Lécureux-Mercier. A class of non-local models for pedestrian traffic. *Mathematical Models and Methods in the Applied Sciences*, 22(4), 2012.

[9] R. M. Colombo, M. Herty, and M. Mercier. Control of the continuity equation with a non local flow. *ESAIM Control Optim. Calc. Var.*, 17(2):353–379, 2011.

[10] R. M. Colombo and M. Lécureux-Mercier. Nonlocal crowd dynamics models for several populations. *Acta Mathematica Scientia*, 32(1):177–196, 2011.

[11] R. M. Colombo and M. Mercier. An analytical framework to describe the interactions between individuals and a continuum. *Journal of Nonlinear Science*, 22(1):39–61, 2012.

[12] R. M. Colombo, M. Mercier, and M. D. Rosini. Stability and total variation estimates on general scalar balance laws. *Commun. Math. Sci.*, 7(1):37–65, 2009.

[13] R. M. Colombo and N. Pogodaev. Confinement strategies in a model for the interaction between individuals and a continuum. *SIAM J. Appl. Dyn. Syst.*, 11(2):741–770, 2012.

[14] R. M. Colombo and N. Pogodaev. On the control of moving sets: Positive and negative confinement results. SIAM Journal on Control and Optimization, 51, 1, 380-401, 2013

[15] R. M. Colombo and M. D. Rosini. Pedestrian flows and non-classical shocks. *Math. Methods Appl. Sci.*, 28(13):1553–1567, 2005.

[16] R. M. Colombo and M. D. Rosini. Existence of nonclassical solutions in a pedestrian flow model. *Nonlinear Anal. Real World Appl.*, 10(5):2716–2728, 2009.

[17] G. Crippa and M. Lécureux-Mercier. Existence and uniqueness of measure solutions for a system of continuity equations with non-local flow. *NODEA*, 2012.

[18] E. Cristiani, B. Piccoli, and A. Tosin. Multiscale modeling of granular flows with application to crowd dynamics. *Multiscale Model. Simul.*, 9(1):155–182, 2011.

[19] C. M. Dafermos. *Hyperbolic conservation laws in continuum physics*, volume 325 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 2010.

[20] M. Di Francesco, P. A. Markowich, J.-F. Pietschmann, and M.-T. Wolfram. On the Hughes' model for pedestrian flow: the one-dimensional case. *J. Differential Equations*, 250(3):1334–1362, 2011.

[21] N. El-Khatib, P. Goatin, and M. D. Rosini. On entropy weak solutions of Hughes model for pedestrian motion. *ZAMP*, 64 (2013), no. 2, 223251.

[22] D. Helbing, A. Johansson, and H. Z. Al-Abideen. Dynamics of crowd disasters: An empirical study. *Phys. Rev. E*, 75:046109, Apr 2007.

[23] D. Helbing, P. Molnár, I. Farkas, and K. Bolay. Self-organizing pedestrian movement. *Environment and Planning B: Planning and Design*, 28:361–383, 2001.

[24] S. N. Kružhkov. First order quasilinear equations with several independent variables. *Mat. Sb. (N.S.)*, 81 (123):228–255, 1970.

[25] B. Maury, A. Roudneff-Chupin, and F. Santambrogio. A macroscopic crowd motion model of gradient flow type. *Math. Models Methods Appl. Sci.*, 20(10):1787–1821, 2010.

[26] B. Piccoli and A. Tosin. Time-evolving measures and macroscopic modeling of pedestrian flow. *Arch. Ration. Mech. Anal.*, 199(3):707–738, 2011.

E-mail address: rinaldo@ing.unibs.it
E-mail address: mauro.garavello@unimib.it
E-mail address: magali.lecureux-mercier@univ-orleans.fr
E-mail address: nickpogo@gmail.com

A HYPERBOLIC MODEL FOR PHASE TRANSITIONS IN POROUS MEDIA

Andrea Corli

Department of Mathematics and Computer Science
University of Ferrara
44121 Ferrara, Italy

Haitao Fan

Department of Mathematics
Georgetown University
Washington, DC 20057, USA

Abstract. We introduce a simple hyperbolic system of three equations, in one space dimension, which models a fluid flow through a porous medium. The model also allows for phase transitions of the fluid: both liquid and vapor phases may be present as well as mixtures of them. At last, an equilibrium pressure makes metastable states possible. The porous medium is modeled by a damping term that depends on the phase and is linear in the velocity.

First, we give some necessary conditions in order that two end-states can be joined by a traveling wave. Then, we provide several sufficient conditions which yield the existence and uniqueness of traveling waves in many different situations. We also verify some structural properties of the model, which imply the global existence of smooth solutions for states close to the stable-liquid or stable-vapor regions.

1. **Introduction.** We propose in this paper a model for dynamic liquid-vapor phase transitions that occur in the fluid flow through a porous medium. We make the assumptions that the fluid is inviscid and that the flow may be assumed to be isothermal. In Lagrangian coordinates, the system is written as

$$\begin{cases} v_t - u_x & = 0\,, \\ u_t + p(v,\lambda)_x & = -\alpha(\lambda)u\,, \\ \lambda_t & = \frac{1}{\tau}\,\left(p(v,\lambda) - p_e\right)\lambda(\lambda-1)\,, \end{cases} \tag{1}$$

for $t > 0$ and $x \in \mathbb{R}$. The notation is standard: $v > 0$ denotes the specific volume, u the velocity, p the pressure, $\lambda \in [0,1]$ the mass-density fraction of the vapor in the fluid. We write for short $w = (v,u,\lambda)$. The flow occurs through a porous medium and the friction force exerted by the medium is assumed to be proportional to the linear momentum, with opposite direction; we emphasize that the proportionality factor $\alpha(\lambda)$ may depend, in a smooth way, on the phase (or the mixtures of phases) of the fluid. The model is completed by a reaction term in the equation for the evolution of the phases: there, the constants $\tau > 0$ and $p_e > 0$ denote a characteristic reaction time and an equilibrium pressure, respectively.

2000 *Mathematics Subject Classification.* Primary: 35C07, 76S05, 82B26; Secondary: 35L65.
Key words and phrases. Traveling waves, Phase transitions, Porous media.

In order to close the model, we assume that the pressure $p(v,\lambda)$ is a given smooth function satisfying the following physical requirements:

$$p_v < 0, \quad p_\lambda > 0, \quad p_{vv} > 0, \quad p_{v\lambda} < 0. \tag{2}$$

As a consequence of the assumption $p_v < 0$, the homogeneous part of 1 is hyperbolic with eigenvalues $\pm\sqrt{-p_v}$ and 0. A simple example of a pressure law satisfying 2 is

$$p = \frac{\kappa_0 + \lambda(\kappa_1 - \kappa_0)}{v^\gamma},$$

for $\gamma \geq 1$ and $\kappa_0 < \kappa_1$.

The system 1 is a particular case of a much more complete model proposed in [10], which we adapted here to the case of a flow in a porous medium. About the homogeneous system, we refer to [10] for traveling waves, to [1] for the Riemann problem and to [4] for many structural properties on the stability of large waves. The existence of global weak solutions to the initial-value problem, with initial data of bounded variation, was first proved in [2] and then in [5] in a different way. Similar results were proved in [3] for the nonhomogeneous system in the undamped case $\alpha = 0$; in that paper, the relaxation limit $\tau \to 0$ was also studied for states close either to $\lambda = 0$ or $\lambda = 1$. Related results about the 2×2 Riemann problem with the constraint $(p - p_e)\lambda(\lambda - 1) = 0$ can be found in [6]. On the other hand, the purely damped system, where the third equation reduces to $\lambda_t = 0$, was considered in [13]; in that case, the damping effect of the source term allowed to prove the existence of (small) smooth global solutions. We mention that if the state variable λ is missing, then the related 2×2 system was studied for instance in [15, 14, 12, 9].

In this note, we provide some results about the existence of traveling waves for 1 in the case that $\alpha(\lambda)$ is bounded away from zero. First, we state some necessary conditions on the end states; then, we provide many sufficient conditions for the existence of such solutions. The sufficient conditions are concerned with the possible mutual intersections of three curves: the equilibrium curve $\mathcal{G}$, the sound-speed curve $\mathcal{S}$ and the equilibrium-pressure curve $\mathcal{P}$ where p equals p_e. We emphasize that, even if our analysis does not cover all possible positions of such curves, nevertheless the method of proof can be used to study specific situations as well. Full proofs can be found in [8] in the case that $\alpha \neq 0$ is constant; they easily adapt to the current case $\alpha(\lambda) > 0$. The case when $\alpha(\lambda)$ is allowed to vanish, for instance if $\alpha(1) = 0$, requires a more detailed analysis; we refer to [7] about this latter case. In the last section we check that system 1 satisfies both the Shizuta-Kawashima and the strict entropy-dissipation conditions on the stable regions of pure liquid and pure vapor. Then, by a result in [11], we deduce the existence of global smooth solutions to the initial-value problem when the initial datum is close to one of these regions.

2. **Traveling waves: existence and uniqueness.**

A traveling wave is a solution to 1 of the form $w(\xi) = w\left(\frac{x-ct}{\tau}\right)$ satisfying $w(\pm\infty) = w_\pm$ and $w'(\pm\infty) = 0$. Here c is the constant speed of propagation of the wave and we may assume that $c \geq 0$. This means that w must be a solution of

$$\begin{cases} -cv' - u' = 0\,, \\ -cu' + p' = -Au\,, \\ -c\lambda' = (p - p_e)\lambda(\lambda - 1)\,, \end{cases} \tag{3}$$

and must satisfy the conditions

$$\left\{\begin{array}{l}(v,u,\lambda)(\pm\infty)=(v_\pm,u_\pm,\lambda_\pm)\,,\\(v',u',\lambda')(\pm\infty)=0\,,\end{array}\right. \tag{4}$$

for $(v_\pm,u_\pm,\lambda_\pm)\in(0,+\infty)\times\mathbb{R}\times[0,1]$. We assume that $\alpha(\lambda)$ is a smooth function satisfying $\alpha(\lambda)>0$ for every $\lambda\in[0,1]$; we denoted $A=A(\lambda)=\alpha(\lambda)\tau$.

The end states are equilibrium points of 3 and then must satisfy either $u_\pm=0$ or $(p_\pm-p_e)\lambda_\pm(\lambda_\pm-1)=0$. By considering the first equation in 3 we find another jump condition, namely $c(v_+-v_-)=0$. We discard the stationary case $c=0$ for brevity (see however [8]) and then focus on the remaining case

$$c>0,\qquad v_-=v_+=:\bar v. \tag{5}$$

We substitute the second equation of 3 into the first one; then, by denoting

$$s(v,\lambda):=c^2+p_v,\qquad g(v,\lambda):=Ac(v-\bar v)+\frac{1}{c}p_\lambda(p-p_e)\lambda(\lambda-1),$$

we re-write 3 as

$$\left\{\begin{array}{l}sv'=g,\\\lambda'=-\frac{1}{c}(p-p_e)\lambda(\lambda-1),\end{array}\right. \tag{6}$$

with the end states, deduced by 4,

$$\left\{\begin{array}{l}(v,\lambda)(\pm\infty)=(\bar v,\lambda_\pm),\\(v',\lambda')(\pm\infty)=0.\end{array}\right. \tag{7}$$

The equilibrium points of 6 are $(\bar v,0)$, $(\bar v,1)$ and the points $(\bar v,\bar\lambda)$ satisfying $p(\bar v,\bar\lambda)=p_e$; we point out that, by 2, for each $\bar v$ there is at most one $\bar\lambda$ satisfying the latter equation. At last, we define the curves

$$\mathcal{S}\colon s(v,\lambda)=0,\qquad \mathcal{G}\colon g(v,\lambda)=0,\qquad \mathcal{P}\colon p(v,\lambda)=p_e.$$

We notice that the whole curve $\mathcal{S}$ is singular for system 6. We denote $p_\pm=p(\bar v,\lambda_\pm)$ and $s_\pm=s(\bar v,\lambda_\pm)$; in the following, we assume for simplicity that

$$s(\bar v,\lambda_\pm)\neq 0. \tag{8}$$

We first give some necessary conditions on the end states in order that 1 has traveling wave solutions. These conditions are a consequence of the presence of the equilibrium pressure p_e: the stable states for the dynamical system 6 are those with $p>p_e$ and $\lambda=0$ or those with $p<p_e$ and $\lambda=1$. Conversely, the states with $p>p_e$ and $\lambda=1$ or those with $p<p_e$ and $\lambda=0$ are unstable. The latter states are also called metastable.

Lemma 2.1. *In order that 6-7 has solutions, the end states $(\bar v,\lambda_\pm)$ must satisfy one of the following conditions,*

(i) $p_\pm>p_e$ and either $\lambda_-=0$ or $\lambda_+=1$;
(ii) $p_\pm<p_e$ and either $\lambda_-=1$ or $\lambda_+=0$;
(iii) $p_+>p_e=p_-$ and either $\lambda_-<1$ or $\lambda_+=1$;
(iv) $p_+<p_e=p_-$ and either $\lambda_->0$ or $\lambda_+=0$;
(v) $p_-=p_+=p_e$ and $\lambda_-=\lambda_+$.

No other case may occur.

The proof of the lemma follows by considering the directions of the field lines of the dynamical system 6. To give an idea of the argument, we show that traveling waves cannot exist if $p_-<p_e<p_+$. Assume by contradiction that a trajectory exists. Then, the condition $p_\lambda>0$ in 2 implies $\lambda_-=0$ and $\lambda_+=1$; therefore, the trajectory $\lambda(\xi)$ would be decreasing in a neighborhood of the end point $(\bar{v}, 0)$, because of the second equation in 6. This clearly is a contradiction.

The following lemma, where we use the numbering introduced in Lemma 2.1, provides some general information about the curves $\mathcal{S}$, $\mathcal{G}$ and $\mathcal{P}$. In particular, we study the mutual positions of the curves $\mathcal{G}$ and $\mathcal{P}$, where both right-hand sides of system 6 vanish. We refer to Figure 1 where, as in the following, we plot graphs in the (λ, v)-plane while we keep the reverse ordering (v, λ) for the arguments of functions.

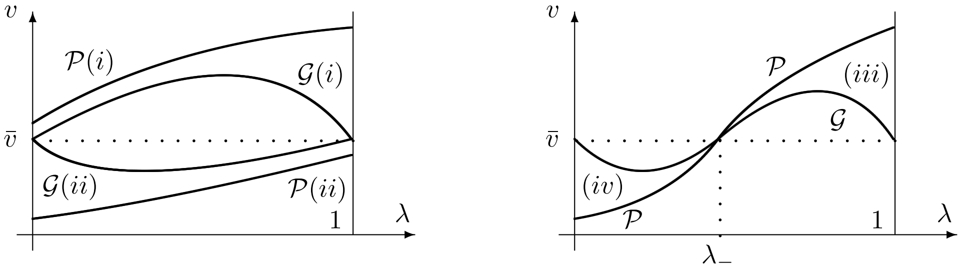


FIGURE 1. Mutual positions of the curves $\mathcal{G}$ and $\mathcal{P}$ in cases *(i)–(iv)*.

Lemma 2.2. *The following holds true for the curves $\mathcal{S}$, $\mathcal{G}$ and $\mathcal{P}$.*

(a) The curves $\mathcal{S}$ and $\mathcal{P}$ define increasing functions of λ.

(b) In cases (i) and (ii) the curves $\mathcal{G}$ and $\mathcal{P}$ do not intersect. In the former case, the curve $\mathcal{G}$ lies below $\mathcal{P}$ and above the line $v = \bar{v}$; in the latter, conversely.

(c) In cases (iii) and (iv) the curves $\mathcal{G}$ and $\mathcal{P}$ intersect only at λ_-. In the former case, the curve $\mathcal{G}$ lies below $\mathcal{P}$ and above the line $v = \bar{v}$, in $[\lambda_-, 1] \times \mathbb{R}$; in the latter, conversely in $[0, \lambda_-] \times \mathbb{R}$.

(d) In case (v) the curves $\mathcal{G}$ and $\mathcal{P}$ intersect only at $\bar{\lambda} \doteq \lambda_\pm$. The curve $\mathcal{G}$ lies above $\mathcal{P}$ and below the line $v = \bar{v}$ in $[0, \bar{\lambda}] \times \mathbb{R}$ and conversely in $[\bar{\lambda}, 1] \times \mathbb{R}$.

The proof follows by simple analytical arguments.

Now, we state our main results about the existence of the traveling wave profiles. We notice that the sufficient conditions below focus on the mutual positions of the curves $\mathcal{S}$ and $\mathcal{G}$, differently from the conditions in Lemma 2.1. In particular, when $\mathcal{S}$ and $\mathcal{G}$ intersect, both sides of the first equation in 6 vanish.

Theorem 2.3. *We consider the traveling-wave system 6 with end data 7 and assume 8. Moreover, referring to the cases listed in Lemma 2.1, we assume that:*

(i) $\mathcal{S}$ and $\mathcal{G}$ intersect at most once;
(ii) $\mathcal{S}$ and $\mathcal{G}$ do not intersect;
(iii) $\mathcal{S}$ and $\mathcal{G}$ intersect at most once in the strip $\mathbb{R} \times (\lambda_-, 1]$;
(iv) $\mathcal{S}$ and $\mathcal{G}$ do not intersect in $\mathbb{R} \times [0, \lambda_-)$.

Here, the intersection points are of multiplicity one.

Then, traveling waves exist and, for each speed c, they are unique up to a shift in ξ. Moreover, in case (v) the only solution is the trivial constant solution.

We give a sketch of the proof in one of the most significant cases, namely when the curves $\mathcal{S}$ and $\mathcal{G}$ intersect (once, by assumption). Consider, for instance, case *(i)* when $s_- > 0 > s_+$, when an intersection must occur, see Figure 2. Let (λ_0, v_0) be the intersection point; by Lemma 2.2, we have $v_0 > \bar{v}$ and $p(v_0, \lambda_0) > p_e$. It is easy to check that the field lines drive the trajectory toward (λ_0, v_0); then, the issue is how to solve the problem

$$\begin{cases} sv' = g, \\ \lambda' = -\frac{1}{c}(p - p_e)\lambda(\lambda - 1), \\ (v, \lambda)(0) = (v_0, \lambda_0). \end{cases} \tag{9}$$

It can be proved that there are exactly two solutions to 9 locally around $\xi = 0$ and both of them cross $\mathcal{S}$ transversally. Moreover, the slope of $\mathcal{S}$ at (λ_0, v_0) is larger than that of $\mathcal{G}$. We denote $\mathcal{S}_\pm = \{(\lambda, v) \colon \pm s > 0\}$ and

$$\Omega_1 = \{(\lambda, v) \in \mathbb{R}^2 \ : \ 0 < \lambda < \lambda_0,\ v > \bar{v},\ p > p_e,\ s > 0\},$$
$$\Omega_2 = \{(\lambda, v) \in \mathbb{R}^2 \ : \ \lambda_0 < \lambda < 1,\ v > \bar{v},\ p > p_e,\ s < 0\}.$$

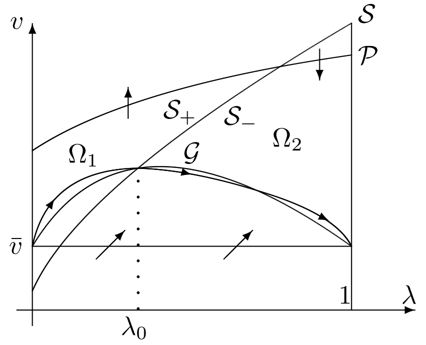


FIGURE 2. Case *(i)*, subcase $s_- > 0 > s_+$.

Let $\big(v(\xi), \lambda(\xi)\big)$ be the C^1 solution to 9 that crosses (v_0, λ_0) from $\mathcal{S}_+$ to $\mathcal{S}_-$. We claim that $\big(v(\xi), \lambda(\xi)\big)$ is a solution to 6.

On the one hand, the trajectory enters Ω_2 because of Lemma 2.2. The trajectory can exit Ω_2 neither through the boundary $\lambda = \lambda_0$, since $\lambda(\xi)$ is increasing, nor through $\lambda = 1$ at some $v \neq \bar{v}$, by the uniqueness of solutions to the initial-value problem. Moreover, the flow directions at both $\partial\Omega_2 \cap \{v = \bar{v}\}$ and $\partial\Omega_2 \cap \mathcal{P}$ are directed inside Ω_2, as well as those at points of Ω_2 close to $\mathcal{S}$. As a consequence, the trajectory is driven to the equilibrium point $(1, \bar{v})$ as $\xi \to +\infty$.

On the other hand, for $\xi < 0$ and close to 0, this trajectory lies in Ω_1. Since $\lambda(\xi)$ is increasing, the trajectory cannot enter Ω_1 through $\lambda = \lambda_0$. Moreover, the flow direction points toward the exterior of Ω_1 both at $\Omega_1 \cap \mathcal{P}$, because $g > 0$ there, and at $\Omega_1 \cap \{v = \bar{v}\}$ or in a neighborhood of $\Omega_1 \cap \mathcal{S}$. The uniqueness of the initial value problem rules out the possibility for the trajectory to enter Ω_1 through $\lambda = 0$ at $v \neq \bar{v}$. Thus, the trajectory must connect to the equilibrium point $(0, \bar{v})$ as $\xi \to -\infty$.

Hence, a traveling wave to 6 exists in this case.

3. **The global existence of smooth solutions.** In [11], the authors considered a system of balance laws, in one space dimension, provided with a strictly convex entropy η. They studied the initial-value problem for an initial datum close to an equilibrium point of the system and proved that the system had smooth solutions, defined globally in time, under two assumptions: the Shizuta-Kawashima and the strict entropy-dissipation condition had to be satisfied at the equilibrium point. We

prove in this section that, if $\alpha(\lambda) > 0$, system 1 satisfies both conditions at the equilibrium points lying either on the stable-liquid or on the stable-vapor curve.

We denote the flux function and the source term of 1 by F and G, respectively, i.e., $F(w) = (-u, p, 0)$ and

$$G(w) = \Big(0, -\alpha(\lambda)u, \big(p(v,\lambda) - p_e\big)\lambda(\lambda - 1)\Big) \doteq \big(0, q(w)\big),$$

where we set $\tau = 1$ for simplicity. We introduce the notation $\gamma_0 = \{(v, 0, 0) \colon v > 0,\ p \neq p_e\}$, $\gamma_1 = \{(v, 0, 1) \colon v > 0,\ p \neq p_e\}$, $\gamma_e = \{(v, 0, \lambda) \colon v > 0,\ p = p_e, \lambda \neq 0, 1\}$. We also write $\gamma_0^s = \{(v, 0, 0) \in \gamma_0 \colon p(v, 0) > p_e\}$ and $\gamma_1^s = \{(v, 0, 1) \in \gamma_1 \colon p(v, 0) < p_e\}$ for the stable-liquid, respectively, stable-vapor equilibrium points.

