HANDBOOK OF DIFFERENTIAL EQUATIONS

Evolutionary Equations

VOLUME 5

Edited by

C.M. Dafermos

Milan Pokorny

NH

North-Holland is an imprint of Elsevier
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK

First edition 2009

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN-13: 978-0-444-53222-0

For information on all North-Holland publications visit our website at books.elsevier.com

Printed and bound in Hungary

09 10 10 9 8 7 6 5 4 3 2 1

Preface

This is the fifth and final volume of the Handbook of Differential Equations, dedicated to Evolutionary Equations. The objective of the series of five volumes was to present a panorama of this amazingly complex and rapidly developing branch of mathematics, and to illustrate some of its applications. In retrospect, the Editors believe that this goal has been achieved.

The present volume follows the tradition set by its predecessors, in collecting contributions that vary in style, mathematical methodology and target of applications, as this is a reflection of the nature of the area of evolutionary partial differential equations. Thus,

Chapter 1, by D. Bresch, provides a thorough discussion of the theory of so-called shallow-water equations, which model water waves in rivers, lakes and oceans. It addresses the issues of modeling, analysis and applications.

The following chapter, by D. Hilhorst, M. Mimura and H. Ninomiya, examines singular limits of reaction-diffusion systems, where the reaction is fast compared to the other processes. Applications range from the theory of the evolution of certain biological processes to the phenomena of Turing and cross-diffusion instability.

Chapter 3, by G. Métivier, provides a detailed study of a number of fascinating problems arising in nonlinear optics, at the high frequency and high intensity regime. Geometric and diffractive optics, and wave interactions are discussed.

Chapter 4, by R. Racke, addresses the issues of existence, blow-up and asymptotic stability of solutions to the equations of linear and nonlinear thermoelasticity, both in their classical form and in their formulation to exhibit "second sound".

The final chapter, by S. Vessela, discusses the important questions of unique spacelike continuation and backward uniqueness for linear second-order parabolic equations.

We are indebted to the authors for their valuable contributions, to the referees for their helpful comments, and to the editors and staff of Elsevier for their assistance.

Constantine Dafermos
Milan Pokorný

List of Contributors

Bresch, D., *Laboratoire de mathématique (LAMA) Unité Mixte de Recherche 5127 CNRS Université de Savoie UFR SFA, Campus Scientifique, 73376 Le Bourget-du-Lac Cedex, France* (Ch. 1)

Hilhorst, D., *Université Paris-Sud B. 425 Laboratoire d'Analyse Numérique Département de Mathématiques 91405 Orsay, France* (Ch. 2)

Métivier, G., *Institut de Mathématiques de Bordeaux Université de Bordeaux 1 Cours de la Libération 33405, Talence Cedex, France* (Ch. 3)

Mimura, M., *Department of Mathematics, School of Science and Technology, Meiji University, 1-1-1 Higashimita, Tamaku, Kawasaki 214-8571, Japan* (Ch. 2)

Ninomiya, H., *Department of Applied Mathematics and Informatics, Ryukoku University, Seta, Otsu, 520-2194, Japan* (Ch. 2)

Racke, R., *Fachbereich Mathematik und Statistik Universtaet Konstanz Fach D 187 78457 Konstanz, Germany* (Ch. 4)

Vessella, S., *Dipartimento di Matematica per le Decisioni via Cesare Lombroso, 6/17 50134 - Firenze (FI), Italy* (Ch. 5)

Contents of Volume I

Contents of Volume II

Contents of Volume III

Contents of Volume IV

CHAPTER 1

Shallow-Water Equations and Related Topics

Didier Bresch

UMR 5127 CNRS, LAMA, Université de Savoie, 73376 Le Bourget-du-Lac, France

Contents

HANDBOOK OF DIFFERENTIAL EQUATIONS
Evolutionary Equations, Volume 5
Edited by C.M. Dafermos and M. Pokorný

DOI: 10.1016/S1874-5717(08)00208-9

Abstract

These notes are devoted to the study of some problems related to shallow-water equations. In particular, we will review results from derivations of shallow-water systems which depend on a-priori conditions, the compressible-incompressible limit around a constant or inconstant height profile (quasi-geostrophic and lake equations), oscillating topography and boundary effects, variable depth vanishing on the shore, open boundary conditions, coupling models such as pollutants, sedimentation, multiphasic models.

Keywords: Shallow water, quasi-geostrophic equations, lake equations, oscillating topography, viscous parametrization, boundary effects, degenerate systems, singular perturbations, coupling models

1. Preface

These notes originate from lectures given during a Summer-School organized by Professor Éric LOMBARDI through the Groupement de Recherche: Équations d'Amplitudes et Propriétés Qualitatives in Roscoff 2004 (FRANCE) and during a post-graduate course given at the Institute of Mathematical Sciences (HONG-KONG) in November 2006 at the invitation of Professor ZHOUPING XIN. Their main objective is to present, at the level of beginners, an introduction to some problems related to shallow-water type equations coming from Navier–Stokes equations with free surface. In the author's opinion, shallow-water equations are not well understood from the modeling, numerical and mathematical points of view. The reason for this is not because shallow-water approximations are hard to find but rather because it is pretty easy to find a great many different ones. The various mathematical results and properties differ from one another, and the choice of one approximation in preference to another may depend on the problems that are to be treated. As an illustration, we will mention in the introduction two-different kinds of shallow-water system which depend on the boundary conditions assumed on the bottom with regard to Navier–Stokes equations with free surface. As a sequel to this, we will focus on these two approximations. We will present some structural properties, well-posed results, asymptotic analysis and open problems. We will also discuss miscellaneous possible extensions/questions such as: Time dependency of bottom, propagation of pollutants, multiphasic flows. We remark that it is impossible to be exhaustive when talking about shallow-water systems and, for instance, will not discuss water-waves and non-rotational flows even if really important results have been recently obtained on such topics. We also refer readers to [25], [27] and [28] and the book [165] with references cited-therein for non-viscous shallow-water type models.

It is a pleasure for me to thank Professors Constantine DAFERMOS and Milan POKORNÝ for their suggestions to write up these notes. I also warmly thank close collaborators and friends namely: L. CHUPIN, T. COLIN, B. DESJARDINS, E. FERNÁNDEZ-NIETO, D. GÉRARD-VARET, J.-M. GHIDAGLIA, M. GISCLON, E. GRENIER, F. GUILLEN-GONZALEZ, R. KLEIN, J. LEMOINE, G. MÉTIVIER, V. MILISIC, P. NOBLE, J. SIMON, M.A. RODRIGUEZ-BELLIDO. I have not forgotten the PDEs team in Chambéry (UMR CNRS 5127/Université de Savoie) namely C. BOURDARIAS, D. BUCUR, D. DUTYKH, S. GERBI, E. OUDET, C. ROBERT and P. VIGNEAUX with whom several mathematical studies are still in progress. Research by the author on the subjects which are presented here was supported by the région Rhône-Alpes in 2004 and by the French Ministry through the ACI: "Analyse mathématique des paramétrisations en océanographie" from 2004 to 2007 (thanks to the LJK and LAMA Secretaries: JUANA, CATHY, CLAUDINE, HÉLÈNE, NADINE). I also wish to thank the MOISE INRIA Project in Grenoble managed by Éric BLAYO.

Dedication. This review paper is dedicated to the memory of Professor Alexander V. KAZHIKHOV, one of the most inventive applied mathematician in compressible fluid mechanics. I also dedicate this chapter to my father Georges BRESCH and to my wife Catherine BRESCH-BAUTHIER.

2. Introduction

The shallow-water equations, generally, model free surface flow for a fluid under the influence of gravity in the case where the vertical scale is assumed to be much smaller than the horizontal scale. They can be derived from the depth-averaged incompressible Navier–Stokes equations and express conservation of mass and momentum. The shallow-water equations have applications to a wide range of phenomena other than water waves, e.g. avalanches and atmosphere flow. In the case of free surface flow when the shallow-water approximation is not valid, it is common to model the surface waves using several layers of shallow-water equations coupled via the pressure, see for instance [120], [121] and [159].

Viscous shallow-water type equations have attracted quite a lot of attention in recent years. The aim of this paper is to discuss several problems related to general forms of various shallow-water type equations coming from Navier–Stokes equations with free surface. Note that many other regimes can be investigated including for instance strong dispersive effects, see [7], however these will not be considered here.