The Shizuta-Kawashima condition (in the hyperbolic setting [11]) holds for 1 at an equilibrium point $\bar{w}$ if

$$\ker DG(\bar{w}) \cap \{\text{eigenspaces of } DF(\bar{w})\} = \{0\}\,. \tag{10}$$

The eigenvalues of DF are $e_1 = -c$, $e_2 = 0$, $e_3 = c$, for $c = c(v, \lambda) = \sqrt{-p_v}$, with eigenvectors

$$r_1 = {}^t(1, c, 0), \quad r_2 = {}^t(p_\lambda, 0, -p_v), \quad r_3 = {}^t(-1, c, 0).$$

Then we have

$$\begin{aligned}(Dg)r_1 = -(Dg)r_3 = (0, -\alpha c, p_v\lambda(\lambda - 1))\,,\\ (Dg)r_2 = (0, -\alpha' u p_v, -\lambda(\lambda - 1)p_\lambda p_v - p_v(p - p_e)(2\lambda - 1))\,.\end{aligned}$$

Therefore $r_i \notin \ker Dg$ if $i = 1, 2, 3$ and $w \in \gamma_0 \cup \gamma_1 \cup \gamma_e$. As a consequence, condition 10 is satisfied on $\gamma_0 \cup \gamma_1 \cup \gamma_e$.

We now recall the conditions on the dissipation of entropy [11]. Let $\eta = \eta(w)$ be a strictly convex entropy for 1 with entropy flux $q = q(w)$; this means that $D\eta DF = Dq$. The system 1 is *entropy dissipative* in a neighborhood ω of an equilibrium point $\bar{w}$ if for every $w \in \omega$

$$\big(D\eta(w) - D\eta(\bar{w})\big) \cdot G(w) \leq 0. \tag{11}$$

Since η is strictly convex we can make the change of variables $W = D\eta(w)$ and write $W = (U, V) \in \mathbb{R} \times \mathbb{R}^2$. By denoting $\Phi = (D\eta)^{-1}$ we write 1 as

$$A_0(W)\partial_t W + A_1(W)\partial_x W = G\left(\Phi(W)\right). \tag{12}$$

Let $\bar{W} = \Phi^{-1}(\bar{w})$. Then, condition 11 becomes

$$\big(V - \bar{V}\big) \cdot Q(W) \leq 0, \tag{13}$$

for every W in a neighborhood $\Omega = \Phi^{-1}(\omega)$ of $\bar{W}$. Here $Q(W) = q(\Phi(W))$.

Definition 3.1. The system 1 is strictly entropy-dissipative in a neighborhood Ω of $\bar{W}$ if there exists a positive-definite 2×2 matrix $B\left(W, \bar{W}\right)$ such that

$$Q(W) = -B\left(W, \bar{W}\right)\left(V - \bar{V}\right)$$

for every $W \in \Omega$.

We notice that the condition of strict entropy-dissipation implies 11, by 13.

System 1 admits the entropy-entropy flux pair (η, q) [3] given by

$$\eta(w) = \frac{u^2}{2} - P(v, \lambda) + \phi(\lambda), \qquad q(w) = up. \tag{14}$$

Here, $P(v,\lambda) = \int_{v_o}^{v} p(z,\lambda)dz$ for some fixed $v_o > 0$ and ϕ is any smooth function. Remark that 14 is the canonical entropy pair for the p-system in the case that p is independent of λ and then $\phi = 0$.

Lemma 3.2. *The entropy η in 14 is strictly convex if*

$$\phi'' - P_{\lambda\lambda} - \frac{(p_\lambda)^2}{p_v} > 0\,. \tag{15}$$

Proof. We have $D\eta(w) = (-p, u, \phi' - P_\lambda)$ and

$$D^2\eta(w) = \begin{pmatrix} -p_v & 0 & -p_\lambda \\ 0 & 1 & 0 \\ -p_\lambda & 0 & \phi'' - P_{\lambda\lambda} \end{pmatrix}.$$

Then, $D^2\eta(w)$ is positive definite if both $\phi'' - P_{\lambda\lambda} > 0$ and $-p_v\left(\phi'' - P_{\lambda\lambda}\right) - (p_\lambda)^2 > 0$ hold true. Condition 15 implies both inequalities. □

For simplicity, we focus on the equilibrium region γ_0^s; below, we shall show how an analogous result can be obtained for γ_1^s. We choose [3]

$$\phi(\lambda) = \frac{C}{2}\lambda^2, \tag{16}$$

for a positive constant C and make the assumption

$$p_\lambda(v,0) = 0. \tag{17}$$

If v varies in a bounded interval $[v_o, v_1]$, then 15 holds for λ close to 0 if C is large enough to have

$$\int_{v_0}^{v_1} p_{\lambda\lambda}(v,0)\,dv < C. \tag{18}$$

Under the above assumptions, it is easy to see that system 1 is entropy dissipative on γ_0^s. In fact, fix $\bar{w} = (\bar{v}, 0, 0) \in \gamma_0^s$ and consider w in a neighborhood of $\bar{w}$. Condition 11 becomes

$$-\alpha u^2 + \big(\phi'(\lambda) - P_\lambda(v,\lambda)\big)\left(p(v,\lambda) - p_e\right)\lambda(\lambda - 1) \le 0.$$

In the same way, one easily see that, with the choices above, system 1 is neither entropy dissipative on the unstable branch of γ_0 nor on γ_e.

Lemma 3.3. *Let η be defined by 14 and 16; moreover, assume 17 and 18. Then, system 1 is strictly entropy-dissipative on γ_0^s.*

Proof. We compute, for $\Delta = \frac{1}{\delta\beta - \gamma^2} = -p_v\left(\phi''(\lambda) - P_{\lambda\lambda}\right) - p_\lambda^2$,

$$\Phi'(W) = \frac{1}{\Delta}\begin{pmatrix} \phi'' - P_{\lambda\lambda} & 0 & p_\lambda \\ 0 & \Delta & 0 \\ p_\lambda & 0 & -p_v \end{pmatrix} \doteq \begin{pmatrix} \delta & 0 & \gamma \\ 0 & 1 & 0 \\ \gamma & 0 & \beta \end{pmatrix}.$$

Then, system 1 can be written as 12 with

$$A_1(W) = Df\left(\Phi(W)\right)\Phi'(W) = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$G\left(\Phi(W)\right) = \begin{pmatrix} 0 \\ Q(W) \end{pmatrix} = \begin{pmatrix} 0 \\ -\alpha\left(\lambda(U,V_2)\right)V_1 \\ h\left(v(U,V_2), \lambda(U,V_2)\right) \end{pmatrix}.$$

We have $\Gamma_0^s = \Phi^{-1}(\gamma_0^s) = \{(-p(v,0),0,0)\colon v>0,\ p(v,0)>p_e\}$. Fix any $\bar{W} = (\bar{U},0,0)\in\Gamma_0^s$; with reference to Definition 3.1 we have $\bar{V}=0$. Then we can write

$$Q(W) = -\begin{pmatrix} \alpha\left(\lambda(U,V_2)\right) & 0 \\ 0 & -\frac{h(v(U,V_2),\lambda(U,V_2))}{V_2} \end{pmatrix}\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = -B(W,\bar{W})\begin{pmatrix} V_1 \\ V_2 \end{pmatrix}.$$

Since $\alpha>0$, we only need to prove that $-h\left(v(U,V_2),\lambda(U,V_2)\right)/V_2 > 0$, which immediately follows by 18. □

An analogous result can be proved for the equilibrium region γ_1^s. In that case, replace 16 with $\phi(\lambda) = \frac{C}{2}(\lambda-1)^2$, 17 with $p_\lambda(v,1)=0$, 18 with $\int_{v_0}^{v} p_{\lambda\lambda}(v,1)\,dv < C$.

Acknowledgments. Andrea Corli thanks Debora Amadori for some discussions on the strict entropy-dissipation condition.

REFERENCES

[1] D. Amadori and A. Corli, *A hyperbolic model of multi-phase flow*, in "Hyperbolic Problems: Theory, Numerics, Applications. Proceedings of the 11^{th} Int. Conf. on Hyperbolic Problems", (eds. S. Benzoni-Gavage and D. Serre), Springer, (2008), 407–414.

[2] D. Amadori and A. Corli, *On a model of multiphase flow*, SIAM J. Math. Anal., **40** (2008), 134–166.

[3] D. Amadori and A. Corli, *Global existence of BV solutions and relaxation limit for a model of multiphase reactive flow*, Nonlinear Anal., **72** (2010), 2527–2541.

[4] D. Amadori and A. Corli, *Solutions for a hyperbolic model of multiphase flow*, in "Proceedings of the Conference: Multiphase Flow in Industrial and Environmental Engineering", Springer, (2012), to appear.

[5] F. Asakura and A. Corli, *Global existence of solutions by path decomposition for a model of multiphase flow*, Quart. Appl. Math. (2012), to appear.

[6] A. Corli and H. Fan, *The Riemann problem for reversible reactive flows with metastability*, SIAM J. Appl. Math., **65** (2004/05), 426–457.

[7] A. Corli and H. Fan, *Traveling waves of phase transitions in porous media with phase-dependent damping*, preprint (2013).

[8] A. Corli and H. Fan, *Traveling waves of phase transitions in porous media*, Appl. Anal. (2013), to appear.

[9] C. M. Dafermos, *A system of hyperbolic conservation laws with frictional damping*, Z. Angew. Math. Phys., **46** (Special Issue) (1995), S294–S307.

[10] H. Fan, *On a model of the dynamics of liquid/vapor phase transitions*, SIAM J. Appl. Math., **60** (2000), 1270–1301.

[11] B. Hanouzet and R. Natalini, *Global existence of smooth solutions for partially dissipative hyperbolic systems with a convex entropy*, Arch. Ration. Mech. Anal., **169** (2003), 89–117.

[12] L. Hsiao and T.-P. Liu, *Convergence to nonlinear diffusion waves for solutions of a system of hyperbolic conservation laws with damping*, Comm. Math. Phys., **143** (1992), 599–605.

[13] L. Hsiao and D. Serre, *Global existence of solutions for the system of compressible adiabatic flow through porous media*, SIAM J. Math. Anal., **27** (1996), 70–77.

[14] M. Luskin, *On the existence of global smooth solutions for a model equation for fluid flow in a pipe*, J. Math. Anal. Appl., **84** (1981), 614–630.

[15] T. Nishida, "Nonlinear hyperbolic equations and related topics in fluid dynamics", Publications Mathématiques d'Orsay, Orsay, 1978.

E-mail address: `andrea.corli@unife.it`
E-mail address: `fanh@georgetown.edu`

SEMI-LAGRANGIAN SCHEMES FOR LINEAR AND FULLY NON-LINEAR HAMILTON-JACOBI-BELLMAN EQUATIONS

Kristian Debrabant

University of Southern Denmark
Department of Mathematics and Computer Science
Campusvej 55
5230 Odense M, Denmark

Espen Robstad Jakobsen

Norwegian University of Science and Technology
NO–7491, Trondheim, Norway

Abstract. We consider the numerical solution of Hamilton-Jacobi-Bellman equations arising in stochastic control theory. We introduce a class of monotone approximation schemes relying on monotone interpolation. These schemes converge under very weak assumptions, including the case of arbitrary degenerate diffusions. Besides providing a unifying framework that includes several known first order accurate schemes, stability and convergence results are given, along with two different robust error estimates. Finally, the method is applied to a super-replication problem from finance.

1. **Introduction.** In this paper we consider the numerical solution of partial differential equations of Hamilton-Jacobi-Bellman type,

$$u_t - \inf_{\alpha\in\mathcal{A}} \Big\{ L^\alpha[u](t,x) + c^\alpha(t,x)u + f^\alpha(t,x) \Big\} = 0 \qquad \text{in} \quad Q_T, \tag{1}$$

$$u(0,x) = g(x) \qquad \text{in} \quad \mathbb{R}^N, \tag{2}$$

where

$$L^\alpha[u](t,x) = \mathrm{tr}[a^\alpha(t,x)D^2u(t,x)] + b^\alpha(t,x)Du(t,x),$$

$Q_T := (0,T]\times\mathbb{R}^N$, and $\mathcal{A}$ is a complete metric space. The coefficients $a^\alpha = \frac{1}{2}\sigma^\alpha\sigma^{\alpha\top}$, b^α, c^α, f^α and the initial data g take values respectively in $\mathbb{S}^N$, the space of $N\times N$ symmetric matrices, $\mathbb{R}^N$, $\mathbb{R}$, $\mathbb{R}$, and $\mathbb{R}$. We will only assume that a^α is positive semi-definite, thus the equation is allowed to degenerate and hence not have smooth solutions in general. By solutions in this paper we will therefore always mean generalized solutions in the viscosity sense, see e. g. [6, 12]. Then the solution coincides with the value function of a finite horizon, optimal stochastic control problem [12].

To ensure comparison and well-posedness of (1)–(2) in the class of bounded x-Lipschitz functions, we will use the following standard assumptions on its data:

2000 *Mathematics Subject Classification.* Primary: 65M12, 65M15, 65M06; Secondary: 35K10, 35K55, 35K65, 49L25, 49L20.

Key words and phrases. Monotone approximation schemes, difference-interpolation methods, stability, convergence, error bound, degenerate parabolic equations, Hamilton-Jacobi-Bellman equations, viscosity solution.

(A1) For any $\alpha \in \mathcal{A}$, $a^\alpha = \frac{1}{2}\sigma^\alpha \sigma^{\alpha\,\top}$ for some $N \times P$ matrix σ^α. Moreover, there is a constant K independent of α such that

$$|g|_1 + |\sigma^\alpha|_1 + |b^\alpha|_1 + |c^\alpha|_1 + |f^\alpha|_1 \leq K,$$

where $|\phi|_1 = \sup_{(t,x)\in Q_T} |\phi(x,t)| + \sup_{(x,t)\neq(y,s)} \frac{|\phi(x,t)-\phi(y,s)|}{|x-y|+|t-s|^{1/2}}$ is a space-time Lipschitz/Hölder-norm.

The following result is standard.

Proposition 1. *Assume that (A1) holds. Then there exist a unique solution u of (1)–(2) and a constant C only depending on T and K from (A1) such that*

$$|u|_1 \leq C.$$

Furthermore, if u_1 and u_2 are sub- and supersolutions of (1) satisfying $u_1(0,\cdot) \leq u_2(0,\cdot)$, then $u_1 \leq u_2$.

2. Semi-Lagrangian schemes.

Following [8] we propose a class of approximation schemes for (1)–(2) which we call Semi-Lagrangian or SL schemes. These schemes converge under very weak assumptions, including the case of arbitrary degenerate diffusions. In particular, these schemes are L^∞-stable and convergent for problems involving diffusion matrices that are not diagonally dominant. This class includes (parabolic versions of) the "control schemes" of Menaldi [11] and Camilli and Falcone [4] and some of the monotone schemes of Crandall and Lions [7]. It also includes SL schemes for first order Bellman equations [5, 9] and some new versions as discussed in the following section.

The schemes are defined on a possibly unstructured family of grids $\{G_{\Delta t,\Delta x}\}$,

$$G = G_{\Delta t,\Delta x} = \{(t_n, x_i)\}_{n\in\mathbb{N}_0, i\in\mathbb{N}} = \{t_n\}_{n\in\mathbb{N}_0} \times X_{\Delta x},$$

for $\Delta t, \Delta x > 0$. Here $0 = t_0 < t_1 < \cdots < t_n < t_{n+1}$ satisfy

$$\max_n \Delta t_n \leq \Delta t \qquad \text{where} \qquad \Delta t_n = t_n - t_{n-1},$$

and $X_{\Delta x} = \{x_i\}_{i\in\mathbb{N}}$ is the set of vertices or nodes for a non-degenerate polyhedral subdivision of $\mathbb{R}^N$.

We consider the following general finite difference approximations of the differential operator $L^\alpha[\phi]$ in (1):

$$L_k^\alpha[\phi](t,x) := \sum_{i=1}^{M} \frac{\phi(t, x + y_{k,i}^{\alpha,+}(t,x)) - 2\phi(t,x) + \phi(t, x + y_{k,i}^{\alpha,-}(t,x))}{2k^2}, \tag{3}$$

for $k > 0$ and some $M \geq 1$. For this approximation we will assume

$$\text{(Y1)} \begin{cases} \displaystyle\sum_{i=1}^{M} [y_{k,i}^{\alpha,+} + y_{k,i}^{\alpha,-}] = 2k^2 b^\alpha + \mathcal{O}(k^4), \\ \displaystyle\sum_{i=1}^{M} [y_{k,i}^{\alpha,+} y_{k,i}^{\alpha,+\,\top} + y_{k,i}^{\alpha,-} y_{k,i}^{\alpha,-\,\top}] = 2k^2 \sigma^\alpha \sigma^{\alpha\,\top} + \mathcal{O}(k^4), \\ \displaystyle\sum_{i=1}^{M} [y_{k,i,j_1}^{\alpha,+} y_{k,i,j_2}^{\alpha,+} y_{k,i,j_3}^{\alpha,+} + y_{k,i,j_1}^{\alpha,-} y_{k,i,j_2}^{\alpha,-} y_{k,i,j_3}^{\alpha,-}] = \mathcal{O}(k^4), \\ \displaystyle\sum_{i=1}^{M} [y_{k,i,j_1}^{\alpha,+} y_{k,i,j_2}^{\alpha,+} y_{k,i,j_3}^{\alpha,+} y_{k,i,j_4}^{\alpha,+} + y_{k,i,j_1}^{\alpha,-} y_{k,i,j_2}^{\alpha,-} y_{k,i,j_3}^{\alpha,-} y_{k,i,j_4}^{\alpha,-}] = \mathcal{O}(k^4), \end{cases}$$

for all $j_1, j_2, j_3, j_4 = 1, 2, \ldots, N$ indicating components of the y-vectors.

Under assumption (Y1), a Taylor expansion shows that L_k^α is a second order consistent approximation satisfying

$$|L_k^\alpha[\phi] - L^\alpha[\phi]| \leq C(|D\phi|_0 + \cdots + |D^4\phi|_0)k^2 \tag{4}$$

for all smooth functions ϕ, where $|\phi|_0 = \sup_{(t,x)\in Q_T} |\phi(x,t)|$.

To relate this approximation to the spatial grid $X_{\Delta x}$, we replace ϕ by its interpolant $\mathcal{I}\phi$, yielding overall a semi-discrete approximation of (1),

$$U_t - \inf_{\alpha\in\mathcal{A}} \Big\{ L_k^\alpha[\mathcal{I}U](t,x) + c^\alpha(t,x)U + f^\alpha(t,x) \Big\} = 0 \quad \text{in} \quad (0,T)\times X_{\Delta x}.$$

We require the interpolation operator $\mathcal{I}$ to fulfill the following two conditions:

(I1) There are $K \geq 0, r \in \mathbb{N}$ such that for all smooth functions ϕ

$$|(\mathcal{I}\phi) - \phi|_0 \leq K|D^r\phi|_0 \Delta x^r.$$

(I2) There is a set of non-negative functions $\{w_j(x)\}_j$ such that

$$(\mathcal{I}\phi)(x) = \sum_j \phi(x_j)w_j(x),$$

and

$$w_j(x) \geq 0, \qquad w_i(x_j) = \delta_{ij}$$

for all $i,j \in \mathbb{N}$.

(I1) implies together with (4) that $L_k^\alpha[\mathcal{I}\phi]$ is a consistent approximation of $L^\alpha[\phi]$ if $\frac{\Delta x^r}{k^2} \to 0$. An interpolation satisfying (I2) is said to be *positive* and is *monotone* in the sense that $U \leq V$ implies that $\mathcal{I}U \leq \mathcal{I}V$. Typically $\mathcal{I}$ will be constant, linear, or multi-linear interpolation (i. e. $r \leq 2$ in (I1)), because higher order interpolation is not monotone in general.

The final scheme can now be found by discretizing in time using a parameter $\theta \in [0,1]$,

$$\delta_{\Delta t_n} U_i^n = \inf_{\alpha\in\mathcal{A}} \Big\{ L_k^\alpha[\mathcal{I}\bar{U}_\cdot^{\theta,n}]_i^{n-1+\theta} + c_i^{\alpha,n-1+\theta}\bar{U}_i^{\theta,n} + f_i^{\alpha,n-1+\theta} \Big\} \tag{5}$$

in G, where $U_i^n = U(t_n,x_i)$, $f_i^{\alpha,n-1+\theta} = f^\alpha(t_{n-1}+\theta\Delta t_n, x_i)$, ... for $(t_n,x_i) \in G$,

$$\delta_{\Delta t}\phi(t,x) = \frac{\phi(t,x)-\phi(t-\Delta t,x)}{\Delta t}, \qquad \text{and} \qquad \bar{\phi}_\cdot^{\theta,n} = (1-\theta)\phi_\cdot^{n-1} + \theta\phi_\cdot^n.$$

As initial conditions we take

$$U_i^0 = g(x_i) \quad \text{in} \quad X_{\Delta x}. \tag{6}$$

For the choices $\theta = 0, 1$, and $1/2$ the time discretization corresponds to respectively explicit Euler, implicit Euler, and midpoint rule. For $\theta = 1/2$, the full scheme can be seen as generalized Crank-Nicolson type discretization.

3. Examples of approximations L_k^α.

1. The approximation of Falcone [9] (see also [5]),

$$b^\alpha D\phi \approx \frac{\mathcal{I}\phi(x+hb^\alpha) - \mathcal{I}\phi(x)}{h},$$

corresponds to our L_k^α if $k = \sqrt{h}$, $y_k^{\alpha,\pm} = k^2 b^\alpha$.

2. The approximation of Crandall-Lions [7],

$$\frac{1}{2}\mathrm{tr}[\sigma^\alpha\sigma^{\alpha\top} D^2\phi] \approx \sum_{j=1}^{P} \frac{\mathcal{I}\phi(x+k\sigma_j^\alpha) - 2\mathcal{I}\phi(x) + \mathcal{I}\phi(x-k\sigma_j^\alpha)}{2k^2},$$

corresponds to our L_k^α if $y_{k,j}^{\alpha,\pm} = \pm k\sigma_j^\alpha$ and $M = P$.

3. The corrected version of the approximation of Camilli-Falcone [4] (see also [11]),

$$\frac{1}{2}\mathrm{tr}[\sigma^\alpha\sigma^{\alpha\top} D^2\phi] + b^\alpha D\phi$$
$$\approx \sum_{j=1}^{P} \frac{\mathcal{I}\phi(x+\sqrt{h}\sigma_j^\alpha + \frac{h}{P}b^\alpha) - 2\mathcal{I}\phi(x) + \mathcal{I}\phi(x-\sqrt{h}\sigma_j^\alpha + \frac{h}{P}b^\alpha)}{2h},$$

corresponds to our L_k^α if $k = \sqrt{h}$, $y_{k,j}^{\alpha,\pm} = \pm k\sigma_j^\alpha + \frac{k^2}{P}b^\alpha$ and $M = P$.

4. The new approximation obtained by combining approximations 1 and 2,

$$\frac{1}{2}\mathrm{tr}[\sigma^\alpha\sigma^{\alpha\top} D^2\phi] + b^\alpha D\phi$$
$$\approx \frac{\mathcal{I}\phi(x+k^2b^\alpha) - \mathcal{I}\phi(x)}{k^2} + \sum_{j=1}^{P} \frac{\mathcal{I}\phi(x+k\sigma_j^\alpha) - 2\mathcal{I}\phi(x) + \mathcal{I}\phi(x-k\sigma_j^\alpha)}{2k^2},$$

corresponds to our L_k^α if $y_{k,j}^{\alpha,\pm} = \pm k\sigma_j^\alpha$ for $j \le P$, $y_{k,P+1}^{\alpha,\pm} = k^2b^\alpha$ and $M = P+1$.