For the reader's convenience, let us mention in this introduction two different shallow-water type systems mainly derived from two different bottom boundary condition choices for Navier–Stokes equations with free surface: friction boundary conditions and no-slip boundary conditions.

Friction boundary conditions. Let Ω be a two-dimensional space domain that will be defined later. A friction shallow-water system, with flat bottom topography, is written in Ω (see for instance J.-F. GERBEAU and B. PERTHAME [96], F. MARCHE [127]) as follows

$$\begin{cases} \partial_t h + \operatorname{div}(hv) = 0, \\ \partial_t(hv) + \operatorname{div}(hv \otimes v) = -h\dfrac{\nabla h}{\mathrm{Fr}^2} - h\dfrac{fv^\perp}{\mathrm{Ro}} \\ \qquad\qquad + \dfrac{2}{\mathrm{Re}}\operatorname{div}(hD(v)) + \dfrac{2}{\mathrm{Re}}\nabla(h\operatorname{div} v) - \mathrm{We}\, h\nabla\Delta h + \mathcal{D} \end{cases} \tag{1}$$

where h denotes the height of the free surface, v the vertical average of the horizontal velocity component of the fluid and f a function which depends on the latitude y in order to describe the variability of the Coriolis force. In the second equation, $\perp$ denotes a rotation of $\frac{\pi}{2}$, namely $G^\perp = (-G_2, G_1)$ when $G = (G_1, G_2)$. The space variable is denoted $\mathrm{x} = (x, y) \in \Omega$, the gradient operator $\nabla = (\partial_x, \partial_y)$ and the Laplacian operator $\Delta = \partial_x^2 + \partial_y^2$. The dimensionless numbers Ro, Re, Fr, We denote the Rossby number, the Reynolds number, the Froude number and the Weber number, respectively. Damping terms $\mathcal{D}$ coming from friction may also be included. We will discuss this point in Section 3 and also the Coriolis force assumption. In 1871, Saint-Venant wrote in a note, see [76], about a system which describes the flow of a river and corresponds to the inviscid shallow-water model written in System (7) neglecting the Coriolis force.

Equations (1), respectively, express the conservation of mass and momentum energy. System (1) is supplemented with initial conditions

$$h|_{t=0} = h_0, \qquad (hv)|_{t=0} = q_0 \quad \text{in } \Omega. \tag{2}$$

The functions h_0, q_0, are assumed to satisfy

$$h_0 \geq 0 \quad \text{a.e. on } \Omega, \quad \text{and} \quad \frac{|q_0|^2}{h_0} = 0 \quad \text{a.e. on } \{x \in \Omega / h_0(x) = 0\}. \tag{3}$$

The formal derivation of such systems from the Navier–Stokes equations with free boundary may be found in [127], generalizing a paper by J.-F. GERBEAU and B. PERTHAME, see [96]. Such formal derivations will be discussed in Section 3.9.

No-slip boundary conditions. Let Ω be a periodic box in two-dimensional space. Let us consider a thin, free surface, liquid down an inclined plane with a slope angle θ. Assuming no-slip boundary conditions on the bottom, we can prove, following word for word the interesting paper written by J.-P. VILA see [158], that a non-slip shallow-water system in Ω is given by:

$$\begin{cases} \partial_t h + \operatorname{div}(hv) = 0, \\ \partial_t(hv) + \operatorname{div}\left(\dfrac{6}{5} hv \otimes v\right) + \nabla\left(\dfrac{c\,h^2}{\mathrm{Re}} - \dfrac{(2s)^2}{75} h^5\right) \\ \qquad - \varepsilon^2 \mathrm{We}\, h\, \nabla\Delta h = \dfrac{1}{\varepsilon\,\mathrm{Re}}\left(2sh - \dfrac{3v}{h}\right) \end{cases} \tag{4}$$

where h denotes the height of the free surface, v is the vertical average of the horizontal velocity component of the fluid. The other quantities c, s, We, ε, Re are, respectively, $\cos\theta$, $\sin\theta$, the Weber coefficient (linked to surface tension), the aspect ratio of the domain and the Reynolds number. System (4) is also supplemented with initial conditions

$$h|_{t=0} = h_0, \qquad (hv)|_{t=0} = q_0 \quad \text{in } \Omega. \tag{5}$$

The functions h_0, q_0, are also assumed to satisfy

$$h_0 \geq 0 \quad \text{a.e. on } \Omega, \quad \text{and} \quad \frac{|q_0|^2}{h_0} = 0 \quad \text{a.e. on } \{x \in \Omega / h_0(x) = 0\}. \tag{6}$$