5. Yet another new approximation,

$$\frac{1}{2}\mathrm{tr}[\sigma^\alpha\sigma^{\alpha\top} D^2\phi] + b^\alpha D\phi \approx \sum_{j=1}^{P-1} \frac{\mathcal{I}\phi(x+k\sigma_j^\alpha) - 2\mathcal{I}\phi(x) + \mathcal{I}\phi(x-k\sigma_j^\alpha)}{2k^2}$$
$$+ \frac{\mathcal{I}\phi(x+k\sigma_P^\alpha + k^2b^\alpha) - 2\mathcal{I}\phi(x) + \mathcal{I}\phi(x-k\sigma_P^\alpha + k^2b^\alpha)}{2k^2},$$

corresponds to our L_k^α if $y_{k,j}^{\alpha,\pm} = \pm k\sigma_j^\alpha$ for $j < P$, $y_{k,P}^{\alpha,\pm} = \pm k\sigma_P^\alpha + k^2b^\alpha$ and $M = P$.

When σ^α does not depend on α but b^α does, approximations 4 and 5 are much more efficient than approximation 3.

4. Linear interpolation SL scheme (LISL).

To keep the scheme (5) monotone, linear or multi-linear interpolation is the most accurate interpolation one can use in general. In this typical case we call the full scheme (5)–(6) the LISL scheme. In the following, we denote by $c^{\alpha,+}$ the positive part of c^α. Then we have the following result by [8]:

Theorem 4.1. *Assume that (A1), (I1), (I2), and (Y1) hold.*

(a) The LISL scheme is monotone if the following CFL conditions hold:

$$(1-\theta)\Delta t\Big[\frac{M}{k^2} - c_i^{\alpha,n-1+\theta}\Big] \le 1 \quad \textit{and} \quad \theta\Delta t\, c_i^{\alpha,n-1+\theta} \le 1 \textit{ for all } \alpha, n, i. \qquad (7)$$

(b) The truncation error of the LISL scheme is $O(|1-2\theta|\Delta t + \Delta t^2 + k^2 + \frac{\Delta x^2}{k^2})$; it is first order accurate for $k = O(\Delta x^{1/2})$, $\Delta t = O(\Delta x)$ ($\Delta t = O(\Delta x^{1/2})$ if $\theta = \frac{1}{2}$).

(c) If $2\theta\Delta t \sup_\alpha |c^{\alpha,+}|_0 \leq 1$ and (7) holds, then there exists a unique bounded and L^∞-stable solution U of the LISL scheme converging uniformly to the solution u of (1)–(2) as $\Delta t, k, \frac{\Delta x}{k} \to 0$.

From this result it follows that the scheme is at most *first order accurate*, has *wide and increasing stencil* and a *good CFL condition*. From the truncation error and the definition of L_k^α the stencil is wide since the scheme is consistent only if $\Delta x/k \to 0$ as $\Delta x \to 0$ and has stencil length proportional to

$$l := \frac{\max\limits_{t,x,\alpha,i}\{|y_{k,i}^{\alpha,-}|, |y_{k,i}^{\alpha,+}|\}}{\Delta x} \sim \frac{k}{\Delta x} \to \infty \quad \text{as} \quad \Delta x \to 0.$$

Here we have used that if (Y1) holds and $\sigma \not\equiv 0$, then typically $y_{k,i}^{\alpha,\pm} \sim k$. Note that if $k = \Delta x^{1/2}$, then $l \sim \Delta x^{-1/2}$. Finally, in the case $\theta \neq 1$ the CFL condition for (5) is $\Delta t \leq Ck^2 \sim \Delta x$ when $k = O(\Delta x^{1/2})$, and it is much less restrictive than the usual parabolic CFL condition, $\Delta t = O(\Delta x^2)$.

Remark 1. The LISL scheme is consistent and monotone for arbitrary degenerating diffusions, without requiring that a^α is diagonally dominant or similar conditions. In comparison to other schemes applicable in this situation, like the ones of Bonnans-Zidani [3], it is much easier to analyze and to implement and faster in the sense that the computational cost for approximating the diffusion matrix is for fixed x, t, α independent of the stencil size.

5. The error estimate of [8].

To simplify the presentation, in the following we restrict to a uniform time-grid, $G = \Delta t\,\{0, 1, \ldots, N_T\} \times X_{\Delta x}$. Let $Q_{\Delta t,T} := \Delta t\,\{0, 1, \ldots, N_T\} \times \mathbb{R}^N$. To apply the regularization method of Krylov [10] we need a regularity and continuous dependence result for the scheme that relies on the following additional (covariance-type) assumptions: Whenever two sets of data σ, b and $\tilde{\sigma}, \tilde{b}$ are given, the corresponding approximations $L_k^\alpha, y_{k,i}^{\alpha,\pm}$ and $\tilde{L}_k^\alpha, \tilde{y}_{k,i}^{\alpha,\pm}$ in (3) satisfy

$$\text{(Y2)} \begin{cases} \displaystyle\sum_{i=1}^{M}[y_{k,i}^{\alpha,+} + y_{k,i}^{\alpha,-}] - [\tilde{y}_{k,i}^{\alpha,+} + \tilde{y}_{k,i}^{\alpha,-}] \leq 2k^2(b^\alpha - \tilde{b}^\alpha), \\ \displaystyle\sum_{i=1}^{M}[y_{k,i}^{\alpha,+} y_{k,i}^{\alpha,+\top} + y_{k,i}^{\alpha,-} y_{k,i}^{\alpha,-\top}] + [\tilde{y}_{k,i}^{\alpha,+} \tilde{y}_{k,i}^{\alpha,+\top} + \tilde{y}_{k,i}^{\alpha,-} \tilde{y}_{k,i}^{\alpha,-\top}] \\ \qquad -[y_{k,i}^{\alpha,+} \tilde{y}_{k,i}^{\alpha,+\top} + \tilde{y}_{k,i}^{\alpha,+} y_{k,i}^{\alpha,+\top} + y_{k,i}^{\alpha,-} \tilde{y}_{k,i}^{\alpha,-\top} + \tilde{y}_{k,i}^{\alpha,-} y_{k,i}^{\alpha,-\top}] \\ \leq 2k^2(\sigma^\alpha - \tilde{\sigma}^\alpha)(\sigma^\alpha - \tilde{\sigma}^\alpha)^\top + 2k^4(b^\alpha - \tilde{b}^\alpha)(b^\alpha - \tilde{b}^\alpha)^\top, \end{cases}$$

when $\sigma, b, y_k^\pm$ are evaluated at (t, x) and $\tilde{\sigma}, \tilde{b}, \tilde{y}_k^\pm$ are evaluated at (t, y) for all t, x, y.

Then one can prove the following error estimate [8]:

Theorem 5.1 (Error Bound I)**.** *Assume that (A1), (I1), (I2), (Y1), and (Y2) hold, and that $\Delta t, \Delta x > 0$, $k \in (0, 1)$ satisfy the CFL conditions (7). If u solves (1)–(2) and U solves (5)–(6), then there is $c_0 > 0$ such that for any $\Delta t \in (0, c_0)$*

$$|u - U| \leq C(|1 - 2\theta|\Delta t^{1/4} + \Delta t^{1/3} + k^{1/2} + \frac{\Delta x}{k^2}) \quad \textit{in} \quad G.$$

This error bound holds also for unstructured grids. For more regular solutions it is possible to obtain better error estimates, but general and optimal results are not available. The best estimate in our case is $O(\Delta x^{1/5})$ which is achieved when $k = O(\Delta x^{2/5})$ and $\Delta t = O(k^2)$. Note that the CFL conditions (7) already imply that

$\Delta t = O(k^2)$ if $\theta < 1$. Also note that the above bound does not show convergence when k is optimal for the LISL scheme ($k = O(\Delta x^{1/2})$).

6. **A new error estimate.**

In the above error estimate, the lower estimate on $u - U$ follows if you can prove regularity and continuous dependence results for the solution of the equation only. The proof of the upper estimate is symmetric and requires such results for the numerical solution. However, it is possible to avoid using such properties of the numerical solution by a clever approximation argument, see e. g. [1]. This allows for error estimates that show convergence for any k such that the scheme is consistent. We need an extra assumption on the coefficients:

(A2) The coefficients σ^α , b^α, c^α, f^α are continuous in α for all x, t.

Theorem 6.1 (Error Bound II)**.** *Assume that (A1), (A2), (I1) with $r = 2$ (∼linear interpolation), (I2), and (Y1) hold, and that $\Delta t, \Delta x > 0$, $k \in (0, 1)$ satisfy the CFL conditions* (7)*. If u solves* (1)*–*(2) *and U solves* (5)*–*(6)*, then there is $c_0 > 0$ such that for any $\Delta t \in (0, c_0)$*

$$u - U \geq C\Big(|1 - 2\theta|\Delta t^{1/4} + \Delta t^{1/3} + k^{1/2} + \frac{\Delta x}{k}\Big) \quad \textit{in} \quad G,$$

$$u - U \leq C\Big(|1 - 2\theta|\Delta t^{1/10} + \Delta t^{1/8} + k^{1/5} + \big(\frac{\Delta x}{k}\big)^{1/2}\Big) \quad \textit{in} \quad G.$$

With optimal k for the LISL scheme, $\Delta t = O(k^2)$ and $k = O(\Delta x^{1/2})$, we find that $u - U = O(\Delta x^{1/10})$.

Proof. By a direct computation the local truncation error of the method is bounded by

$$\frac{|1-2\theta|}{2}|\phi_{tt}|_0\Delta t + C\Big(\Delta t^2\left(|\phi_{tt}|_0 + |\phi_{ttt}|_0 + |D\phi_{tt}|_0 + |D^2\phi_{tt}|_0\right) + |D^2\phi|_0\frac{\Delta x^2}{k^2} + (|D\phi|_0 + \cdots + |D^4\phi|_0)k^2\Big)$$

for smooth ϕ (cf. Lemma 4.1 in [8]). Moreover if also $\partial_t^{k_1} D_x^{k_2}\phi = O(\varepsilon^{1-2k_1-k_2})$ for any $k_1, k_2 \in \mathbb{N}_0$, then the truncation error is of order

$$(1 - 2\theta)\Delta t\varepsilon^{-3} + \Delta t^2\varepsilon^{-5} + k^2\varepsilon^{-3} + \frac{\Delta x^2}{k^2}\varepsilon^{-1} =: E(\varepsilon).$$

Since the scheme is monotone (under the CFL condition) and condition (A1) holds, it now follows from Theorem 3.1 in [1] that

$$C\inf_{\varepsilon>0}\big(\varepsilon + E(\varepsilon)\big) \leq u - U \leq C\inf_{\varepsilon>0}\big(\varepsilon^{1/3} + E(\varepsilon)\big),$$

and we complete the proof optimizing over ε (as e. g. in [1, 8]). □

7. **Convergence test for a super-replication problem.**

We consider a test problem from [2] which was used to test convergence rates for numerical approximations of a super-replication problem from finance. The corresponding PDE is

$$\inf_{\alpha_1^2+\alpha_2^2=1}\left\{\alpha_1^2 u_t(t,x) - \frac{1}{2}\mathrm{tr}\left(\sigma^\alpha(t,x)\sigma^{\alpha\top}(t,x)D^2u(t,x)\right)\right\} = f(t,x) \tag{8}$$

with $0 \leq x_1, x_2 \leq 3$, $\sigma^\alpha(t,x) = \begin{pmatrix} \alpha_1 x_1 \sqrt{x_2} \\ \alpha_2 \eta(x_2) \end{pmatrix}$ and $\eta(x) = x(3-x)$. We take $u(t,x) = 1 + t^2 - e^{-x_1^2 - x_2^2}$ as exact solution as in [2], and then f is forced to be

$$f(t,x) = \frac{1}{2}\Bigg(u_t - \frac{1}{2}x_1^2 x_2 u_{x_1x_1} - \frac{1}{2}x_2^2(3-x_2)^2 u_{x_2x_2}$$
$$-\sqrt{\left(-u_t + \frac{1}{2}x_1^2 x_2 u_{x_1x_1} - \frac{1}{2}x_2^2(3-x_2)^2 u_{x_2x_2}\right)^2 + \left(x_1\sqrt{x_2}^3(3-x_2)u_{x_1x_2}\right)^2}\Bigg).$$

In [2] $\eta(x) = x$, while we take $\eta(x) = x(3-x)$ to prevent the LISL scheme from overstepping the boundaries. Note that changing η does *not* change the solutions as long as $\eta > 0$ in the interior of the domain, see [2], and hence the above equation is equivalent to the equation used in [2]. The initial values and Dirichlet boundary values at $x_1 = 0$ and $x_2 = 0$ are taken from the exact solution. As in [2], at $x = 3$ and $y = 3$ homogeneous Neumann boundary conditions are implemented. To approximate the values of α_1, α_2, the Howard algorithm is used (see [2]), which requires an implicit time discretization, so we choose $\theta = 1$. We choose $k = \sqrt{\Delta x}$ and a regular triangular grid. The numbers of time steps are chosen as $\frac{1}{\Delta x}$.

The results at $t = 1$ are given in Table 1. The numerical order of convergence is approximately one.

Δx	$\lvert u - U\rvert_0$	rate
1.50e-1	2.01e-1	
7.50e-2	9.49e-2	1.08
3.75e-2	4.29e-2	1.15
1.87e-2	1.94e-2	1.15

TABLE 1. Results for the convergence test for the super-replication problem at $t = 1$

Remark 2. Equation (8) can not be written in a form (1) satisfying the assumptions of this paper, so the results of this paper do not apply to this problem. However, it seems possible to extend them to cover this problem using comparison results from [2] along with L^∞-bounds on the numerical solution that follow from the maximum principle.

8. **A super-replication problem.**

We apply our method to solve a problem from finance, the super-replication problem under gamma constraints considered in [2]. It consists of solving equation (8) with $f \equiv 0$, Neumann boundary conditions and σ^α as in Subsection 7, and initial and Dirichlet conditions given by

$$u(t,x) = \max(0, 1 - x_1), \quad t = 0 \quad or \quad x_1 = 0 \quad or \quad x_2 = 0.$$

The solution obtained with the LISL scheme is given in Figure 1 and coincides with the solution found in [2]. It gives the price of a put option of strike and maturity 1, and x_1 and x_2 are respectively the price of the underlying and the price of the forward variance swap on the underlying.

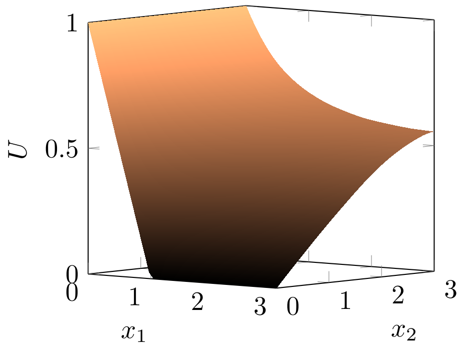


FIGURE 1. Numerical solution of super-replication problem at $t = 1$

REFERENCES

[1] Guy Barles and Espen Robstad Jakobsen, *Error bounds for monotone approximation schemes for parabolic Hamilton-Jacobi-Bellman equations*, Math. Comp., **76** (2007), 1861–1893.

[2] Olivier Bokanowski, Benjamin Bruder, Stefania Maroso and Hasnaa Zidani, *Numerical approximation for a superreplication problem under gamma constraints*, SIAM J. Numer. Anal., **47** (2009), 2289–2320.

[3] Joseph Frédéric Bonnans, Élisabeth Ottenwaelter and Hasnaa Zidani, *A fast algorithm for the two dimensional HJB equation of stochastic control*, M2AN Math. Model. Numer. Anal., **38** (2004), 723–735.

[4] Fabio Camilli and Maurizio Falcone, *An approximation scheme for the optimal control of diffusion processes*, RAIRO Modél. Math. Anal. Numér., **29** (1995), 97–122.

[5] Italo Capuzzo Dolcetta, *On a discrete approximation of the Hamilton-Jacobi equation of dynamic programming*, Appl. Math. Optim., **10** (1983), 367–377.

[6] Michael G. Crandall, Hitoshi Ishii and Pierre-Louis Lions, *User's guide to viscosity solutions of second order partial differential equations*, Bull. Amer. Math. Soc. (N.S.), **27** (1992), 1–67.

[7] Michael G. Crandall and Pierre-Louis Lions, *Convergent difference schemes for nonlinear parabolic equations and mean curvature motion*, Numer. Math., **75** (1996), 17–41.

[8] Kristian Debrabant and Espen Robstad Jakobsen, *Semi-Lagrangian schemes for linear and fully non-linear diffusion equations*, Math. Comp., **82** (2013), 1433–1462.

[9] Maurizio Falcone, *A numerical approach to the infinite horizon problem of deterministic control theory*, Appl. Math. Optim., **15** (1987), 1–13.

[10] Nicolai V. Krylov, *On the rate of convergence of finite-difference approximations for Bellman's equations with variable coefficients*, Probab. Theory Related Fields, **117** (2000), 1–16.

[11] José-Luis Menaldi, *Some estimates for finite difference approximations*, SIAM J. Control Optim., **27** (1989), 579–607.

[12] Jiongmin Yong and Xun Yu Zhou, "Stochastic Controls", vol. 43 of *Applications of Mathematics*, Springer-Verlag, New York, 1999.

E-mail address: `debrabant@imada.sdu.dk`
E-mail address: `erj@math.ntnu.no`

ON A SOLUTION RECONSTRUCTION AND LIMITING PROCEDURE FOR UNSTRUCTURED FINITE VOLUMES

A.I. Delis

Department of Sciences, Division of Mathematics
Technical University of Crete, University Campus
Chania, Crete 73100, Greece

I.K. Nikolos

Department of Production Engineering & Management
Technical University of Crete, University Campus
Chania, Crete 73100, Greece

Abstract. The present paper deals with the continuous work of extending the application of a multidimensional-type solution reconstruction and limiting procedure into the finite volume (FV) computation of 2D compressible flows. Based on a MUSCL-type technique, this reconstruction procedure identifies and cures poor connected grids without suffering from loss of accuracy by taking into account geometrical characteristics of computational triangular and hybrid meshes and is independent of the Riemann solver used. Monotonicity in the solution is enforced by a limiting strategy that implements well-known edge-type limiters hence avoiding the procedure of solving any minimization problems. Through several test cases, it is observed that the methodology provides quite desirable performances in retaining the formal order of accuracy, controlling numerical oscillations as well as capturing key flow features.

1. **Introduction.** In multidimensional high-resolution FV schemes, on unstructured meshes, numerous solution reconstruction and limiting strategies have been developed to resolve complex flows. Many of these strategies involve the construction of an appropriate linear representation of the solution variables within each FV element, which is then limited as to enforce positivity and stability constraints on the scheme, usually based on the satisfaction of the Maximum Principle [4]. Although current reconstruction and limiting methods have enjoyed success in many CFD applications, there is no consensuses on a global optimal reconstruction strategy that fulfills a high-level of accuracy, robustness and convergence. As such, the search for efficient reconstruction and limiting processes, in a multidimensional context, is still an active field of research, see for example [1, 6, 5, 3] and references therein. Recently, in [2] and [3], a cell-centered finite volume (CCFV) scheme of the Godunov-type with a novel solution reconstruction and limiting procedure was developed and tested for smooth and non-smooth shallow water flow computations. This novel alternative procedure was developed in order to apply in the reconstruction an edge-based limiting strategy that takes into account geometrical characteristics of the computational mesh. Grids with poor connectivity [4, 1], i.e.

2000 *Mathematics Subject Classification.* Primary: 65M08, 70M12; Secondary: 35L65.
Key words and phrases. Finite volumes, linear reconstruction, limiting, Euler equations.

those whose flux integration points do not coincide with the location to which reconstructed values are computed, can be properly treated. The use of edge-type limiters avoids the procedure of solving any minimization problems or the need to use any tunable parameters.

In this presentation, the reconstruction and limiting procedure proposed in [2, 3] is applied to the approximation of the Euler equations and compares its performance when implementing truly multidimensional limiters. Two well-known approximate Riemann solvers are implemented, demonstrating the procedure's independence to the solver used, as well as its universal applicability, efficiency and robustness to either conservative or primitive variable reconstructions. Its applicability to some node-centered (vertex-centered) FV schemes is also briefly sketched.

2. **The Euler equations.** In their conservative form the 2D Euler equations read as

$$\partial_t \mathbf{U} + \nabla \cdot \mathcal{H}(\mathbf{U}) = 0 \quad \text{on} \quad \Omega \times [0,t] \subset \mathbb{R}^2 \times \mathbb{R}^+, \tag{1}$$

where $\Omega \times [0,t]$ is the space-time Cartesian domain over which solutions are sought, $\mathbf{U}$ is the vector of the conserved variables and $\mathcal{H} = [\mathbf{F}, \mathbf{G}]$ are the nonlinear flux vectors defined as

$$\mathbf{U} = \begin{bmatrix} \rho \\ \rho u \\ \rho v \\ \rho e_T \end{bmatrix}, \quad \mathbf{F}(\mathbf{U}) = \begin{bmatrix} \rho u \\ \rho u^2 + p \\ \rho u v \\ (\rho e_T + p) u \end{bmatrix}, \quad \mathbf{G}(\mathbf{U}) = \begin{bmatrix} \rho v \\ \rho u v \\ \rho v^2 + p \\ (\rho e_T + p) v \end{bmatrix}. \tag{2}$$

Here, ρ is the flow density, p is the flow pressure, e_T is the specific total energy, and $\mathbf{u} = [u, v]^{\mathrm{T}}$ are the velocity components. The system is completed by the equation of state for a perfect gas, $p = (\gamma_s - 1)\left(\rho e_T - \frac{1}{2}\rho \|\mathbf{u}\|^2\right)$, where $\gamma_s = 1.4$ is the ratio of the specific heats. Hyperbolic system (1) is supplemented by the initial condition $\mathbf{U}(x,y,0) = \mathbf{U}_0(x,y)$, $x, y \in \Omega$ and by appropriate boundary conditions (periodic, inflow, outflow, slip) on the boundary $\partial\Omega$ of Ω.

3. **FV framework and the MUCL-type reconstruction.** By considering a conforming triangulation $\mathcal{T}^{h_N}$ of Ω, with characteristic length h_N, to be a set of finitely many triangular cells[1] $T_p \subset \Omega$, $p = 1, 2, \ldots, N$, we construct FV approximations on each T_p, see Fig. 1. Flow variables are placed at the barycenter of T_p, and we denote the set of indices of the neighboring triangles of T_p by $K(p) := \{q \in \mathbb{N} \mid \partial T_p \cap \partial T_q \text{ is a face of } T_p\}$. Then the semi-discretized form of (1) over each T_p can be written as follows

$$|T_p| \frac{\partial \mathbf{U}_p}{\partial t} + \sum_{q \in K(p)} \mathcal{H}^\star(\mathbf{U}_p^L, \mathbf{U}_q^R) \cdot \mathbf{n}_q = 0, \tag{3}$$

where $|T_p|$ is the area of T_p, $\mathbf{n}_q$ is the outward normal vector and $\mathcal{H}^\star$ is the numerical flux vector at each face's midpoint M, evaluated using the left and right states existing at the two sides of M, denoted as $\mathbf{U}_p^L$ and $\mathbf{U}_q^R$. Although a one-point quadrature rule is used here the ideas presented next can be applied to high-order integration, e.g. using Gauss quadrature. In general, this numerical flux function can be calculated as an exact or approximate local solution of the Riemann problem posed at a cell's face. In this work, Roe's approximate Riemann solver and the HLLC one have been utilized.

[1] Our discussion is based on the cell-centered FV (CCFV) approach, but it also holds for node-centered FV when the word "cell" is simply replaced by the "control volume", C_P.

3.1. **Solution reconstruction and limiting procedure.** For achieving second-order spatial accuracy most FV implementations calculate the $\mathbf{U}_p^L$ and $\mathbf{U}_q^R$ values assuming that the solution varies linearly in each cell, starting from given average solution values of adjacent cells, i.e. MUSCL-type local reconstructions. Three choices exist for this reconstruction; primitive, conservative and characteristic variables. Here, reconstruction on primitive and conservative variables has been investigated. To prevent oscillations from developing in the solution by controling the total variation of the reconstructed field, slope limiting has to be applied. In the present work, an alternative approach, compared to these usually implemented in unstructured CCFV schemes is used following [2, 3] and is briefly repeated next for completeness. Strict monotonicity in the reconstruction is enforced by the use of edge-based limiter functions, usually reserved for node-centered FV schemes of the median-dual type [2]. One such limiter is the modified Van Albada-Van Leer which is differentiable for linearly varying flow variables. Continuous differentiability helps in achieving smooth transition between discontinuous jumps with first-order representation and sharp but continuous gradients which require second-order consistency.

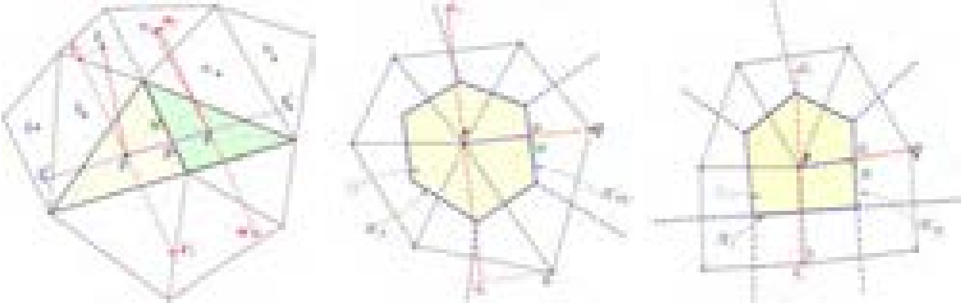


FIGURE 1. Cell-centered FV (left), centroid-dual FV (middle) and hybrid mesh (right)

Starting with a constant piecewise approximation of the $i-$th component of $\mathbf{W}_p$ (representing either primitive or conservative variables) and since we wish to apply edge-based limiters, one is forced to compute reconstructed values at the intersection point D of face $\partial T_q \cap \partial T_p$ and $\overline{pq}$, see Fig. 1, as to compare with the reference gradient value $w_{i,q} - w_{i,p}$. This choice seems natural also from a geometrical point of view since it corresponds to the linear interpolation between p and q. Therefore, we start by computing left and right extrapolated values at D as,

$$(w_{i,p})_D^L = w_{i,p} + \mathbf{r}_{pD} \cdot \nabla w_{i,p} \quad \text{and} \quad (w_{i,q})_D^R = w_{i,q} - \mathbf{r}_{Dq} \cdot \nabla w_{i,q}, \tag{4}$$

where $\mathbf{r}$ is a position vector relative to the centroid of the cell and $\nabla(\cdot)$ a gradient operator. Note that point D does not coincide, in general, with the face's midpoint M. Then a limiter function, $\Phi(\cdot,\cdot)$, has to be applied and its arguments have to be consecutive gradients of the solution defined in an upwind manner around face $\partial T_q \cap \partial T_p$. Thus, a virtual value q' has to be defined as a node upwind of q. The main difficulty now lies in the need to define this virtual q' value (similar we can define a p' value for p). We first denote the local centered reference gradient as $(\nabla w_{i,q})^{\text{cnt}} \cdot \mathbf{r}_{pq} = w_{i,q} - w_{i,p}$, and we compute the upwind gradient $w_{i,q} - w_{i,q'}$ by expressing the virtual unknown q' value using known ones by assuming that q' is chosen such that it lies along edge $\overline{pq}$ and that q is at the center of $\overline{pq'}$. In this case,

$$\begin{aligned} w_{i,q'} - w_{i,q} &= (w_{i,q'} - w_{i,p}) - (w_{i,q} - w_{i,p}) = (\nabla w_{i,q}) \cdot \mathbf{r}_{pq'} - (w_{i,q} - w_{i,p}) \\ &= 2\,(\nabla w_{i,q}) \cdot \mathbf{r}_{pq} - (\nabla w_{i,q})^{\text{cnt}} \cdot \mathbf{r}_{pq}. \end{aligned}$$

Then, to invoke monotonicity, limiting is performed and the ratio of the corresponding lengths has to be used, resulting in the left and right states at face $\partial T_q \cap \partial T_p$ as

$$(w_{i,q})_D^R = w_{i,q} - \frac{||\mathbf{r}_{Dq}||}{||\mathbf{r}_{pq}||}\Phi\left((\nabla w_{i,q})^{\text{upw}} \cdot \mathbf{r}_{pq}, (\nabla w_{i,q})^{\text{cnt}} \cdot \mathbf{r}_{pq}\right); \tag{5}$$

$$(w_{i,p})_D^L = w_{i,p} + \frac{||\mathbf{r}_{pD}||}{||\mathbf{r}_{pq}||}\Phi\left((\nabla w_{i,p})^{\text{upw}} \cdot \mathbf{r}_{pq}, (\nabla w_{i,p})^{\text{cnt}} \cdot \mathbf{r}_{pq}\right), \tag{6}$$

where the upwind limiter arguments are given as

$$(\nabla w_{i,q})^{\text{upw}} = 2\nabla w_{i,q} - (\nabla w_{i,q})^{\text{cnt}} \quad \text{and} \quad (\nabla w_{i,p})^{\text{upw}} = 2\nabla w_{i,p} - (\nabla w_{i,p})^{\text{cnt}}.$$

In an ideal unstructured grid the variables are extrapolated to the center M of a cell's face and as such the numerical integration of the exact flux with the midpoint rule will be exact for linear functions along $\partial T_q \cap \partial T_p$. If the variables are extrapolated to a different location, e.g. at point D, then the one-point interpolation is expected to be only first-order accurate, especially for types of grids where the distance between optimal location M and the extrapolated location D is large [4, 2, 3]. This inconsistency had to be corrected thus, a correction was proposed and tested for smooth shallow water flow conditions in [2]. Hence, after the reconstructed values (5) and (6) at D have been computed, a directional correction is applied in order to compute reconstructed values at M, as follows,

$$(w_{i,p})_M^L = (w_{i,p})_D^L + \mathbf{r}_{DM} \cdot \nabla w_{i,p} \quad \text{and} \quad (w_{i,q})_M^R = (w_{i,q})_D^R + \mathbf{r}_{DM} \cdot \nabla w_{i,q}. \tag{7}$$

Since the gradient estimates used in the above correction terms are unlimited it was shown that, accurate gradient computations would result in an accurate correction, in the sense of retaining second order accuracy for smooth flows, on poor connected grids where the distance between D and M is large [2], even for highly stretched meshes. However, further considerations had to be taken into account if shocks are to be present in the flow field. As such, the correction terms in (7) have to be properly limited along the direction of $\overline{DM}$ as proposed in [3], thus enhancing the multidimensional character of the reconstruction. Achieving this is not trivial since proper reference values have to be defined, as to calculate limiter arguments that are physically meaningful. Assuming we want to properly limit the directional correction added to $(w_{i,p})_M^L$ in (7), we first identify the set of indices $l_j, j = 1, 2, 3$ of the triangles T_{l_j} that have now a common vertex with T_p in the direction of $\overline{DM}$. We choose as a reference triangle the one for which $\overline{pl_j}$ has the smallest angle with $\overline{DM}$, which is T_{l_2} in Fig. 1, and project its cell center in the direction of $\overline{DM}$, with $\overline{pk_2}$ being that projection. Now, the extrapolated value at k_2 is calculated from the value at the barycenter l_2 as

$$w_{i,k_2} = w_{i,l_2} + \mathbf{r}_{l_2 k_2} \cdot \nabla w_{i,l_2}.$$

We can now define the local central reference gradient as $(\nabla w_{i,p})^{\text{cnt}} \cdot \mathbf{r}_{pk_2} = w_{i,k_2} - w_{i,p}$, and compute the upwind gradient $w_{i,p} - w_{i,k_2'}$ by expressing the virtual unknown k_2' value using known values as detailed before. This leads to the left reconstructed value (now corrected and limited) at the flux integration point M,

$$(w_{i,p})_M^L = (w_{i,p})_D^L + \frac{||\mathbf{r}_{DM}||}{||\mathbf{r}_{pk_2}||}\Phi\left((\nabla w_{i,p})^{\text{upw}} \cdot \mathbf{r}_{pk_2}, (\nabla w_{i,p})^{\text{cnt}} \cdot \mathbf{r}_{pk_2}\right), \tag{8}$$

where now

$$(\nabla w_{i,p})^{\text{upw}} = 2\nabla w_{i,p} - (\nabla w_{i,p})^{\text{cnt}}.$$

With similar reasoning, the right limited reconstructed value at M is computed as

$$(w_{i,q})_M^R = (w_{i,q})_D^R + \frac{||\mathbf{r}_{DM}||}{||\mathbf{r}_{qm_2}||}\Phi\left((\nabla w_{i,q})^{\text{upw}} \cdot \mathbf{r}_{qm_2}, (\nabla w_{i,q})^{\text{cnt}} \cdot \mathbf{r}_{qm_2}\right). \tag{9}$$

Similar rationale can be followed to correct the inconsistency between points D and M for node-centered FV schemes of the centroid-dual type, as depicted in Fig. 1.

Finally, it remains to define appropriate gradient operators for the reconstruction presented above for the CCFV approach. Here, the gradient is computed in the closed path defined for every T_p by connecting the barycenters of the triangles having a common vertex with T_p, taking into account the assumption that the gradient is constant, i.e. Green-Gauss (GG) linear reconstruction. As it was demonstrated in [3], when this (wide) stencil is used for gradient computations an almost identical convergence behavior is achieved on different grid types, with no reduction on the asymptotic convergence rate while for steady-state calculations convergence is greatly improved. This is due to the fact that the data points involved in this gradient computation satisfy the so-called good neighborhood for Van Leer limiting.

4. **Numerical Tests and Results.** In our scheme, named CCFVw2L, the Van Albada-Van Leer limiter was implemented on the tests to follow, while a second-order explicit SSP Runge-Kutta time integration is used under the usual CFL stability condition for the time step computation. Ghost cells are implemented for boundary treatment. To test performance and convergence properties regular and irregular grids, shown in Fig. 2, have been used. These grids exhibit different connectivity behaviors for internal and boundary cells [3]. Comparisons between the results obtained with the Venkatakrishnan [7] V-limiter and the MLPu2 [6] one are also presented. It is noted that, these two truly multidimensional limiters have a tunable parameter[2], K, that is problem dependent and needs to be carefully chosen.

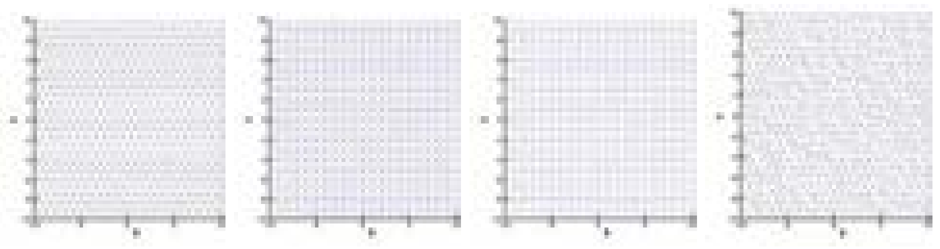


FIGURE 2. Regular and irregular grids: types I-IV from left to right

[2] If $K = 0$ the limiter is always active, and strict monotonicity is maintained, while a very large value of K essentially means no limiting and monotonicity is violated.

4.1. **Isentropic Vortex Problem.** This 2D vortex problem is often used as a benchmark for comparing numerical methods for fluid dynamics. The flow-field is smooth and the exact solution is known. The mean flow is with $\rho_\infty = 1, p_\infty = 1$, and $[u_\infty, v_\infty]^{\mathrm{T}} = [1,1]^{\mathrm{T}}$ in a domain $\Omega = [-10,10] \times [-10,10]$ with periodic boundary conditions. An isentropic vortex is added [6, 5] and the exact solution of the problem is the initial solution shifted by $(u_\infty t, v_\infty t)$ thus, numerical phase (dispersion) and amplitude (dissipation) errors are easy to identify. Roe's Riemann solver is implemented for this test case applying reconstruction on the primitive variables. In Fig. 3 the density contours of computed solutions compared with the exact solution at time $t = 2T$ are presented for the CCFVw2L scheme and those using the V and MLPu2 limiters with $K = 5$ on a type-IV distorted grid. Next and in Fig 4 we report grid convergence studies on all different grids, having been consistently refined.

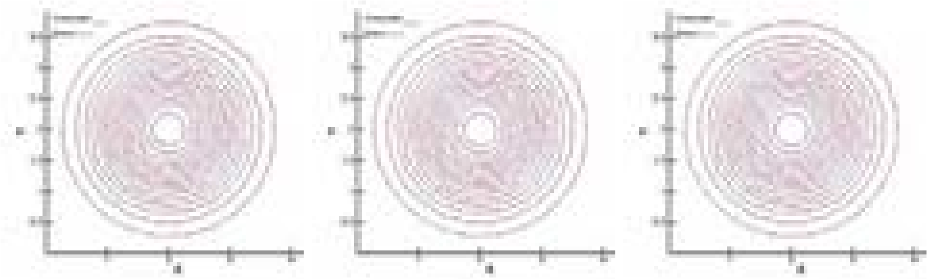


Figure 3. Isentropic vortex at $t = 2T$: V-limiter (left), MLPu2 (middle) and CCFVw2L (right)

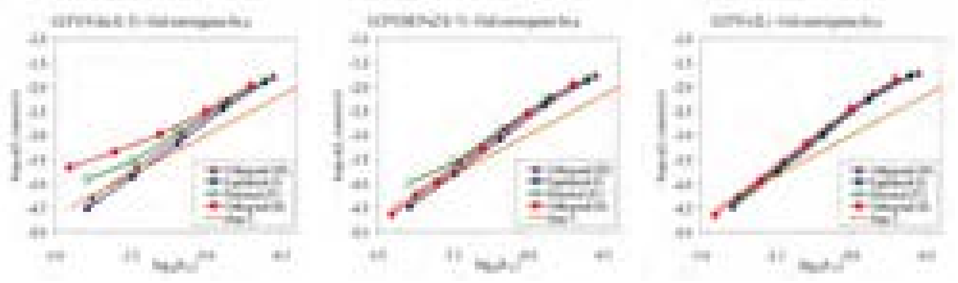


Figure 4. Isentropic vortex: convergence results on all grid types

4.2. **Shock tube problems.** These (one-dimensional in nature) test cases are chosen as to test the capability of the proposed methodology in resolving various linear and non-linear waves on unstructured grids. The computational domain is $[0,1] \times [0,0.1]$ with a triangulation of $N = 16,000$ cells on a type-II mesh. Three cases are considered, namely, (a) Sod's problem, (b) Harten-Lax problem and (c)

Supersonic expansion, with Riemann-type initial conditions given respectively as

(a) $(\rho_L, u_L, v_L, p_L) = (1, 0, 0, 1)$ and $(\rho_R, u_R, v_R, p_R) = (1, 0, 0, 1)$,

(b) $(\rho_L, u_L, v_L, p_L) = (0.445, 0.698, 0, 3.528)$ and
$(\rho_R, u_R, v_R, p_R) = (0.5, 0, 0, 0.571)$,

(c) $(\rho_L, u_L, v_L, p_L) = (1, -2, 0, 0.4)$ and $(\rho_R, u_R, v_R, p_R) = (1, 2, 0, 0.4)$

and the HLLC approximate Riemann solver is used in all problems applying reconstruction on the conservative variables. Comparisons between the proposed CCFVw2L scheme and the MLPu2 are shown in Fig. 5.

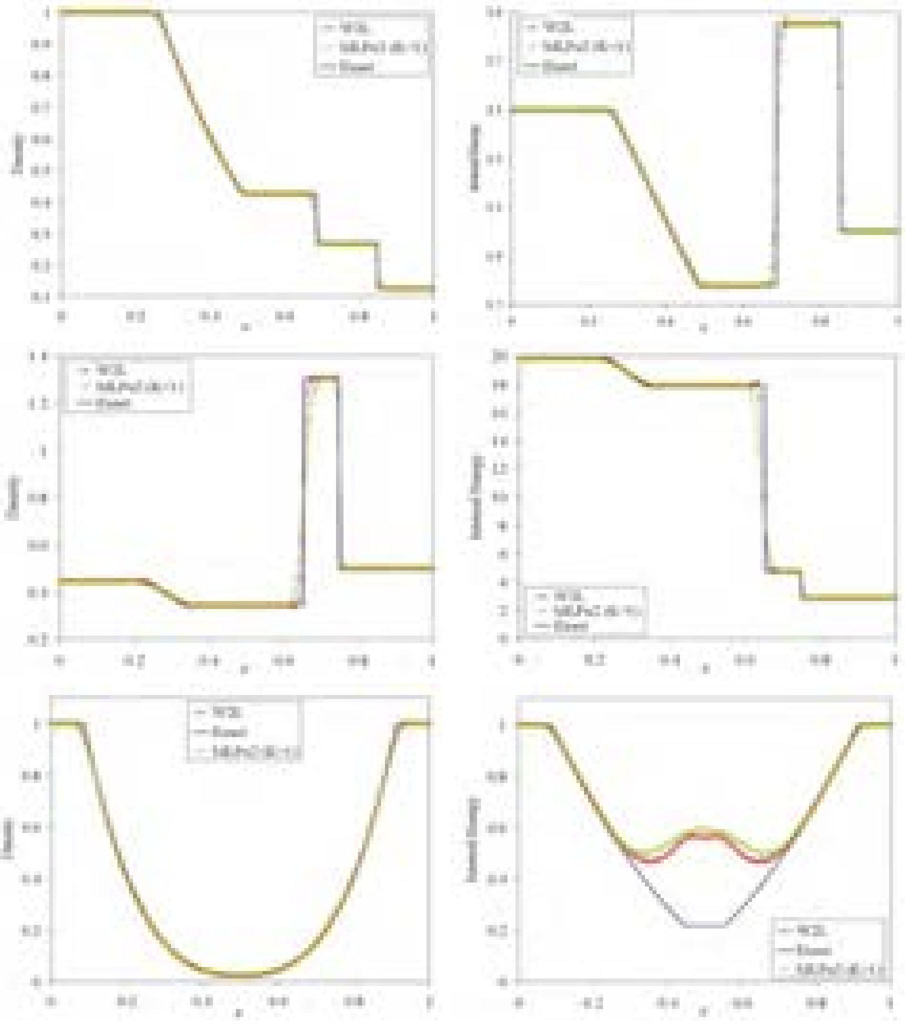


FIGURE 5. Shock tube problems: Sod problem (top), Lax problem (middle) and supersonic expansion problem (bottom)

4.3. **Transonic flow around NACA 0012 airfoil.** The case of transonic flow around NACA 0012 airfoil is considered here with Mach number number 0.8 and

$\alpha = 1.25°$ as to obtain a convergent solution. The number of triangular cells used was $N = 6,492$ with 200 surface grid points. The grid and convergence history obtained using the HLLC solver and reconstruction on the primitive variables are shown in Fig. 6 while Fig. 7 shows a comparison of the surface pressure coefficient between the CCFVw2L and MLPu2 schemes.

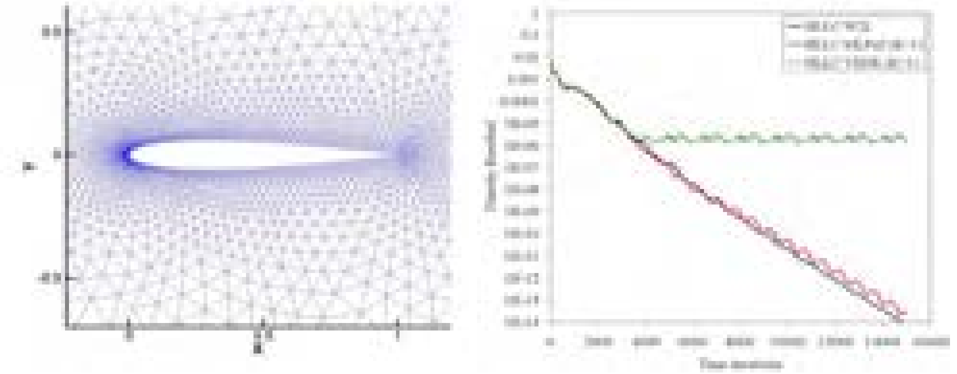


FIGURE 6. Grid distribution and convergence history (NACA0012 airfoil)

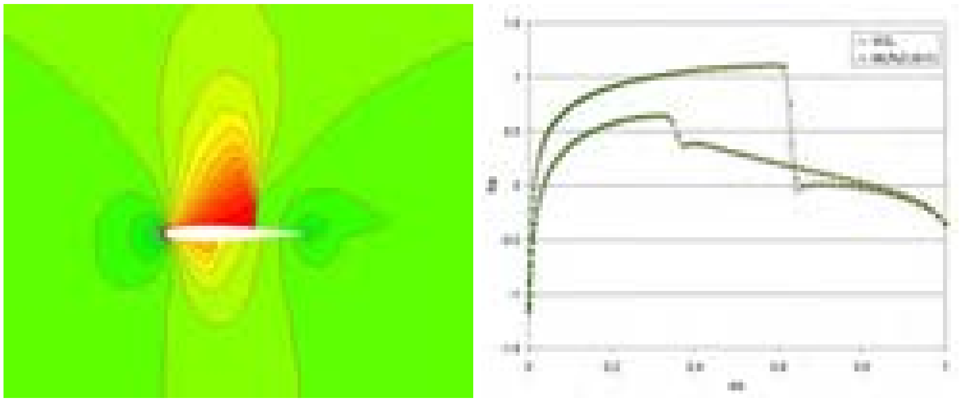


FIGURE 7. Mach contours and surface pressure over NACA0012 airfoil

REFERENCES

[1] S. Clain and V. Clauzon, *Monoslope and multislope MUSCL methods for unstructured meshes*, J. Comp. Phys., **229** (2010), 3745–3776.

[2] A. I. Delis, I.K. Nikolos and M. Kazolea, *Performance and comparizon of cell-centered and node centered unstructured finite volume discretizations for shallow water free surface flows*, Archives of Computational Methods in Engineering, **18** (2011), 57–108.

[3] A. I. Delis and I.K. Nikolos, *A novel multi-dimensional solution reconstruction and edge-based limiting procedure for unstructured cell-centered finite volumes with application to shallow water dynamics*, Int. J. Numer. Meth. Fluids, **71** (2013), 584-633.

[4] M.E. Hubbard, *Multi-dimensional slope limiters for MUSCL-type finite volume schemes on unstructured grids*, J. Comp. Phys., **155** (1999), 54–74.