In this chapter, we will focus on the two kinds of shallow-water systems already defined, namely (1) and (4) in different horizontal domains Ω. We will try as much as possible to consider the validity of the formal asymptotic, some mathematical properties and we will also discuss several deduced or related models.

3. A friction shallow-water system

3.1. *Conservation of potential vorticity*

The inviscid case. The inviscid shallow-water equations coincide with the Euler equations of gas dynamics in the case of isentropic gas flow with a power pressure law

$p(\rho) = \rho^\gamma/2\mathrm{Fr}^2$ with $\gamma = 2$. Adding Coriolis force, it reads

$$\begin{cases} \partial_t h + \operatorname{div}(hv) = 0, \\ \partial_t(hv) + \operatorname{div}(hv \otimes v) + h\dfrac{\nabla h}{\mathrm{Fr}^2} + h\dfrac{f v^\perp}{\mathrm{Ro}} = 0. \end{cases} \tag{7}$$

Thus the 2×2 shallow-water equations often serve as a simpler test compared with the full 3×3 system of the Euler system. As a consequence we can expect all the difficulties of hyperbolic systems in classical mechanics: namely linear dispersive waves for small amplitudes, rarefaction, shock waves for a nonlinear regime.

Potential vorticity equation. For the two-dimensional inviscid shallow-water equations, the vorticity ω is a scalar, defined as

$$\omega = \partial_x v_2 - \partial_y v_1.$$

To derive an evolution equation for ω, we take the curl of the momentum equation (7)$_2$ divided by h, and we get

$$\begin{aligned} 0 &= \operatorname{curl}\left(\partial_t v + v \cdot \nabla v + f\frac{v^\perp}{\mathrm{Ro}}\right) \\ &= \partial_t \omega + v \cdot \nabla\omega + \left(\omega + \frac{f}{\mathrm{Ro}}\right)\operatorname{div} v + v_2 \frac{\partial_y f}{\mathrm{Ro}}. \end{aligned}$$

We can eliminate the term involving $\operatorname{div} v$ by multiplying the previous vorticity equation by h and the conservation of mass equation (7)$_1$ by $\omega + f/\mathrm{Ro}$ and subtracting them. This yields

$$h\left(\partial_t\left(\omega + \frac{f}{\mathrm{Ro}}\right) + v \cdot \nabla\left(\omega + \frac{f}{\mathrm{Ro}}\right)\right) - \left(\omega + \frac{f}{\mathrm{Ro}}\right)(\partial_t h + v \cdot \nabla h) = 0.$$

Therefore we get

$$\partial_t\left(\frac{\omega_R}{h}\right) + v \cdot \nabla\left(\frac{\omega_R}{h}\right) = 0, \quad \omega_R := \omega + \frac{f}{\mathrm{Ro}} \tag{8}$$

where ω_R is the *relative vorticity*. This equation means that the ratio of ω_R and the effective depth h is conserved along the particle trajectories of the flow. This constraint is called the *potential vorticity*, namely $\omega_P = \omega_R/h$. It provides a powerful constraint in large-scale motions of the atmosphere. If $\omega + f/\mathrm{Ro}$ is constant initially, the only way that it can remain constant at a latter time is if h itself is constant. In general, the conservation of potential vorticity tells us that if h increases then $\omega + f/\mathrm{Ro}$ must increase, and conversely, if h decreases, then $\omega + f/\mathrm{Ro}$ must also decrease. We refer the reader to the introductory book written by A. MAJDA for more information on the inviscid rotating shallow-water system, see [125]. Mathematically speaking, multiplying (8) by $\omega_R = \omega + f/\mathrm{Ro}$ and using the conservation of mass equation (7)$_1$ we have the following conservation equality