[5] W. Li, Y.-X. Ren G. Lei and H. Luo, *The multi-dimensional limiters for solving hyperbolic conservation laws on unstructured grids*, J. Comp. Phys., **230** (2011), 7775-7795.

[6] J.S. Park, S.-H. Yoon and C. Kim, *Multi-dimensional limiting process for hyperbolic conservation laws on unstructured grids*, J. Comp. Phys., **229** (2010), 788-812.

[7] V. Venkatakrishan, *Convergence to steady state of the Euler Equations on unstructured grids with limiters*, J. Comp. Phys., **118** (1995), 120-130.

E-mail address: adelis@science.tuc.gr
E-mail address: jnikolo@dpem.tuc.gr

A STRONGLY COUPLED PDE-ODE SYSTEM MODELING MOVING DENSITY CONSTRAINTS IN TRAFFIC FLOW

Maria Laura Delle Monache and Paola Goatin

Inria Sophia-Antipolis-Méditerranée - EPI OPALE
2004, route des Lucioles - BP 93
06902, Sophia Antipolis Cedex, France

Abstract. We prove the existence of solutions of a coupled PDE-ODE system modeling the interaction of a large slow moving vehicle with the surrounding traffic flow. The model consists in a scalar conservation law with moving density constraint describing traffic evolution coupled with an ODE for the slow vehicle trajectory. The constraint location moves due to the surrounding traffic conditions, which in turn are affected by the presence of the slower vehicle, thus resulting in a strong non-trivial coupling.

1. **Introduction.** A slow moving large vehicle, like a bus or a truck, reduces the road capacity and thus generates a moving bottleneck for the surrounding traffic flow. From the macroscopic point of view this can be modeled by a PDE-ODE coupled system consisting in a scalar conservation law with moving density constraint and an ODE describing the slower vehicle motion, i.e.,

$$\begin{cases} \partial_t \rho + \partial_x f(\rho) = 0, & (t,x) \in \mathbb{R}^+ \times \mathbb{R}, \\ \rho(0,x) = \rho_0(x), & x \in \mathbb{R}, \\ \rho(t, y(t)) \leq \alpha R, & t \in \mathbb{R}^+, \\ \dot{y}(t) = \omega(\rho(t, y(t)+)), & t \in \mathbb{R}^+, \\ y(0) = y_0. \end{cases} \tag{1}$$

Above, $\rho = \rho(t,x) \in [0,R]$ is the scalar conserved quantity representing the mean traffic density, R is the maximal density allowed on the road and the flux function $f : [0,R] \to \mathbb{R}^+$ is a strictly concave function such that $f(0) = f(R) = 0$. It is given by the following flux-density relation

$$f(\rho) = \rho v(\rho),$$

where v is a smooth decreasing function denoting the mean traffic speed and here set to be $v(\rho) = V(1 - \frac{\rho}{R})$, V being the maximal velocity allowed on the road.

The time-dependent variable y denotes the slower vehicle position, which moves with a traffic density dependent speed of the form

$$\omega(\rho) = \begin{cases} V_b & \text{if } \rho \leq \rho^* \doteq R(1 - \frac{V_b}{V}), \\ v(\rho) & \text{otherwise}, \end{cases} \tag{2}$$

2000 *Mathematics Subject Classification.* Primary: 58J45, 35L65; Secondary: 90B20.

Key words and phrases. Conservation laws with constraints, Traffic flow modeling, PDE-ODE coupling, Wave-front tracking, Riemann Problem.

The authors are supported by ERC Starting Grant 2010 under the project *"TRAffic Management by Macroscopic Models"*.

that is, it moves with constant speed $V_b < V$ as long as it is not slowed down by downstream traffic conditions. When this happens, it moves with the mean traffic speed.

Finally, the constant coefficient $\alpha \in \,]0,1[$ gives the reduction rate of the road capacity due to the presence of this large vehicle.

For our analytical purposes, it is not restrictive to assume that $R = V = 1$, so that the model becomes

$$\begin{cases} \partial_t \rho + \partial_x(\rho(1-\rho)) = 0, & (t,x) \in \mathbb{R}^+ \times \mathbb{R}, \\ \rho(0,x) = \rho_0(x), & x \in \mathbb{R}, \\ \rho(t,y(t)) \le \alpha, & t \in \mathbb{R}^+, \\ \dot{y}(t) = \omega(\rho(t,y(t)+)), & t \in \mathbb{R}^+, \\ y(0) = y_0. \end{cases} \tag{3}$$

The above model was introduced in [12] to model the effect of urban transport systems, such as buses, in a road network. Other macroscopic models for moving bottlenecks in road traffic were recently proposed by [7, 14]. Compared to those approaches, the model described by (1) offers a more realistic definition of the slower vehicle speed and a description of its impact on traffic conditions which is simpler to handle both from the analytical and the numerical point of view.

From the analytical point of view, model (1) can be viewed as a generalization to moving constraints of the problem consisting in a scalar conservation law with a (fixed in space) constraint on the flux, introduced and studied in [1, 8, 9]. In the present case, the constraint location moves due to the surrounding traffic conditions, which in turn are modified by the presence of the slower vehicle, thus resulting in a strong non-trivial coupling between the conservation equation and the trajectory of the vehicle.

The study of coupled PDE-ODE systems is not new in the conservation laws framework, we refer the reader to [6, 5, 10, 14]. Nevertheless, the problem posed here is slightly different. On one side, we deal with a strong coupling with the PDE and the ODE affecting each other, unlike [5, 10], where the PDE solution does not depend on the ODE. On the other side, even if the ODE has discontinuous right-hand side, the particular definition of the model allows us to consider classical Carathéodory solutions as in [6, 7, 2, 5] instead of the weaker Filippov's generalized solutions needed in [10, 14].

This paper presents an existence result for solutions of (1) constructed by wavefront tracking approximations, as stated by Theorem 3.2 in Section 3. Details on the proofs can be found in [11].

2. **The Riemann problem with moving density constraint.**

Consider (3) with the particular choice

$$y_0 = 0 \quad \text{and} \quad \rho_0(x) = \begin{cases} \rho_L & \text{if } x < 0, \\ \rho_R & \text{if } x > 0. \end{cases} \tag{4}$$

We aim at defining a Riemann solver for the conservation law with moving density constraint. Therefore we consider the following Riemann problem

$$\begin{cases} \partial_t \rho + \partial_x f(\rho) = 0, \\ \rho(0,x) = \begin{cases} \rho_L & \text{if } x < 0, \\ \rho_R & \text{if } x > 0, \end{cases} \end{cases} \tag{5}$$

under the constraint

$$\rho(t, V_b t) \leq \alpha. \tag{6}$$

Let $f_\alpha : [0, \alpha] \to \mathbb{R}^+$ be the function describing the constrained flow at $x = y(t)$, i.e.,

$$f_\alpha(\rho) = \rho\left(1 - \frac{\rho}{\alpha}\right),$$

and $\rho_\alpha \in \,]0, \alpha/2[$ such that $f'(\rho_\alpha) = V_b$, i.e.,

$$\rho_\alpha = \frac{\alpha}{2}(1 - V_b).$$

Problem (5), (6) can be recasted in the framework of conservation laws with flux constraint studied in [1, 9]. Rewriting the equations in the bus reference frame (setting $X = x - V_b t$), we get

$$\begin{cases} \partial_t \rho + \partial_X \left(f(\rho) - V_b \rho\right) = 0, \\ \rho(0, X) = \begin{cases} \rho_L & \text{if } X < 0, \\ \rho_R & \text{if } X > 0, \end{cases} \end{cases} \tag{7}$$

under the constraint

$$\rho(t, 0) \leq \alpha. \tag{8}$$

Remark that solving problem (7), (8) is equivalent to solving (7) under the corresponding constraint on the flux

$$f(\rho(t, 0)) - V_b \rho(t, 0) \leq f_\alpha(\rho_\alpha) - V_b \rho_\alpha \doteq F_\alpha.$$

We are now ready to define the Riemann solver for (3), (4) following [12, §V]. Denote by $\mathcal{R}$ the standard Riemann solver (i.e., without the constraint (6)) for (5), i.e., the (right continuous) map $(t, x) \mapsto \mathcal{R}(\rho_L, \rho_R)(\frac{x}{t})$ is the standard weak entropy solution to (5). Moreover, let $\check{\rho}_\alpha$ and $\hat{\rho}_\alpha$, with $\check{\rho}_\alpha \leq \hat{\rho}_\alpha$, be the intersections of the flux function $f(\rho)$ and the line $f_\alpha(\rho_\alpha) + V_b(\rho - \rho_\alpha)$ (see Figure 1).

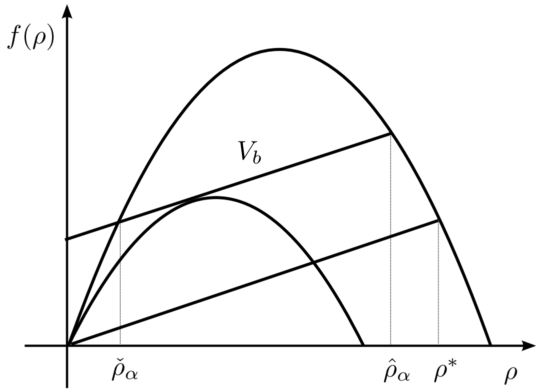


FIGURE 1. Flux function

Definition 2.1. The constrained Riemann solver $\mathcal{R}^\alpha$ for (3), (4) is defined as follows.

1. If $f(\mathcal{R}(\rho_L,\rho_R)(V_b)) > F_\alpha + V_b\mathcal{R}(\rho_L,\rho_R)(V_b)$, then
$$\mathcal{R}^\alpha(\rho_L,\rho_R)(x) = \begin{cases} \mathcal{R}(\rho_L,\hat\rho_\alpha) & \text{if } x < V_b t, \\ \mathcal{R}(\check\rho_\alpha,\rho_R) & \text{if } x \geq V_b t, \end{cases} \quad \text{and} \quad y(t) = V_b t.$$
2. If $V_b\mathcal{R}(\rho_L,\rho_R)(V_b) \leq f(\mathcal{R}(\rho_L,\rho_R)(V_b)) \leq F_\alpha + V_b\mathcal{R}(\rho_L,\rho_R)(V_b)$, then
$$\mathcal{R}^\alpha(\rho_L,\rho_R) = \mathcal{R}(\rho_L,\rho_R) \quad \text{and} \quad y(t) = V_b t.$$
3. If $f(\mathcal{R}(\rho_L,\rho_R)(V_b)) < V_b\mathcal{R}(\rho_L,\rho_R)(V_b)$, then
$$\mathcal{R}^\alpha(\rho_L,\rho_R) = \mathcal{R}(\rho_L,\rho_R) \quad \text{and} \quad y(t) = v(\rho_R)t.$$

Note that, when the constraint is enforced (point 1. in the above definition), a nonclassical shock arises, which satisfies the Rankine-Hugoniot condition but violates the Lax entropy condition.

Remark 1. The above definition is well-posed even if the classical Riemann solution $\mathcal{R}(\rho_L,\rho_R)(x/t)$ displays a shock at $x = V_b t$. In fact, due to Rankine-Hugoniot equation, we have
$$f(\rho_L) = f(\rho_R) + V_b(\rho_L - \rho_R)$$
and hence
$$f(\rho_L) > f_\alpha(\rho_\alpha) + V_b(\rho_L - \rho_\alpha) \quad \Longleftrightarrow \quad f(\rho_R) > f_\alpha(\rho_\alpha) + V_b(\rho_R - \rho_\alpha).$$

Remark 2. The density constraint $\rho(t,y(t)) \leq \alpha$ does not appear explicitly in Definition 2.1, and in the following Definition 3.1. It is handled by the corresponding condition on the flux
$$f(\rho(t,y(t))) - \omega(\rho(t,y(t)))\rho(t,y(t)) \leq F_\alpha. \tag{9}$$
The corresponding density on the reduced roadway at $x = y(t)$ is found taking the solution to the equation
$$f(\rho_y) + \omega(\rho_y)(\rho - \rho_y) = \rho\left(1 - \frac{\rho}{\alpha}\right),$$
closer to $\rho_y \doteq \rho(t,y(t)))$.

3. The Cauchy problem: Existence of solutions.

The aim of this section is to study the existence of the solutions of problem (1), (3). A bus travels along a road modeled by
$$\begin{cases} \partial_t\rho + \partial_x(\rho(1-\rho)) = 0, \\ \rho(0,x) = \rho_0(x), \\ \rho(t,y(t)) \leq \alpha. \end{cases} \tag{10}$$
The bus influences the traffic along the road but it is also influenced by it. The bus position $y = y(t)$ then solves
$$\begin{cases} \dot y(t) = \omega(\rho(t,y(t)+)), \\ y(0) = y_0. \end{cases} \tag{11}$$
Solutions to (11) will be intended in Carathéodory sense, i.e., as absolutely continuous functions which satisfy (11) for a.e. $t \geq 0$. In our setting, due to the strong PDE-ODE coupling, we will prove existence of both solutions to (10) and (11) at the same time. We start giving our definition of solution.

Definition 3.1. A couple $(\rho,y) \in \mathcal{C}^0\left(\mathbb{R}^+;\mathbf{L^1} \cap \mathrm{BV}(\mathbb{R};[0,R])\right) \times \mathbf{W^{1,1}}(\mathbb{R}^+;\mathbb{R})$ is a solution to (3) if

1. ρ is a weak solution of the conservation law, i.e., for all $\varphi \in \mathcal{C}^1_c(\mathbb{R}^2;\mathbb{R})$

$$\int_{\mathbb{R}^+}\int_{\mathbb{R}} (\rho\partial_t\varphi + f(\rho)\partial_x\varphi)\, dx\; dt + \int_{\mathbb{R}} \rho_0(x)\varphi(0,x)\; dx = 0\;; \tag{12a}$$

moreover, ρ satisfies the Kružhkov entropy conditions [13] on $(\mathbb{R}^+\times\mathbb{R})\setminus\{(t,y(t))\colon t\in\mathbb{R}^+\}$, i.e., for every $k\in[0,1]$ and for all $\varphi\in\mathcal{C}^1_c(\mathbb{R}^2;\mathbb{R}^+)$ and $\varphi(t,y(t))=0$, $t>0$,

$$\begin{aligned}\int_{\mathbb{R}^+}\int_{\mathbb{R}} (|\rho-k|\partial_t\varphi + \operatorname{sgn}(\rho-k)\,(f(\rho)-f(k))\,\partial_x\varphi)\, dx\; dt\\ + \int_{\mathbb{R}} |\rho_0-k|\varphi(0,x)\; dx \geq 0\;;\end{aligned} \tag{12b}$$

2. y is a Carathéodory solution of the ODE, i.e., for a.e. $t\in\mathbb{R}^+$

$$y(t) = y_0 + \int_0^t \omega(\rho(s,y(s)+))\; ds\;; \tag{12c}$$

3. the constraint is satisfied, in the sense that for a.e. $t\in\mathbb{R}^+$

$$\lim_{x\to y(t)\pm} (f(\rho)-\omega(\rho)\rho)\,(t,x) \leq F_\alpha. \tag{12d}$$

Remark that the above traces exist because $\rho(t,\cdot)\in \mathrm{BV}(\mathbb{R})$ for all $t\in\mathbb{R}^+$.

Remark 3. Our choice of a Carathéodory solution of the ODE is motivated by the particular bus velocity defined by (2). With this choice it is not possible for a bus to end up trapped in a queue unless its speed is equal to V_b, in which case $\omega(\rho(t,y(t)+)) = \omega(\rho(t,y(t)-)) = V_b$. Therefore Carathéodory solutions are always well defined.

We are now ready to state the main result of the paper.

Theorem 3.2. *Let $\rho_0 \in BV(\mathbb{R};[0,R])$, then the problem (1) admits a solution in the sense of Definition 3.1.*

The rest of the section is devoted to the proof of Theorem 3.2. In particular, we will construct a sequence of approximate solutions via the wave-front tracking method, and prove its convergence. Finally, we will check that the limit functions satisfy conditions (12a)-(12d) of Definition 3.1.

3.1. **Wave-front tracking.** To construct piecewise constant approximate solutions, we adapt the standard wave-front tracking method, see for example [3, §6]. Fix a positive $n\in\mathbb{N}$, $n>0$ and introduce in $[0,1]$ the mesh $\mathcal{M}_n = \{\rho^n_i\}_{i=0}^{2^n}$ defined by

$$\mathcal{M}_n = \left(2^{-n}\mathbb{N}\cap[0,1]\right)\cup\{\check\rho_\alpha,\hat\rho_\alpha\}.$$

In order to include the critical points $\check\rho_\alpha,\hat\rho_\alpha$, we modify the above mesh as follows:

- if $\min_i|\check\rho_\alpha-\rho^n_i| = 2^{-n-1}$, then we simply add the new point to the mesh:
$$\widetilde{\mathcal{M}}_n = \mathcal{M}_n\cup\{\check\rho_\alpha\};$$
- if $|\check\rho_\alpha-\rho^n_l| = \min_i|\check\rho_\alpha-\rho^n_i| < 2^{-n-1}$, then we replace ρ^n_l by $\check\rho_\alpha$:
$$\widetilde{\mathcal{M}}_n = \mathcal{M}_n\cup\{\check\rho_\alpha\}\setminus\{\rho^n_l\};$$
- we perform the same operations for $\hat\rho_\alpha$.

In this way, the distance between two points of the mesh $\widetilde{\mathcal{M}}_n = \{\tilde{\rho}_i^n\}_{i=0}^{2^n}$ satisfies the lower bound $\left|\tilde{\rho}_i^n - \tilde{\rho}_j^n\right| \geq 2^{-n-1}$.

Let f^n be the piecewise linear function which coincides with f on $\widetilde{\mathcal{M}}_n$, and let ρ_0^n be a piecewise constant function defined by

$$\rho_0^n = \sum_{j\in\mathbb{Z}} \rho_{0,j}^n \; \chi_{]x_{j-1},x_j]} \quad \text{with } \rho_{0,j}^n \in \widetilde{\mathcal{M}}_n,$$

which approximates ρ_0 in the sense of the strong $\mathbf{L^1}$ topology, that is

$$\lim_{n\to\infty} \|\rho_0^n - \rho_0\|_{\mathbf{L^1}(\mathbb{R})} = 0,$$

and such that $\mathrm{TV}(\rho_0^n) \leq \mathrm{TV}(\rho_0)$. Above, we set $x_0 = y_0$.

For small times $t > 0$, a piecewise approximate solution (ρ^n, y_n) to (3) is constructed piecing together the solutions to the Riemann problems

$$\left\{\begin{array}{l} \partial_t\rho + \partial_x\left(f^n(\rho)\right) = 0, \\ \rho(0,x) = \left\{\begin{array}{ll} \rho_0 & \text{if } x < y_0, \\ \rho_1 & \text{if } x > y_0, \end{array}\right. \\ \rho(t, y_n(t)) \leq \alpha, \end{array}\right. \qquad \left\{\begin{array}{l} \partial_t\rho + \partial_x\left(f^n(\rho)\right) = 0, \\ \rho(0,x) = \left\{\begin{array}{ll} \rho_j & \text{if } x < x_j, \\ \rho_{j+1} & \text{if } x > x_j, \end{array}\right. \\ j \neq 0, \end{array}\right. \tag{13}$$

where y_n satisfies

$$\left\{\begin{array}{l} \dot{y}_n(t) = \omega(\rho^n(t, y_n(t)+)), \\ y_n(0) = y_0. \end{array}\right. \tag{14}$$

Note that the solutions to the constrained Riemann problem in (13), left, coupled with (14), is constructed by means of $\mathcal{R}^\alpha$, see Definition 2.1.

The approximate solution ρ^n constructed above can be prolonged up to the first time $\bar{t} > 0$, where two discontinuities collide, or a discontinuity hits the bus trajectory. In both cases, a new Riemann problem arises and its solution, obtained in the former case with $\mathcal{R}$ and in the latter case with the constrained Riemann solver $\mathcal{R}^\alpha$, allows to extend ρ^n further in time.

3.2. **Bounds on the total variation.**

Given an approximate solution $\rho^n = \rho^n(t,\cdot)$ constructed by the wave-front tracking method, we define the Glimm type functional

$$\Upsilon(t) = \Upsilon(\rho^n(t,\cdot)) = \mathrm{TV}(\rho^n) + \gamma(\rho^n) = \sum_j \left|\rho_{j+1}^n - \rho_j^n\right| + \gamma(\rho^n), \tag{15}$$

where γ is given by

$$\gamma(\rho^n) = \gamma(\rho^n(t)) = \left\{\begin{array}{ll} 0 & \text{if } \rho^n(t, y_n(t)-) = \hat{\rho}_\alpha,\ \rho^n(t, y_n(t)+) = \check{\rho}_\alpha \\ 2|\hat{\rho}_\alpha - \check{\rho}_\alpha| & \text{otherwise.} \end{array}\right. \tag{16}$$

The value of γ is chosen to have the following uniform bound on Υ.

Lemma 3.3. *For any $n \in \mathbb{N}$, the map $t \mapsto \Upsilon(t) = \Upsilon(\rho^n(t,\cdot))$ at any interaction either decreases by at least 2^{-n}, or remains constant and the number of waves does not increase.*

Lemma 3.3 in particular implies that the wave-front tracking procedure can be prolonged to any time $T > 0$.

Proof. In order to obtain a uniform bound on the total variation, we consider different types of interactions separately. In particular, we assume that at any interaction time $t = \bar{t}$ either two waves interact or a single wave hits the bus trajectory. In each case an estimation for $\Upsilon(t)$ was found. For details, we refer the reader to [11]. □

3.3. **Convergence of approximate solutions.** In this section we prove that the limit of wave-front tracking approximations provides a solution (ρ, y) of the PDE-ODE model (1) in the sense of Definition 3.1.

We start showing the convergence of the wave-front tracking approximations.

Lemma 3.4. *Let ρ^n and y_n, $n \in \mathbb{N}$, be the wave-front tracking approximations to (1) constructed as detailed in Section 3.1, and assume $TV(\rho_0) \leq C$ be bounded, $0 \leq \rho_0 \leq 1$. Then, up to a subsequence, we have the following convergences*

$$\rho^n \to \rho \qquad \text{in } \mathbf{L^1_{loc}}(\mathbb{R}^+ \times \mathbb{R}); \tag{17a}$$

$$y_n(\cdot) \to y(\cdot) \qquad \text{in } \mathbf{L}^\infty([0,T]), \text{ for all } T > 0; \tag{17b}$$

$$\dot{y}_n(\cdot) \to \dot{y}(\cdot) \qquad \text{in } \mathbf{L^1}([0,T]), \text{ for all } T > 0; \tag{17c}$$

for some $\rho \in \mathcal{C}^0\left(\mathbb{R}^+; \mathbf{L^1} \cap BV(\mathbb{R})\right)$ and $y \in \mathbf{W^{1,1}}(\mathbb{R}^+)$.

Proof. Lemma 3.3 gives a uniform bound on the total variation of approximate solutions. Thus, we have $\mathrm{TV}(\rho^n(t,\cdot)) \leq \Upsilon(t) \leq \Upsilon(0)$. A standard procedure based on Helly's Theorem (see [3, Theorem 2.4]) ensures the existence of a subsequence converging to some function $\rho \in \mathcal{C}^0\left(\mathbb{R}^+; \mathbf{L^1} \cap \mathrm{BV}(\mathbb{R})\right)$, proving (17a).

Since $|\dot{y}_n(t)| \leq V_b$, the sequence $\{y_n\}$ is uniformly bounded and equicontinuous on any compact interval $[0,T]$. By Ascoli-Arzelà Theorem, there exists a subsequence converging uniformly, giving (17b). In order to prove (17c), we have to show that $\mathrm{TV}\left(\dot{y}_n; [0,T]\right)$ is uniformly bounded. In fact, the analysis performed in Section 3.2 shows that $\dot{y}_n$ can change only at interactions with waves coming from its right. We can estimate the speed variation at interactions times $\bar{t}$ by the size of the interacting front:

$$|\dot{y}_n(\bar{t}+) - \dot{y}_n(\bar{t}-)| = |\omega(\rho_l) - \omega(\rho_r)| \leq |\rho_l - \rho_r|.$$

In particular, $\dot{y}_n$ is non-increasing at interactions with shock fronts and non-decreasing at interactions with rarefaction fronts, which must be originated at $t = 0$. In fact, the analysis performed in Section 3.2 shows that no new rarefaction front can arise at interactions. Therefore,

$$\mathrm{TV}\left(\dot{y}_n; [0,T]\right) \leq 2\,\mathrm{PV}\left(\dot{y}_n; [0,T]\right) + \|\dot{y}_n\|_{\mathbf{L}^\infty([0,T])} \leq 2\,\mathrm{TV}(\rho_0) + V_b$$

is uniformly bounded. Above, $\mathrm{PV}\left(\dot{y}_n; [0,T]\right)$ denotes the positive variation of $\dot{y}_n$, i.e., the total amount of positive jumps in the interval $[0,T]$. □

3.3.1. *Proof of* (12a) *and* (12b). Since ρ^n converge strongly to ρ in $\mathbf{L^1_{loc}}(\mathbb{R}^+ \times \mathbb{R})$, it is straightforward to pass to the limit in the weak formulation of the conservation law, proving that the limit function ρ satisfies (12a). Kružhkov entropy condition (12b) can be recovered in the same way.

3.3.2. *Proof of* (12c). We prove that

$$\lim_{n\to\infty} \rho^n(t, y_n(t)+) = \rho^+(t) = \rho(t, y(t)+) \quad \text{for a.e. } t \in \mathbb{R}^+. \tag{18}$$

At this purpose, we use the weak convergence of measure in [4, Lemma 15] and we proceed like in [4, §4], for the details see [11].

Combining (17c) and (18) we get $\dot{y}(t) = \omega(\rho(t, y(t)+))$ for a.e. $t > 0$.

3.3.3. *Proof of* (12d). In order to verify that the limit solutions satisfy the constraint (12d), we can use directly the convergence result (18).

REFERENCES

[1] B. Andreianov, P. Goatin and N. Seguin, *Finite volume scheme for locally constrained conservation laws*, Numer. Math, **115** (2010), 609–645.

[2] A. Bressan, *Unique solutions for a class of discontinuous differential equations*, Proc. Amer. Math. Soc., **104** (1988), 772–778.

[3] A. Bressan, "Hyperbolic systems of conservation laws," 1^{st} edition, Oxford University Press, Oxford, 2000.

[4] A. Bressan and P. G. LeFloch, *Structural stability and regularity of entropy solutions to hyperbolic systems of conservation laws*, Indiana univ. Math. J. , **48** (1999), 43–84.

[5] A. Bressan and W. Shen, *Uniqueness for discontinuous ODE and conservation laws*, Nonlinear Anal. , **34** (1998), 637–652.

[6] R. Borsche, R. M. Colombo and M. Garavello, *On the coupling of systems of hyperbolic conservation laws with ordinary differential equations*, Nonlinearity, **23** (2010), 2749–2770.

[7] R. Borsche, R. M. Colombo and M. Garavello, *Mixed systems: ODEs - balance laws*, J. Differential Equations, **252** (2012), 2311–2338.

[8] C. Chalons, P. Goatin and N. Seguin, *General constrained conservation laws. Application to pedestrian flow modeling.*, preprint http://hal.inria.fr/hal-00713609.

[9] R. M. Colombo and P. Goatin, *A well posed conservation law with a variable unilateral constraint*, J. Differential Equations, **234** (2007), 654–675.

[10] R. M. Colombo and A. Marson, *A Hölder continuous ODE related to traffic flow*, Proc. Roy. Soc. Edinburgh Sect. A, **133** (2003), 759–772.

[11] M. L. Delle Monache and P. Goatin, *Scalar conservation laws with moving density constraints arising in traffic flow modeling*, Inria Research Report no. 8119 (2012)

[12] Florence Giorgi, "Prise en compte des transports en commune de surface dans la modélisation macroscopique de l'écoulement du trafic," Ph.D thesis, Insitut National des Sciences Appliquèes de Lyon, 2002.

[13] S. N. Kružhkov, *First order quasilinear equations with several independent variables*, Math. Sb. (N. S.), **81** (1970), 228–255.

[14] C. Lattanzio, A. Maurizi and B. Piccoli, *Moving bottleneck in car traffic flow: a PDE-ODE coupled model*, SIAM J. Math. Anal., **43** (2011), 50–67.

E-mail address: maria-laura.delle_monache@inria.fr
E-mail address: paola.goatin@inria.fr

A PRIORI ANALYSIS OF ASYMPTOTIC PRESERVING SCHEMES WITH THE MODIFIED EQUATION

Bruno Després

Laboratoire Jacques-Louis Lions, Université Pierre et Marie Curie
5252 Paris Cedex 05, France

Christophe Buet and Emmanuel Franck

CEA, DAM, DIF
91297 Arpajon Cedex, France

Abstract. We analyze the modified equation of a model hyperbolic heat equation, which can be used as a guide for numerical methods. The main result is a uniform estimate of accuracy for a particular form of the modified equation.

1. **Introduction.** Our model problem is the hyperbolic heat equation or Cattaneo's equation in dimension one $x \in \mathbb{R}$

$$\begin{cases} \partial_t u_\varepsilon + \dfrac{1}{\varepsilon}\partial_x v_\varepsilon = 0, \\ \partial_t v_\varepsilon + \dfrac{1}{\varepsilon}\partial_x u_\varepsilon = -\dfrac{\sigma}{\varepsilon^2} v_\varepsilon. \end{cases} \tag{1}$$

The scaling parameter is $0 < \varepsilon \leq 1$ which can takes value arbitrarily in $]0, 1]$. The other coefficient $\sigma > 0$ is constant: the case with a non constant σ will be examined in the last part of this work. Hilbert expansion of all quantities with respect to ε (that is $f = f_0 + \varepsilon f_1 + O(\varepsilon^2)$) shows that the limit equation writes

$$\partial_t u - \frac{1}{\sigma}\partial_{xx} u = 0. \tag{2}$$

Indeed the second equation in (1) can be rewritten formerly as $v_\varepsilon = -\frac{\varepsilon}{\sigma}\partial_x u_\varepsilon + O(\varepsilon^2)$. Plugged in the first equation it yields $\partial_t u_\varepsilon - \frac{1}{\sigma}\partial_{xx} u_\varepsilon = O(\varepsilon)$ which is further simplified in (2). In other words the diffusion equation (2) is the **asymptotic limit** of the system (1).

At the numerical level, Asymptotic Preserving (AP) techniques [1, 2] are useful to discretize such problems with accuracy uniform with respect to ε. This family of methods and schemes is particularly appealing for the numerical discretization of physical problems with very different scales. It has motivated recent contributions [3, 4], see also [5] in a different context. A general observation is that the numerical structure of Asymptotic Preserving schemes is not so simple to understand. As a result comparisons between different methods is also difficult.

That is why we desire to develop, for the model problem (1), an *a priori* understanding of asymptotic preserving techniques. Here *a priori* means that we do not want to rely on the standard method, that is a) development of a new numerical method, and b) analysis of the pros and cons of the method. On the contrary we

The first author is supported by CEA contract.

desire to understand *a priori* what are the mathematical properties that the desired numerical method should satisfy. Following [2], we use the modified equation which is a natural and powerful tool to develop such a priori understanding. The main result of this work is a new estimate of accuracy in theorem 3.1. This estimate is uniform with respect to ε. It gives insights into the mathematical and numerical structure of the Gosse-Toscani scheme [1] and the Jin-Levermore scheme [2].

2. **The modified equation.**

The starting point is the modified equation

$$\begin{cases} \partial_t u_{\varepsilon,\alpha} + \dfrac{1}{\varepsilon}\left(\partial_x v_{\varepsilon,\alpha} - \alpha \partial_{xx} u_{\varepsilon,\alpha}\right) = 0, \\ \partial_t v_{\varepsilon,\alpha} + \dfrac{1}{\varepsilon}\left(\partial_x u_{\varepsilon,\alpha} - \alpha \partial_{xx} u_{\varepsilon,\alpha}\right) = -\dfrac{\sigma}{\varepsilon^2} v_{\varepsilon,\alpha}, \end{cases} \tag{3}$$

where $\alpha \approx \frac{\Delta x}{2}$ is the coefficient of the numerical viscosity characteristic of first order Finite Volume techniques. See [2, 3] for a justification. Since this system admits two different small parameters ε and α, the behavior of the the solutions depends on the competition between these parameters. Hilbert expansion of all quantities with respect to ε shows that the limit equation, the parameter α being kept constant, is

$$\partial_t u_0 - \left(\frac{1}{\sigma} + \frac{\alpha}{\varepsilon}\right) \partial_{xx} u_0 = 0 \tag{4}$$

which is of course non correct in the regime $\varepsilon << \alpha$ because the diffusion coefficient has been modified: $\frac{1}{\sigma} << \frac{1}{\sigma} + \frac{\alpha}{\varepsilon}$. Therefore something has to be done in order to preserve the correct asymptotic limit.

2.1. **A modified equation of the first kind.**

Starting from this consideration a very natural idea is to modify the first equation and to replace (3) with

$$\begin{cases} \partial_t \overline{u}_{\varepsilon,\alpha} + \dfrac{M}{\varepsilon}\left(\partial_x \overline{v}_{\varepsilon,\alpha} - \alpha \partial_{xx} \overline{u}_{\varepsilon,\alpha}\right) = 0, \\ \partial_t \overline{v}_{\varepsilon,\alpha} + \dfrac{1}{\varepsilon}\left(\partial_x \overline{u}_{\varepsilon,\alpha} - \alpha \partial_{xx} \overline{v}_{\varepsilon,\alpha}\right) = -\dfrac{\sigma}{\varepsilon^2} \overline{v}_{\varepsilon,\alpha}, \end{cases} \tag{5}$$

where the Magic coefficient is the ratio of the true viscosity over the spurious one

$$M = \frac{\frac{1}{\sigma}}{\frac{1}{\sigma} + \frac{\alpha}{\varepsilon}} = \frac{\varepsilon}{\varepsilon + \sigma\alpha} \in]0, 1].$$

This system is obtained as the modified equation of the Jin-Levermore scheme [2]. The Hilbert expansion, $\overline{v}_{\varepsilon,\alpha} = \overline{v}^0_\alpha + \varepsilon \overline{v}^1_{\varepsilon,\alpha} + \dots$ and so on, yields the limit equation

$$\partial_t \overline{u}^0_\alpha - \frac{1}{\sigma} \partial_{xx} \overline{u}^0_\alpha = O(\varepsilon)$$

which is now correct. Nevertheless modifying (3) into (5) alters the whole mathematical structure of the original system. For example the energy identity attached to (5) writes

$$\partial_t \frac{\overline{u}^2_{\varepsilon,\alpha} + M\overline{v}^2_{\varepsilon,\alpha}}{2} + \partial_x \frac{M \overline{u}_{\varepsilon,\alpha} \overline{v}_{\varepsilon,\alpha}}{\varepsilon} - \partial_{xx} \frac{M\alpha\left(\overline{u}^2_{\varepsilon,\alpha} + \overline{v}^2_{\varepsilon,\alpha}\right)}{2\varepsilon}$$
$$= -\frac{\sigma M}{\varepsilon^2} \overline{v}^2_{\varepsilon,\alpha} - \frac{M\alpha}{\varepsilon}\left(\left(\partial_x \overline{u}_{\varepsilon,\alpha}\right)^2 + \left(\partial_x \overline{v}_{\varepsilon,\alpha}\right)^2\right) \le 0. \tag{6}$$

It yields

$$\int_{\mathbb{R}} \frac{\overline{u}_{\varepsilon,\alpha}(t)^2 + M\overline{v}_{\varepsilon,\alpha}(t)^2}{2} dx \le \int_{\mathbb{R}} \frac{\overline{u}_{\varepsilon,\alpha}(0)^2 + M\overline{v}_{\varepsilon,\alpha}(0)^2}{2} dx. \tag{7}$$

This inequality shows the L^2 stability of the system. However the L^2 control on the variable v may be quite weak in regimes such that $M \to 0^+$ which are precisely our concern.

2.2. **A modified equation of the second kind.**

With this regard the following system is much better

$$\begin{cases} \partial_t \widehat{u}_{\alpha,\varepsilon} + \dfrac{M}{\varepsilon}\left(\partial_x \widehat{v}_{\alpha,\varepsilon} - \alpha \partial_{xx} \widehat{u}_{\alpha,\varepsilon}\right) = 0, \\ \partial_t \widehat{v}_{\alpha,\varepsilon} + \dfrac{M}{\varepsilon}\left(\partial_x \widehat{u}_{\alpha,\varepsilon} - \alpha \partial_{xx} \widehat{v}_{\alpha,\varepsilon}\right) = -\dfrac{\sigma M}{\varepsilon^2} \widehat{v}_{\alpha,\varepsilon}. \end{cases} \tag{8}$$

This system is obtained as the modified equation of the Gosse-Toscani scheme [1]. The formal asymptotic diffusion equation is still the correct one. Evident manipulations show that the energy identity writes now

$$\int_{\mathbb{R}} \frac{\widehat{u}_{\alpha,\varepsilon}(t)^2 + \widehat{v}_{\alpha,\varepsilon}(t)^2}{2} dx \le \int_{\mathbb{R}} \frac{\widehat{u}_{\alpha,\varepsilon}(0)^2 + \widehat{v}_{\alpha,\varepsilon}(0)^2}{2} dx \tag{9}$$

which is better balanced than (7) in the sense that both variables have the same weights.

In other words, the modified equation of the second kind has the correct asymptotic limit and has a correct energy law. The remaining part of this work is dedicated to a proof of a uniform accuracy estimate for the second kind modified equation.

3. **Main result.**

To simplify the analysis, we will consider well prepared initial data for the system (1). Such well prepared data are easy to construct: for example one first takes a sufficiently smooth data $u_\varepsilon(t=0) = u_0 \in H^3(\mathbb{R})$. Then one picks $v_\varepsilon(t=0) = v_0 = -\frac{\varepsilon}{\sigma}\partial_x u_0$. Since the system (1) satisfies an energy identity similar to (9), one first has that $u(t)$, $v(t)$, $\partial_x u(t)$, $\partial_x v(t)$ are uniformly bounded in $L^2(\mathbb{R})$. One also has that $\partial_t u(t=0) = -\frac{1}{\varepsilon}\partial_x v_0 = \frac{1}{\sigma}\partial_{xx} u_0$ and $\partial_t v(t=0) = 0$ are bounded in $L^2(\mathbb{R})$: it implies that $\partial_t u(t)$, $\partial_t v(t)$ are also uniformly bounded in $L^2(\mathbb{R})$. More useful estimates satisfied by well prepared data are the following.

Proposition 1. *Assume that $u_0 \in H^3(\mathbb{R})$. Then the solution of (1) satisfies*

$$\|\partial_{tx} u_\varepsilon\|_{L^\infty((0,T):L^2(\mathbb{R}))} \le \frac{1}{\sigma}\|\partial_{xxx} u_0\|_{L^2(\mathbb{R})} \tag{10}$$

and

$$\|\partial_t v_\varepsilon\|_{L^2((0,T)\times\mathbb{R})} \le \frac{\varepsilon}{\sigma}\|\partial_{xx} u_0\|_{L^2(\mathbb{R})}. \tag{11}$$

- Set $w = \partial_{tx} u_\varepsilon$ and $z = \partial_{tx} v_\varepsilon$. Since the pair (w,z) satisfies the same homogeneous system (1) with the initial data (w_0, z_0)

$$w_0 = (\partial_{tx} u_\varepsilon)(t=0) = -\frac{\partial_{xx} v_0}{\varepsilon} = \frac{1}{\sigma}\partial_{xxx} u_0,$$

$$z_0 = (\partial_{tx} v_\varepsilon)(t=0) = -\partial_x\left(\frac{1}{\varepsilon}\partial_x u_0 + \frac{\sigma}{\varepsilon^2} v_0\right) = 0,$$

the first inequality (10) naturally holds using the L^2 stability property (9).

- Now we set $w = \partial_t u_\varepsilon$ and $z = \partial_t v_\varepsilon$ Since the pair (w,z) satisfies the same homogeneous system (1) with the initial data (w_0, z_0)

$$w_0 = (\partial_t u_\varepsilon)(t=0) = -\frac{\partial_x v_0}{\varepsilon} = \frac{1}{\sigma}\partial_{xx} u_0,$$

and $z_0 = (\partial_t v_\varepsilon)(t=0) = 0$, one obtains after integration

$$\int_{\mathbb{R}} \frac{w(t)^2 + z(t)^2}{2} + \frac{\sigma}{\varepsilon^2}\int_0^T \int_{\mathbb{R}} z(s)^2 ds = \int_{\mathbb{R}} \frac{w_0^2 + z_0^2}{2}.$$

It yields $\|z\|_{L^2((0,T)\times\mathbb{R})} \le \frac{\varepsilon}{\sigma}\|w_0\|_{L^2(\mathbb{R})}$ which ends the proof of (11).

From now on, we use the same initial data for the modified system, that is

$$\widehat{u}_{\alpha,\varepsilon}(t=0) = u_\varepsilon(t=0) = u_0 \text{ and } \widehat{v}_{\alpha,\varepsilon}(t=0) = v_\varepsilon(t=0) = v_0 = -\frac{\varepsilon}{\sigma}\partial_x u_0. \quad (12)$$

Theorem 3.1. *For well prepared data (12), there exists a constant independent of ε and α such that one has the accuracy inequality $\underline{\text{independent of } \varepsilon}$*

$$\|u_\varepsilon(t) - \widehat{u}_{\alpha,\varepsilon}(t)\|_{L^2(\mathbb{R})} + \|v_\varepsilon(t) - \widehat{v}_{\alpha,\varepsilon}(t)\|_{L^2(\mathbb{R})} \le C\alpha, \qquad t \le T. \quad (13)$$

Remark 1. The important fact is that the error is no more spoiled by $\frac{\alpha}{\varepsilon}$ terms. Similar behavior is displayed by numerical schemes, such as the Gosse-Toscani scheme [1], which are compatible with the second kind modified equation, see [3].

Set $e = u_\varepsilon - \widehat{u}_{\alpha,\varepsilon}$ and $f = v_\varepsilon - \widehat{v}_{\alpha,\varepsilon}$. By construction

$$\begin{cases} \partial_t e + \dfrac{M}{\varepsilon}(\partial_x f - \alpha\partial_{xx} e) = r, \\ \partial_t f + \dfrac{M}{\varepsilon}(\partial_x f - \alpha\partial_{xx} f) + \dfrac{\sigma M}{\varepsilon^2} f = s, \end{cases} \quad (14)$$

where the truncation errors r and s are defined by

$$\begin{cases} r = \partial_t u_\varepsilon + \dfrac{M}{\varepsilon}(\partial_x v_\varepsilon - \alpha\partial_{xx} u_\varepsilon) & = (1-M)\partial_t u_\varepsilon - \frac{M\alpha}{\varepsilon}\partial_{xx} u_\varepsilon, \\ s = \partial_t v_\varepsilon + \dfrac{M}{\varepsilon}(\partial_x u_\varepsilon - \alpha\partial_{xx} v_\varepsilon) + \dfrac{\sigma M}{\varepsilon^2} v_\varepsilon & = (1-M)\partial_t v_\varepsilon - \frac{M\alpha}{\varepsilon}\partial_{xx} v_\varepsilon. \end{cases}$$

The second truncation error, s, is naturally small since

$$s = \left[\frac{\partial_t v_\varepsilon}{\varepsilon}\right](1-M)\varepsilon + [M\partial_{tx} u_\varepsilon]\,\alpha$$

Due to proposition 1 each term in square brackets is bounded in $L^2((0,T)\times\mathbb{R})$, and $\varepsilon(1-M) = \frac{\varepsilon}{\varepsilon+\sigma\alpha}\sigma\alpha = O(\alpha)$. Therefore $\|s\|_{L^2((0,T)\times\mathbb{R})} = O(\alpha)$. The first term r needs a little more manipulations. Since $\partial_x u_\varepsilon = -\varepsilon\partial_t u_\varepsilon - \frac{\sigma}{\varepsilon} v_\varepsilon$, one has the identity

$$\partial_{xx} u_\varepsilon = -\varepsilon\partial_{tx} u_\varepsilon - \frac{\sigma}{\varepsilon}\partial_x v_\varepsilon = -\varepsilon\partial_{tx} u_\varepsilon + \sigma\partial_t u_\varepsilon$$

from which we deduce that

$$r = (1 - M - M\frac{\alpha\sigma}{\varepsilon})\partial_t u_\varepsilon + M\alpha\partial_{tx} u_\varepsilon = [M\partial_{tx} u_\varepsilon]\,\alpha$$

since $1 - M - M\frac{\alpha\sigma}{\varepsilon} = 0$ by construction. Once again the term in square brackets is bounded in $L^2((0,T)\times\mathbb{R})$. Therefore $\|r\|_{L^2((0,T)\times\mathbb{R})} = O(\alpha)$.

Next we multiply the first equation in (14) by e, the second by f. Using the L^2 stability, one obtains (using notation such as $\|e,f\|_{L^2(\mathbb{R})} = \sqrt{\|e\|^2_{L^2(\mathbb{R})} + \|f\|^2_{L^2(\mathbb{R})}}$)

$$\frac{d}{dt}\frac{\|e(t),f(t)\|^2_{L^2(\mathbb{R})}}{2} \le \|r(t),s(t)\|_{L^2(\mathbb{R})}\|e(t),f(t)\|_{L^2(\mathbb{R})}$$

from which we deduce that

$$\|e(t),f(t)\|_{L^2(\mathbb{R})} \le \int_0^T \|r(\tau),s(\tau)\|_{L^2(\mathbb{R})}\,d\tau$$

$$\leq \sqrt{T}\left(\int_0^T \|r(\tau), s(\tau)\|^2_{L^2(\mathbb{R})}\, d\tau\right)^{\frac{1}{2}} = \sqrt{T}\,\|r, s\|_{L^2((0,T)\times\mathbb{R})}\,.$$

Since $\|r, s\|_{L^2((0,T)\times\mathbb{R})} = O(\alpha)$ the proof of the claim is finished.

4. **Other modified equations.**

We modify the parameters of the equation or of the scheme to analyze the potential impact of such modifications.

4.1. **Changing the sink.**

The first one is already considered in the seminal work [2]. It is intermediate between the first kind and second kind modified equations. It writes

$$\begin{cases} \partial_t \widehat{u} + \dfrac{M}{\varepsilon}\left(\partial_x \widehat{v} - \alpha \partial_{xx}\widehat{u}\right) = 0, \\ \partial_t \widehat{v} + \dfrac{M}{\varepsilon}\left(\partial_x \widehat{u} - \alpha \partial_{xx}\widehat{v}\right) = -\dfrac{\sigma}{\varepsilon^2}\widehat{v}. \end{cases} \tag{15}$$

That is the Magic coefficient is everywhere except in the sink. The formal asymptotic limit $\widehat{v} \approx -\frac{\varepsilon M}{\sigma}\partial_x \widehat{u}$. Plugging in the first equation one gets the limit diffusion equation

$$\partial_t u - \left(\frac{M^2}{\sigma} + \frac{M\alpha}{\varepsilon}\right)\partial_{xx} u.$$

The diffusion coefficient is

$$\frac{M^2}{\sigma} + \frac{M\alpha}{\varepsilon} = \frac{1}{\sigma} + \frac{M(M-1)}{\sigma}.$$

The correct coefficient is modified by a factor proportional to $M(M-1)$. This term vanishes for $M \approx 0$ or $M \approx 1$, which means that either $\varepsilon << \alpha$ or $\alpha << \varepsilon$. If $\frac{\alpha}{\varepsilon} = O(1)$ then the correction is non zero as well.

This is why such modification (15) is not recommended in practice for numerical methods, see also [2].

4.2. **Non constant coefficients.**

Another case often encountered in practice is when the coefficients are non constant. Let us assume that $\sigma = \sigma(x) > 0$ is non constant, positive and smooth. A modified equation that corresponds to this situation is

$$\begin{cases} \partial_t \widehat{u} + \partial_x\left(\dfrac{M(x)}{\varepsilon}\widehat{v}\right) - \partial_x\left(\dfrac{M(x)\alpha}{\varepsilon}\partial_x \widehat{u}\right) = 0, \\ \partial_t \widehat{v} + \dfrac{M(x)}{\varepsilon}\partial_x \widehat{u} - \partial_x\left(\dfrac{M(x)\alpha}{\varepsilon}\partial_x \widehat{v}\right) = -\dfrac{\sigma M(x)}{\varepsilon^2}\widehat{v}. \end{cases} \tag{16}$$

It is an easy matter to check that this system satisfies an energy identity with the standard L^2 energy. It also rewrites in a more conservative way as

$$\begin{cases} \partial_t \widehat{u} + \partial_x\left(\dfrac{M(x)}{\varepsilon}\widehat{v}\right) - \partial_x\left(\dfrac{M(x)\alpha}{\varepsilon}\partial_x \widehat{u}\right) = 0, \\ \partial_t \widehat{v} + \partial_x\left(\dfrac{M(x)}{\varepsilon}\widehat{u}\right) - \partial_x\left(\dfrac{M(x)\alpha}{\varepsilon}\partial_x \widehat{v}\right) = -\dfrac{\sigma M(x)}{\varepsilon^2}\widehat{v} + \dfrac{\partial_x M(x)}{\varepsilon}u. \end{cases} \tag{17}$$

In theory the previous method method can be used to analyze the limit $\varepsilon \to 0$ of this system.

4.3. **Open problems.** We hope, and believe, that this study may provide a basis for the design and analysis of AP schemes for more complex equations and systems. For example for larger systems (3 equations and more), it is possible a priori to study systems of modified equations having in mind that the coefficient M becomes a matrix, the ultimate goal still being to design astute M such that error estimates are independent of the stiffness parameter ε. This is largely an open problem in the general case.

REFERENCES

[1] L. Gosse and G. Toscani, *An asymptotic-preserving well-balanced scheme for the hyperbolic heat equations*, C. R. Acad. Sci Paris, Ser. I **334** (2002) 337-342.

[2] S. Jin and D. Levermore, *Numerical schemes for hyperbolic conservation laws with stiff relaxation terms*, JCP **126**, (1996) 449-467.

[3] C. Buet, E. Franck and B. Després, *Design of asymptotic preserving Finite Volume schemes for the hyperbolic heat equation on unstructured meshes*, Numer. Math. **122**, (2012) 227-278.

[4] M. Lemou and L. Mieussens, *A new symptotic preserving scheme based on micro-macro formulation for linear kinetic equations in the diffusion limit.*, SIAM J. Sci. COMPUT. **31**, 1, (2008) 334-368.

[5] P. Degond, F. Deluzet, A. Sangam and M-H. Vignal, *An asymptotic preserving scheme for the Euler equations in a strong magnetic field*, J. Comput. Phys., **228**, (2009) 3540-3558.

E-mail address: despres@ann.jussieu.fr
E-mail address: christophe.buet@cea.fr
E-mail address: efranck21@gmail.com

WAVE-WAVE INTERACTIONS OF A GASDYNAMIC TYPE

Liviu Florin Dinu

Institute of Mathematics of the Romanian Academy
P.O.Box 1-764, Bucharest, RO-014700, Romania

Abstract. Two distinct contexts [isentropic; strictly anisentropic] are considered for a hyperbolic quasilinear system of a gasdynamic type [Euler]. For each of these two contexts two genuinely nonlinear, geometrical and analytical approaches are considered [of a Burnat type and, respectively, of a Martin type]. Each of these two approaches leads particularly to a pair of classes of solutions [wave solutions; wave-wave regular interaction solutions]. We finally parallel the mentioned approaches by using various comparisons between the mentioned pairs of classes.

1. "Algebraic" approach of a Burnat type. Genuine nonlinearity restrictions.

1.1. **Introduction.** For the multidimensional first order hyperbolic system of a gasdynamic type

$$\sum_{j=1}^{n}\sum_{k=0}^{m} a_{ijk}(u)\frac{\partial u_j}{\partial x_k}=0, \qquad 1\le i\le n \tag{1.1}$$

the "algebraic" approach (Burnat [1]) starts with identifying *dual* pairs of directions $\vec{\beta},\vec{\kappa}$ [we write $\vec{\kappa}\rightharpoondown\vec{\beta}$] connecting [via their duality relation] the hodograph [= in the hodograph space H of the entities u] and physical [= in the physical space E of the independent variables] characteristic details. The duality relation at $u^*\in H$ has the form:

$$\sum_{j=1}^{n}\sum_{k=0}^{m} a_{ijk}(u^*)\beta_k\kappa_j=0, \qquad 1\le i\le n. \tag{1.2}$$

Here $\vec{\beta}$ is an *exceptional* direction [= *normal characteristic* direction (orthogonal in the physical space E to a characteristic character)]. A direction $\vec{\kappa}$ dual to an exceptional direction $\vec{\beta}$ is said to be a *hodograph characteristic* direction. A real character of the exceptional / hodograph characteristic directions implied in (1.2) is concurrent with the hyperbolicity of (1.1).

Example 1 (Lax [7]). For the *one-dimensional* strictly hyperbolic version of the system (1.1) a *finite* number n of dual pairs $\vec{\kappa}_i\rightharpoondown\vec{\beta}_i$ consisting in $\vec{\kappa}_i=\vec{R}_i$ and $\vec{\beta}_i=\Theta_i(u)[-\lambda_i(u),1]$, where $\vec{R}_i$ is a right eigenvector of the $n\times n$ matrix a and λ_i is an eigenvalue of a, are available ($i=1,...,n$). Each dual pair associates in this case, at each $u^*\in\mathcal{R}$ [for a suitable region $\mathcal{R}\subset H$], to a vector $\vec{\kappa}$ a *single* dual vector $\vec{\beta}$.

2000 *Mathematics Subject Classification.* 35A30, 35L60, 35L65, 35Q35, 76N15.

Key words and phrases. Burnat type geometrical approach, Martin type geometrical approach, two-dimensional isentropic solutions, quantifiable "amount" of genuine nonlinearity, multidimensional regular interaction.

Example 2 (Peradzyński [10]). For the *two-dimensional* isentropic version of (1.1) an *infinite* number of dual pairs are available at each $u^* \in H$. Each dual pair associates, at the mentioned u^*, to a vector $\vec{\kappa}$ a *single* dual vector $\vec{\beta}$.

Example 3 (Peradzyński [11]). For the isentropic description corresponding to the *three-dimensional* version of (1.1) an *infinite* number of dual pairs are available at each $u^* \in H$. Each dual pair associates, at the mentioned u^*, to a vector $\vec{\kappa}$ a *finite* [constant, $\neq 1$] number of k independent exceptional dual vectors $\vec{\beta}_j$, $1 \leq j \leq k$; and therefore has the structure $\vec{\kappa} \rightharpoonup (\vec{\beta}_1,\ldots,\vec{\beta}_k)$.

Definition 1.1. (Burnat [1]) A curve $\mathcal{C} \subset H$ is said to be *characteristic* if it is tangent at each point of it to a characteristic direction $\vec{\kappa}$. A hypersurface $\mathcal{S} \subset H$ is said to be *characteristic* if it possesses at least a characteristic system of coordinates.

1.2. **Genuine nonlinearity. Simple waves solutions.**

Remark 1. In case of an one-dimensional strictly hyperbolic version of (1.1) a hodograph characteristic curve $\mathfrak{C} \subset \mathfrak{R} \subset H$, of index i, is said to be *genuinely nonlinear (gnl)* if the dual constructive pair $\vec{\kappa}_i \rightharpoonup \vec{\beta}_i$ is restricted [the restriction is on the *pair* !] by $\vec{\kappa}_i(u) \diamond \vec{\beta}_i(u) \equiv \vec{R}_i(u) \cdot \mathrm{grad}_u \lambda_i(u) \neq 0$ in $\mathfrak{R}$; see Example 1. This condition transcribes the requirement $\frac{\mathrm{d}\vec{\beta}}{\mathrm{d}\alpha} \neq 0$ along *each* hodograph characteristic curve $\mathfrak{C}$.

Definition 1.2. We naturally extend the *gnl* character of a hodograph characteristic curve $\mathfrak{C}$ to the cases corresponding to Examples 2 and 3, by requiring along $\mathfrak{C}$: $\left|\frac{\mathrm{d}\vec{\beta}}{\mathrm{d}\alpha}\right| \neq 0$ and, respectively, $\sum\limits_{\mu=1}^{k} \left|\frac{\mathrm{d}\vec{\beta}_\mu}{\mathrm{d}\alpha}\right| \neq 0$.

Definition 1.3. A solution of (1.1) whose [one-dimensional] hodograph is laid along a *gnl* characteristic curve is said to be a *simple waves solution* (here below also called *wave*). The *gnl* character implies a *nondegeneracy* [in the sense of a "funning out"] of such a solution.

Here are three types of simple waves solutions, presented in an implicit form − respectively associated, in presence of a *gnl* character, to the Examples 1−3 above [$\alpha(x,t)$ results from the implicit function theorem; the solution is structured by (1.2)]

$$\left.\begin{array}{c} u = U[\alpha(x,t)];\ \alpha = \theta(\xi),\ \ \xi = x - \zeta_i(\alpha)t, \\ u = U[\alpha(x,t)];\ \alpha = \theta(\xi),\ \xi = \sum_{\nu=0}^{m} \beta_\nu[U(\alpha)]x_\nu = \sum_{\nu=0}^{m} \beta_\nu \left\{U[\theta(\xi)]\right\} x_\nu, \\ u = U[\alpha(x,t)];\ \alpha = \theta(\xi_1,\ldots,\xi_k),\ \xi_j = \sum_{\nu=0}^{m} \beta_{j\nu}[U(\alpha)]x_\nu;\ 1 \leq j \leq k \end{array}\right\} \begin{array}{c} \dfrac{\mathrm{d}U}{\mathrm{d}\alpha} = \vec{\kappa} \\ \text{along } \mathfrak{C}. \end{array}$$

1.3. **Genuine nonlinearity: A constructive extension. Riemann−Burnat invariants. A subclass of the wave-wave regular interaction solutions.**

Remark 2. Let $R_1,\ldots,R_p$ be *gnl* characteristic coordinates on a given p-dimensional characteristic region $\mathfrak{R}$ of a hodograph hypersurface $\mathfrak{S}$ with the normal $\vec{n}$. Solutions of the ***intermediate*** system (Peradzyński [10])

$$\frac{\partial u_l}{\partial x_s} = \sum_{k=1}^{p} \eta_k \kappa_{kl}(u) \beta_{ks}(u),\ u \in \mathcal{R};\ 1 \leq l \leq n,\ 0 \leq s \leq m;\ \ \vec{\kappa}_k \perp \vec{n},\ \ 1 \leq k \leq p \qquad (1.3)$$

appear to concurrently satisfy the system (1.1) [we carry (1.3) into (1.1) and take into account (1.2)]. This indicates a key importance of the "algebraic" concept of dual pair.

Definition 1.4. A solution of (1.1) whose hodograph is laid on a characteristic hypersurface is said to correspond to a *wave-wave regular interaction* if its hodograph possesses a *gnl* system of coordinates *and* there exists a set of ***Riemann—Burnat invariants*** $R(x)$, structuring the dependence on x of the solution u by a *regular* interaction representation

$$u_l = u_l[R_1(x_0, ..., x_m), ..., R_p(x_0, ..., x_m)],\ 1 \le l \le n. \tag{1.4}$$

Remark 3. We consider next a ***subclass*** of the wave-wave solutions of (1.1). This subclass results whether (1.1) is replaced by (1.3) in Definition 1.4 [because, the solutions of (1.3) concurrently satisfy (1.1); cf. Remark 2]. To construct this subclass we have to put together (1.3) and (1.4). We compute from (1.4)

$$\frac{\partial u_l}{\partial x_s} = \sum_{k=1}^{p} \frac{\partial u_l}{\partial R_k} \cdot \frac{\partial R_k}{\partial x_s} = \sum_{k=1}^{p} \kappa_{kl}(u) \frac{\partial R_k}{\partial x_s},\quad 1 \le l \le n;\ 0 \le s \le m \tag{1.5}$$

and compare (1.5) with (1.3), taking into account the independence of the characteristic directions $\vec{\kappa}_k$. It results that for a wave-wave regular interaction solution in the mentioned subclass, $R_i(x)$ in (1.4) must fulfil a *reasonable* (overdetermined and Pfaff) system

$$\frac{\partial R_k}{\partial x_s} = \eta_k \beta_{ks}[u(R)],\quad 1 \le k \le p,\ 0 \le s \le m. \tag{1.6}$$

Sufficient restrictions for solving (1.6) are proposed in [5], [6], [10], [11].

1.4. **Wave-wave regular interaction solutions: Some remarks. Quantifiable “amount” of genuine nonlinearity.** Four circumstances appear to be significant for solutions with a characteristic hodograph: • the case of a characteristic hodograph surface for which *all* the coordinate systems are *gnl*, • the case of a characteristic hodograph surface for which only *a part* of the coordinate systems are *gnl*, • the case of a characteristic hodograph surface for which all the coordinate systems are *linearly degenerate (ldg)* (“$\neq 0$” is replaced by “$= 0$” in Definition 1.2), • the case of a hodograph surface *which is not Burnat characteristic* (Definition 1.1); for such a circumstance a characteristic character of the hodograph surface *may persist* in an *alternative* sense (ex. in a “differential” (Martin type) sense; cf. section 2 bellow). • The first two cases above could be exemplified by two significant analytical and exact self-similar two-dimensional gasdynamic solutions [in usual notations; c is the sound velocity].

A first significant solution

$$v_x = \frac{1}{\gamma}\xi + \mathfrak{c},\ \ v_y = \frac{1}{\gamma}\eta + \bar{\mathfrak{c}};\ \text{ arbitrary } \mathfrak{c}, \bar{\mathfrak{c}},\quad c^2 = \frac{1}{2}\left[\left(\frac{\gamma-1}{\gamma}\xi - \mathfrak{c}\right)^2 + \left(\frac{\gamma-1}{\gamma}\eta - \bar{\mathfrak{c}}\right)^2\right],$$

A second significant solution [for $\frac{3-\gamma}{\gamma+1} < \mathfrak{a} < 1$]

$$v_x = \mathfrak{a}\xi \pm \eta\sqrt{(1-\mathfrak{a})\left(\mathfrak{a} - \frac{3-\gamma}{\gamma+1}\right)} + K\sqrt{1-\mathfrak{a}},\quad K = \frac{\mathfrak{c}}{\sqrt{1-\mathfrak{a}}} = \mp\frac{\bar{\mathfrak{c}}}{\sqrt{\mathfrak{a} - \frac{3-\gamma}{\gamma+1}}}$$

$$v_y = \pm\xi\sqrt{(1-\mathfrak{a})\left(\mathfrak{a} - \frac{3-\gamma}{\gamma+1}\right)} + \eta\left(\frac{4}{\gamma+1} - \mathfrak{a}\right) \mp K\sqrt{\mathfrak{a} - \frac{3-\gamma}{\gamma+1}}$$

$$c = \varepsilon\sqrt{\frac{(3-\gamma)(\gamma-1)}{2(\gamma+1)}}\left(\xi\sqrt{1-\mathfrak{a}} \mp \eta\sqrt{\mathfrak{a} - \frac{3-\gamma}{\gamma+1}} - K\right),\quad \varepsilon = \pm 1\,,$$

where we denote
$$\xi = \frac{x - x_0}{t - t_0}, \quad \eta = \frac{y - y_0}{t - t_0}\ .$$

To the self-similar form of the two-dimensional anisentropic version of (1.1) we associate the *self-similar Mach number* ([12]):
$$\widetilde{M} = \frac{1}{c}\sqrt{(v_x - \xi)^2 + (v_y - \eta)^2}.$$

Remark 4. For the two significant solutions mentioned above we compute $\widetilde{M}\equiv\sqrt{2}>1$ and respectively $\widetilde{M} \equiv \text{constant} = \frac{2}{\sqrt{3-\gamma}} > 1$. This is in contrast with the details of Zhang and Zheng *irregular* interaction ([12]) which allows a *coexistence* of pseudo supersonic regions ($\widetilde{M} > 1$) and pseudo subsonic regions ($\widetilde{M} < 1$).

Remark 5. Quantifiable "amount" of genuine nonlinearity. The two exemplifying solutions above are taken from an exhaustive list of solutions with a suggestive suitable structure ([2]). • The hodograph of the first of these solutions is structured by ***three*** characteristic ***genuinely nonlinear*** fields (two families of ***conical helices*** and a family of ***horizontal circles***) – it appears to be a characteristic surface with three *gnl* characteristic systems of coordinates. We are able in this case to present the first significant solution in ***three distinct manners*** as a regular interaction of multidimensional simple waves solutions. The first mentioned solution appears to be associated locally to a set of ***three concurrent and distinct structures*** [multidimensional wave-wave regular interaction structures; Riemann–Burnat representations] ([2]). • The hodograph of the second of these solutions is structured by *three* characteristic fields too (three families of ***straightlined*** hodograph characteristic fields) of which two are ***non-horizontal*** and appear to be ***genuinely nonlinear*** and the third is ***horizontal*** and shows a ***linearly degenerate*** character. In this case we dispose of a ***single gnl*** hodograph system of characteristic coordinates. The mentioned solution appears to be associated locally to a ***single*** wave-wave regular interaction structure [Riemann–Burnat representation] ([2]).

2. "Differential" approach of a Martin type.

2.1. **Anisentropic context. Details of a Martin type approach.** An *anisentropic* flow is present in the region behind a shock discontinuity which propagates with a non-uniform velocity [unsteady one-dimensional; curved shock path; cf. the theory of the compressive piston] or in the region behind a curved shock [steady two-dimensional; rotational]. Because, the jump of the entropy through the shock depends on the velocity of the shock [one-dimensional] or on the inclination of the shock line [two-dimensional; rotational].

The Burnat type approach 1.1–1.3, leading to a wave-wave regular interaction structure, is valid for both isentropic and anisentropic versions of (1.1). In particular for the case of two independent variables. This approach provides a pair of classes of solutions of this mentioned system [Burnat type wave and wave-wave regular interaction solutions]. • In a strictly anisentropic context the Burnat type pair of classes will be *parallelled* here below [sections 2.2–2.4] by a Martin type pair of classes.

In a Martin type approach ([8]) we interchange the pairs of independent variables x, t [or x, y] with ψ, p [usual notations] on the [natural] assumptions $\frac{\partial p}{\partial t}\frac{\partial \psi}{\partial x} - \frac{\partial p}{\partial x}\frac{\partial \psi}{\partial t} \neq 0$ [or, respectively, $\frac{\partial p}{\partial x}\frac{\partial \psi}{\partial y} - \frac{\partial p}{\partial y}\frac{\partial \psi}{\partial x} \neq 0$]. A Martin type approach represents the solution of the system (1.1) *in two independent variables* through an entity ξ [unsteady one-dimensional] or through a couple of entities ξ, η [steady two-dimensional; rotational] fulfilling each a Monge−Ampère type equation in the independent variables ψ, p ([4]).

Unsteady one-dimensional representation

$$v_x = \frac{\partial \xi}{\partial \psi},\ \ t = \frac{\partial \xi}{\partial p},\ \ x = \int \left(\frac{\partial \xi}{\partial \psi} \frac{\partial^2 \xi}{\partial p \partial \psi} + \frac{1}{\rho} \right) \mathrm{d}\psi + \left(\frac{\partial \xi}{\partial \psi} \frac{\partial^2 \xi}{\partial p^2} \right) \mathrm{d}p. \tag{2.1}$$

where ξ fulfils the Monge−Ampère type equation ([8])

$$\frac{\partial^2 \xi}{\partial p^2} \frac{\partial^2 \xi}{\partial \psi^2} - \left(\frac{\partial^2 \xi}{\partial p \partial \psi} \right)^2 = -\zeta^2(p, \psi) \equiv \frac{\partial}{\partial p} \left(\frac{1}{\rho} \right) \equiv -\frac{1}{\rho^2 c^2} \tag{2.2}$$

with $\rho = \rho(p, \psi)$ and $c(p, \psi) = \sqrt{\left(\frac{\partial \rho}{\partial p} \right)_S^{-1}}$ an ad hoc sound speed.

Steady two-dimensional rotational supersonic representation

$$x = \frac{\partial \eta}{\partial p},\ \ y = -\frac{\partial \xi}{\partial p},\ \ v_x = \frac{\partial \xi}{\partial \psi},\ \ v_y = \frac{\partial \eta}{\partial \psi}, \tag{2.3}$$

where ξ and η fulfil the same Monge−Ampère type equation ([4])

$$4\mathfrak{F}\left[\left(\frac{\partial^2 \xi}{\partial p \partial \psi} \right)^2 - \frac{\partial^2 \xi}{\partial p^2} \frac{\partial^2 \xi}{\partial \psi^2} \right] - 4 \left(\frac{\partial \xi}{\partial \psi} \frac{\partial \mathfrak{F}}{\partial p} \right) \frac{\partial^2 \xi}{\partial p \partial \psi} + 2 \left(\frac{\partial \xi}{\partial \psi} \frac{\partial \mathfrak{F}}{\partial \psi} \right) \frac{\partial^2 \xi}{\partial p^2} + \left\{ \left(\frac{\partial \mathfrak{F}}{\partial p} \right)^2 - 2 \left[\mathfrak{F} - \left(\frac{\partial \xi}{\partial \psi} \right)^2 \right] \frac{\partial^2 \mathfrak{F}}{\partial p^2} \right\} = 0. \tag{2.4}$$

The entities η and ξ are connected yet: a given solution ξ of (2.4) is paired by a computed [in terms of ξ] solution η of (2.4).

Remark 6. The representations (2.1)−(2.4) result in presence of a conservation laws form of the gasdynamic system: a *restriction*, in contrast with the Burnat constructive availability. • The steady two-dimensional rotational and supersonic representation corresponds to a *natural adaptation* of the Martin's unsteady one-dimensional details ([4]).

The hyperbolicity of (1.1) results in the hyperbolicity of the Monge−Ampère type equation associated. In the steady two-dimensional rotational case we have to add, to provide hyperbolicity, the restriction of a supersonic character of the flow.

Remark 7. The anisentropic description has a *pseudo isentropic* character. Precisely: in each of the two mentioned cases [unsteady one-dimensional; steady two-dimensional, rotational] the degenerate fields [particle lines; streamlines] are not characteristic ([4]).

Remark 8. We are left [cf. Remark 7], for each of the mentioned cases, with two genuinely nonlinear characteristic fields − which are in correspondence with the two characteristic fields of a Monge−Ampère type equation ([4]). For example, in the steady two-dimensional, rotational and supersonic $(V^2 - c^2 > 0)$ case the characteristic directions for (2.4) in the plane p, ψ are given by

$$\left(\frac{\mathrm{d}p}{\mathrm{d}\psi} \right)_{\pm} = \frac{2\mathfrak{F} \frac{\partial^2 \xi}{\partial p \partial \psi} - \frac{\partial \mathfrak{F}}{\partial p} \frac{\partial \xi}{\partial \psi} \pm \sqrt{\Delta}}{-2\mathfrak{F} \frac{\partial^2 \xi}{\partial p^2}}\ ,\quad \Delta = \frac{4}{\rho^2 c^2} v_y^2 \left(V^2 - c^2 \right)\ ,\quad V^2 = v_x^2 + v_y^2\ . \tag{2.5}$$

We have

$$\begin{array}{ll} -V^2 \mathrm{d}y + \dfrac{1}{\rho c} \left[c v_x \pm v_y \sqrt{V^2 - c^2} \right] \mathrm{d}\psi = 0 & \\ & \text{along the characteristics } \overline{C}_{\pm} \\ & \text{of (2.4).} \\ v_y \left[v_y \mathrm{d}v_x - v_x \mathrm{d}v_y \mp \dfrac{1}{\rho c} \sqrt{V^2 - c^2} \mathrm{d}p \right] = 0 & \end{array}\ . \tag{2.6}$$

The Mach lines $C_\pm$ of the gasdynamic system in the physical plane and the characteristics $\overline{C}_\pm$ [(2.5)] of the Monge$-$Ampère type equation (2.4) are in correspondence. In fact, we get from $(2.6)_1$ and $\mathrm{d}\psi = -(\rho v_y)\mathrm{d}x + (\rho v_x)\mathrm{d}y$

$$-V^2\mathrm{d}y + \frac{1}{\rho c}\left[cv_x \pm v_y\sqrt{V^2-c^2}\right](v_x\mathrm{d}y + v_y\mathrm{d}x) = 0 \quad \text{along the characteristics } \overline{C}_\pm \text{ of (2.4)}$$

which results in

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{cv_x \pm v_y\sqrt{V^2-c^2}}{cv_y \mp v_x\sqrt{V^2-c^2}} = \frac{v_xv_y \pm c\sqrt{V^2-c^2}}{v_x^2 - c^2} = \lambda_\pm \quad \text{along the characteristics } \overline{C}_\pm \text{ of (2.4)}$$

where λ_+ and λ_- are the Mach eigenvalues of the gasdynamic system.

2.2. **Martin type linearization.** • Finding a solution to a Monge$-$Ampère type equation [(2.2) or (2.4)] is a hard task generally. Incidentally, such a solution can be constructed in presence of a ***geometrical Martin type linearization***. • A Martin type linearization appears to be an ***intermediate*** constructive element [parallel to (1.3)]. • A linearization of a Martin type is available whether there exists for (2.2) or (2.4) at least a pair of *linear in* ξ intermediate integrals $[\mathfrak{I}_+, \mathfrak{I}_-]$ $-$ constant along a characteristic of the Monge$-$Ampère type equation considered

$$\mathfrak{I}_+ = R_+ = \text{constant}_+ (\text{along } \overline{C}_+), \quad \mathfrak{I}_- = R_- = \text{constant}_- (\text{along } \overline{C}_-). \tag{2.7}$$

2.3. **Nature of a Martin type linearization. Unsteady one-dimensional gas dynamics. Burnat type approach vs. Martin type approach.** • In the unsteady one-dimensional *isentropic* gas dynamics a pair of intermediate integrals of (2.2) $\mathcal{F}_\pm \equiv \frac{\partial\xi}{\partial\psi} \pm \int\zeta(p)\mathrm{d}p$ can be identified directly, for $\zeta=\zeta(p)$ in (2.2), by using details of the Monge$-$Ampère representation (2.1) $[v_x = \frac{\partial\xi}{\partial\psi}]$ to transcribe the well known Riemann invariance relations: $v_x \pm \int\zeta(p)\mathrm{d}p = R_\mp$. Two approaches [Burnat type (associated to the Riemann invariance relations), Martin type (associated to the pair of intermediate integrals of (2.2) which transcribe the isentropic Riemann invariance relations)] appear to be *coincident* in an isentropic context. • Next, the two halves of this coincidence [Burnat half, Martin half] are extended separately: cf. section 1 [Burnat half: a dimensional (possibly anisentropic) extension] or cf. section 2 [Martin *linearized* half: an anisentropic extension]. • In order to construct an *anisentropic* extension of the Martin linearization approach we have to identify some anisentropic pairs of linear in ξ intermediate integrals of the Monge$-$Ampère equation associated to (1.1). • There are ***few*** (six) cases of Martin linearization (see [9] for the one-dimensional case, and [4] for the steady, rotational and supersonic two-dimensional case); in contrast with the availability of a Burnat type construction. • Finally, we have to compare the two parallel separate anisentropic extensions [Burnat type, Martin type] which are still initiated by an isentropic coincidence: to identify some significant consonances and, concurrently, some nontrivial contrasts.

2.4. **Details of a Martin type linearization. Pseudo simple waves solution. Pseudo interaction solution. Riemann$-$Martin invariants.** • If in (2.7) R_+, R_- depend on the characteristic $[\overline{C}_+, \overline{C}_-]$ we construct a ***pseudo interaction solution*** by considering R_+, R_- as new independent variables to show (Martin [9]) that in these independent variables the entities of the flow fulfil some *linear* Euler$-$Poisson$-$Darboux equations [for which the solutions are well known]; we present these representations by

$$p\!=\!p(R_+,\!R_-),\ \psi\!=\!\psi(R_+,\!R_-),\ v_x\!=\!v_x(R_+,\!R_-),\ t\!=\!t(R_+,R_-),\ x\!=\!x(R_+,R_-) \quad (2.8)$$

where $x(R_+,R_-)$ results by quadratures. Reversing $(2.8)_{4,5}$ into $R_\pm = R_\pm(x,t)$ will induce a form of solution (2.8), parallel to (1.4) [as $R_\pm$ have a characteristic nature]. We call $R_\pm(x,t)$ ***Riemann−Martin invariants***.

• If in (2.7) R_+ or R_- are overall constants then a solution of the *linear in* ξ equation $\mathfrak{I}_+ \equiv R_+$ [or $\mathfrak{I}_- \equiv R_-$] appears to be automatically a solution of the mentioned Monge−Ampère type equation. We use this solution to get a solution of (1.1) and call the computed solution a ***pseudo simple waves solution***.

Remark 9. We prove that the pseudo interaction hodograph (2.8) is not a Burnat characteristic surface. Still, incidentally and essentially for the linearized approach, this hodograph appears to be associated with an example of surface for which a characteristic character persists − in a Martin sense. In the unsteady one-dimensional case this results from the following relation between the Burnat type ($\vec{\kappa}$) and the Martin type ($\vec{\mu}$) hodograph characteristic directions

$$\vec{\mu}_\pm\!=\!\left(\frac{\partial p}{\partial R_\pm},\!\frac{\partial v_x}{\partial R_\pm},\!\frac{\partial S}{\partial R_\pm}\right)^t\!=\eta_\mp\vec{\kappa}_\mp+\widetilde{\eta}_\mp\vec{\kappa}_0$$

with

$$S(R_+,\!R_-)\equiv F[\psi(R_+,\!R_-)],\ \ \eta_\mp\!=\!\frac{1}{\Lambda_\mp}\frac{\partial v_x}{\partial R_\pm},\ \ \widetilde{\eta}_\mp\!=\!\frac{\partial S}{\partial R_\pm}\ .$$

In the isentropic context the two approaches are coincident as $\widetilde{\eta}_\mp \equiv 0$.

Remark 10. In contrast with the Burnat type construction, a pseudo simple waves solution has a two-dimensional hodograph and for it none of the characteristic fields $C_\pm$ in the physical plane is made of straightlines generally.

Example 4. To $\zeta\!=\!\frac{\psi^{\nu-1}}{p^{\nu+1}}$ $\left(\nu\!=\!-\frac{\gamma-1}{2\gamma},\text{ integral } \nu,\ \nu\!\neq\!0{,}1\right)$ in (2.2) two intermediate integrals $\mathfrak{I}_\pm\!\equiv\!p\frac{\partial\xi}{\partial p}\!+\!\psi\frac{\partial\xi}{\partial\psi}\!-\!\xi\!\pm\!\frac{1}{\nu}\left(\frac{\psi}{p}\right)^\nu$ correspond. We satisfy $\mathfrak{I}_+\!\equiv\!R_+\!=\!0$ by $\xi\!=\!\frac{1}{\nu}\left(\frac{\psi}{p}\right)^\nu$, and calculate from (2.1)

$$p=-\left(\frac{\nu+1}{2\nu+1}\right)^\nu\frac{t^{2\nu-1}}{(-x)^\nu},\quad v_x=\frac{2\nu+1}{\nu+1}\frac{x}{t},\quad \psi=-\left(\frac{\nu+1}{2\nu+1}\right)^{\nu+1}\frac{t^{2\nu+1}}{(-x)^{\nu+1}}. \quad (2.9)$$

This is a [local] pseudo simple waves solution of the gasdynamic system corresponding to a certain region $\mathcal{D}\!\subset\!E$ (for example, a region of $t\!>\!0$, $x\!<\!0$). For this solution we compute

$$|x|\!=\!K_\alpha|t|^{k_\alpha};\ \ K_\alpha\!=\!\log\frac{|x^*|}{|t^*|^{k_\alpha}},\ \ \alpha\!=\!-,0,+,\ \ \text{along } C_\alpha\!\ni\!(x^*,t^*) \quad (2.10)$$

where for $1\!<\!\gamma\!<\!\frac{5}{3}$ we have

$$k_-\!=\!\frac{2\nu}{\nu\!+\!1}\!=\!-2\frac{\gamma\!-\!1}{\gamma\!+\!1},\ \ k_0\!=\!\frac{2\nu\!+\!1}{\nu\!+\!1}\!=\!\frac{2}{\gamma\!+\!1},\ \ k_+\!=\!2\,.$$

Remark 11. The hodograph of a formal regular interaction of pseudo simple waves solutions will be then made by glueing, along suitable *M*- characteristics [induced on the hodograph surface by the characteristics of the Monge−Ampère type equation considered], a hodograph of a pseudo interaction solution with some suitable hodographs of pseudo simple waves solutions.

3. Final remarks.

Consonances of the two approaches [Burnat type; Martin type]. • The two approaches depend on the constructive presence of an *intermediate element:* the system (1.3) for the Burnat type and, respectively, the possibility of a geometrical linearization for the Martin type. • A Martin type approach may provide a hodograph characteristic character in absence of a Burnat type characteristic character [Remarks 5,9]. • Each of the two approaches result in a representation structured by Riemann invariants [Riemann−Burnat or Riemann−Martin]. Cf. (1.4) or section 2.4.

Contrasts of the two approaches. • The Martin approach has a *fragile* character. It requires *dimensional restrictions* [two independent variables], it is based on *few* cases of geometrical linearization, it is initiated from a conservation laws form of the system of equations (Remark 6). • The structure of a pseudo simple waves solution is in contrast with the structure of a Burnat type simple waves solution [two-dimensional hodograph (Remark 10) vs. one-dimensional hodograph (Definition 1.3)].

Other remarks. • Our analysis in section 1 suggests that a *regular* character reflects a *multidimensional* and *skew* interaction construction generally. In Zhang and Zheng *irregular* interaction structure ([12]) the contributing waves are *one-dimensional* and *orthogonal.* • In two independent variables a simple waves solution appears to be constant along linear characteristics. This behaviour appears to be lost in a *Burnat−Peradzyński dimensional hierarchy* [see Examples 1−3 and section 1.2] or in a *Martin anisentropic hierarchy* [cf. Remark 10].

REFERENCES

[1] M. Burnat, *The method of characteristics and Riemann's invariants for multidimensional hyperbolic systems*, Sibirsk. Math. J., **11** (1970), 279−309.

[2] L.F. Dinu, *Multidimensional wave-wave regular interactions and genuine nonlinearity*, Preprint Series of Newton Institute for Math. Sci., Cambridge UK, No. 29, 2006. This work has been essentially considered in the classifying paper J. Li and Y. Zheng, *Interaction of four rarefaction waves in the bi-symmetric class of the two-dimensional Euler equations*, Commun. Math. Phys., **296** (2010), 303−321

[3] L.F. Dinu and M.I. Dinu, *Nondegeneracy, from the prospect of wave-wave regular interactions of a gasdynamic type*, Proceedings of the International Conference on Mathematical and Numerical Aspects of Waves, University of Reading UK, p.461, 2007.

[4] L.F. Dinu and M.I. Dinu, *Martin's "differential" approach*, Preprint Series of Newton Institute for Math. Sci., Cambridge UK, No. 21, 2009.

[5] E.V. Ferapontov and K.R. Khusnutdinova, *On integrability of (2+1)-dimensional quasilinear systems*, Commun. Math. Phys., **248** (2004), 187−206, .

[6] E.V. Ferapontov and K.R. Khusnutdinova, *The Haahtjes tensor and double waves in multidimensional systems of hydrodynamic type: a necessary condition for integrability*, Proc. Roy. Soc., **A 462** (2006), 1197-1219,

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

[8] M.H. Martin, *A new approach to problems in two dimensional flow*, Quarterly Appl. Math., **8** (1950), 137−150.

[9] M.H. Martin, *The Monge−Ampère partial differential equation $rt - s^2 + \lambda^2 = 0$*, Pacific J. Math., **3** (1953), 165−187.

[10] Z. Peradzyński, *Nonlinear plane k-waves and Riemann invariants*, Bull. Acad. Polon. Sci., Ser. Sci. Tech., **19** (1971), 625−631.

[11] Z. Peradzyński, *Riemann invariants for the nonplanar k-waves*, Bull. Acad. Polon. Sci., Ser. Sci. Tech., **19** (1971), 717−724.

[12] T. Zhang and Y.-X. Zheng, Conjecture on structure of solutions of Riemann problem for two-dimensional gas dynamics, SIAM J. Math. Anal., **21** (1990), 593−630.

E-mail address: Liviu.Dinu@imar.ro

Author Index

Abreu, Eduardo 233
Achleitner, Franz 241
Alberti, Giovanni 399
Alessandro Morando 1007
Almeida, Luís 293
Amadori, Debora 641
Ancona, Fabio 249, 447
Andreianov, Boris 259
Angot, Philippe 277
Aregba-Driollet, Denise 269, 365
Arun, Koottungal Revi 439
Auphan, Thomas 277

Bae, Hantaek 285
Bae, Soohyun 125
Bagnerini, Patrizia 293
Ballew, Joshua 301
Bärwolff, Günter 317
Bauzet, Caroline 309
Benfield, Michael 749
Berger, Marsha J. 675
Berres, Stefan 317
Berselli, Luigi C. 325, 967
Berthon, Christophe 541
Bianchini, Stefano 399
Blokhin, Aleksander 991
Borsche, Raul 333, 575
Boscarino, Sebastiano 61
Boudin, Laurent 341
Bourdarias, Christian 615
Boutin, Benjamin 349
Bürger, Raimund 381
Bressan, Alberto 357
Briani, Maya 365
Buet, Christophe 509

Čanić, Sunčica 389
Caravenna, Laura 399
Carvalho, Filipe 407
Castañeda, Pablo 415, 791
Castelli, Pierre 709
Cavalli, Fausto 423
Chalons, Christophe 607
Chemetov, Nikolai 431
Chen, Guoxian 439
Chiodaroli, Elisabetta 1
Choi, Sun-Ho 125
Cipriano, Fernanda 431
Coclite, Giuseppe Maria 447
Coco, Armando 455
Colombo, Rinaldo M.147, 467, 575, 641, 889
Coquel, Frédéric 349, 631
Corli, Andrea 475
Coron, Jean-Michel 1023
Crippa, Gianluca 109, 957

de Almeida, Valmor F. 623
De Lellis, Camillo 1
Debrabant, Kristian 483
Degond, Pierre 179
Delis, A.I. 491
Després, Bruno 509
Dimarco, Giacomo 865
Dinu, Liviu Florin 515
Donadello, Carlotta 957
Donatelli, Donatella 525
Duan, Renjun 533
Duran, Arnaud 541
Dziuk, Gerhard 775

Elling, Volker 873
Ernest, Jan 551

Falcone, Maurizio 559
Fan, Haitao 475
Feireisl, Eduard 11
Festa, Adriano 559
Franck, Emmanuel 509
Frid, Hermano 567
Frouvelle, Amic 179
Fu, Xuedong 685
Fuchs, Franz 41

Furtado, Frederico 415

Gérard-Varet, David 29
Garavello, Mauro 467, 575
Gasser, Ingenuin 907
Gersbacher, Christoph 591
Giesselmann, Jan 599
Girardin, Mathieu 607
Gisclon, Marguerite 615
Glass, Olivier 249
Glimm, James 623
Goatin, Paola 501, 889
Godlewski, Edwige 631
Gowda, G. D. Veerappa 733
Grunert, Katrin 193
Guès, Olivier 277
Guermond, Jean-Luc 119
Guerra, Graziano 641

Ha, Seung-Yeal 125
Hantke, Maren 651
Hashimoto, Itsuko 659
Haspot, Boris 667
Helzel, Christiane 675
Herty, Michael 147
Hittmeir, Sabine 241
Holden, Helge 193
Hong, Zhenying 685
Huth, Frank 317

Jakobsen, Espen Robstad 483
James, François 693
Jenssen, Helge Kristian 749
Jiu, Quansen 701
Junca, Stéphane 615, 709

Kabil, Buğra 373
Kalise, Dante 717
Kall, Jochen 333
Karimou, Mohamed Gazibo 583
Kemm, Friedemann 725
Klingenberg, Christian 741
Kogan, Irina A. 749
Kokh, Samuel 607
Koley, Ujjwal 741, 759
Kotov, Dmitry 1047
Kränkle, Mirko 767
Kröner, Dietmar 775
Kumar, K. Sudarshan 733

Lattanzio, Corrado 163
Lécureux-Mercier, Magali 467

LeFloch, Philippe G. 349

Lim, Hyunkyung 623
Lin, Chi-Kun 1039
Liu, Jian-Guo 179
Lukáčová-Medvid'ová, Maria 931

Müller, Thomas 775, 801
Mailybaev, Alexei A. 783
Marcati, Pierangelo 525
Marche, Fabien 541
Marchesin, Dan 415, 791, 881, 949
Masmoudi, Nader 29
Mathiaud, Julien 341
Matos, Vítor 791
May, Georg 915
McMurry , Andrew 41
Meister, Andreas 823
Mishra, Siddhartha 41
Monache, Maria Laura Delle 501
Morosi, Carlo 847
Muha, Boris 389
Mulet, Pep 381
Natalini, Roberto 365
Nguyen, Khai T. 249
Nikolos, I.K. 491
Nishibata, Shinya 817
Noelle, Sebastian 439
Nolte, Martin 809

Ohnawa, Masashi 817
Ortleb, Sigrun 823

Panov, Evgeniy Yu. 831
Paola Trebeschi 1007
Pareschi, Lorenzo 865

Pelanti, Marica 839
Pernici, Mario 847
Pfeiffer, Axel 801
Pizzocchero, Livio 847
Plaza, Ramón G. 857
Popov, Bojan 119
Praveen, C. 733
Privat, Yannick 89
Puppo, Gabriella 923

Raynaud, Xavier 193
Reintjes, Moritz 209
Risebro, Nils Henrik 41, 759, 1031
Rispoli, Vittorio 865
Roberts, Joseph 873
Rodriguez-Bermudez, Panters 881
Rosini, Massimiliano D. 889
Rozanova, Olga 899
Russo, Giovanni 61, 455
Rybicki, Martin 907

Saye, Robert 81
Schütz, Jochen 915
Schmeiser, Christian 241
Schwandt, Hartmut 317
Seguin, Nicolas 631
Semplice, Matteo 923
Sethian, James 81
Sfakianakis, Nikolaos 931
Shao, Zhi-Qiang 941
Shearer, Michael 221
Shen, Wen 641
Shyue, Keh-Ming 839
Silva, Julio Daniel 949
Sjögreen, Bjorn 1047
Soares, Ana Jacinta 407
Spinolo, Laura V. 957

Spirito, Stefano 967
Strani, Marta 975
Suzuki, Masahiro 817

Terracina, Andrea 983
Thein, Ferdinand 651
Tkachev, Dmitry 991
Trélat, Emmanuel 89
Trakhinin, Yuri 999
Trivisa, Konstantina 285
Turpault, Rodolphe 541
Tzavaras, Athanasios E. 163, 599

Vauchelet, Nicolas 693
Villada, Luis M. 381

Waagan, Knut 1015
Wang, Yi 701
Wang, Zhiqiang 1023
Weber, Franziska 1031
Wollner, Winnifried 907
Woopen, Michael 915
Wu, Kung-Chien 1039

Xin, Zhouping 701

Yee, H.C. 1047
Yuan, Guangwei 685
Yuri Trakhinin 1007

Zanelli, Lorenzo 1059
Zhou, Yijie 623
Zuazua, Enrique 89

F. Ancona, A. Bressan, P. Marcati and A. Marson (Editors)
Hyperbolic Problems: Theory, Numerics, Applications

The International Conference devoted to Theory, Numerics and Applications of Hyperbolic Problems, HYP2012, was held in Padova on June 24–29, 2012. The conference was the fourteenth in a highly successful series of bi-annual meetings that has become one of the most important international events in Applied Mathematics. The volume contains more than 110 contributions that were presented in this conference, including plenary presentations by C. De Lellis, E. Feireisl, N. Masmoudi, S. Mishra, G. Russo, J. Sethian, E. Zuazua, and a contribution by the keynote speaker J. Glimm. These contributions cover a wide range of topics. A very partial list includes: new methods for constructing turbulent solutions to multi-dimensional systems of conservation laws based on Baire category, transport equations with non-Lipschitz velocity fields, relative entropy functionals and the stability of fluid systems, numerical methods for hyperbolic systems with stiff relaxation terms and for multiphase flow, new advances in homogenization theory, optimal sensor location for solutions to multidimensional wave equations, singularities in general relativity. The volume should appeal to researchers, students and practitioners with general interest in time-dependent problems governed by hyperbolic equations.

aimsciences.org