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega h\left|\frac{(\omega + f/\mathrm{Ro})}{h}\right|^2 = 0.$$

The viscous case. Let us now consider System (1) with h assumed to be constant equal to 1, $\mathcal{D} = 0$ and $f(y) = \beta y$ with β a constant. We obtain a divergence free velocity field v, namely div$v = 0$. Then there exists a stream function Ψ, such that $v = \nabla^\perp\Psi = (-\partial_y\Psi, \partial_x\Psi)$. A system with the relative vorticity $\omega + f$ and the stream function Ψ is easily written. It reads

$$\partial_t(\omega + f) + v\cdot\nabla(\omega + f) - \frac{1}{\mathrm{Re}}\Delta(\omega + f) = 0,$$
$$-\Delta\Psi = \omega, \qquad v = \nabla^\perp\Psi.$$

The reader interested by such cases is referred to [126].
When h is not constant, even when it does not depend on time, there are no simple equations for ω_R or ω_P. The term

$$\mathrm{curl}\left(\frac{1}{h}\left(\frac{2}{\mathrm{Re}}\mathrm{div}(hD(v)) + \frac{2}{\mathrm{Re}}\nabla(h\mathrm{div}\, v)\right)\right)$$

gives cross terms and readers are referred to [111,106] for relevant calculations.

As a consequence, even in the case where h is a given function b, depending only on x, we are not able to get global existence and uniqueness of a strong solution for this viscous model if we allow b to vanish on the shore. The lack of an equation for the vorticity also induces problems in the proof of convergence from the viscous case to the inviscid case when the Reynolds number tends to infinity when b degenerates close to the shore. This is an interesting mathematical open problem. For the non-degenerate case, the reader is referred to [106].

Other interesting questions are to analyse the domain of validity of viscous shallow-water equations in a bounded domain, and to determine relevant boundary conditions when h vanishes on the shore. Such a study has been done in the recent paper [39] when the shore is fixed.

3.2. *The inviscid shallow-water equations*

Inviscid shallow-water equations are not the topic of this Handbook. We refer the interested reader to [165,129,19,139]. We will only give some important properties for the sake of completeness.

(a) *Discontinuous shock waves.* Note that the inviscid shallow-water system is degenerate close to a vacuum. Sometimes, in one-dimensional space, authors portray an inviscid shallow-water system by the momentum equation divided by h, namely the following system

$$\begin{cases} \partial_t h + \partial_x(hv) = 0, \\ \partial_t v + v\partial_x v + \dfrac{1}{\mathrm{Fr}^2}\partial_x h = 0 \end{cases} \tag{9}$$

with a Froude number defined by $\mathrm{Fr} = v_{\mathrm{char}}/\sqrt{gh_{\mathrm{char}}}$ with v_{char} a characteristic velocity, h_{char} a characteristic height and g the gravity constant. Considering discontinuous shock waves, this system becomes

$$\partial_t U + A(U)\partial_x U = 0$$

with $U = \binom{h}{v}$ and $A(U) = \begin{pmatrix} v & h \\ 1/\mathrm{Fr}^2 & v \end{pmatrix}$ and the eigenvalues of $A(U)$ is given by

$$\lambda_\pm = v \pm \frac{1}{\mathrm{Fr}}\sqrt{h}.$$

When $\mathrm{Fr} < 1$ the flow is subcritical (fluvial) and subsonic and when $\mathrm{Fr} > 1$ it corresponds to a super-critical (torrential) and supersonic regime.

Shock waves are discontinuous solutions of hyperbolic systems

$$\partial_t U + \partial_x F(U) = 0$$

where

$$U(t,x) = \begin{cases} U_l & \text{for } x \le \sigma t, \\ U_r & \text{for } x \ge \sigma t \end{cases} \tag{10}$$

with the shock speed σ characterized by the Rankine-Hugoniot condition

$$\sigma[U_l - U_r] = [F(U_l) - F(U_r)].$$

In order to select a solution, one has to add an entropy condition corresponding to the energy dissipated by shocks and characterized by the following inequality

$$\partial_t S(U) + \partial_x \eta^S(U) \le 0$$

for convex entropy, namely

$$S(U) = \frac{1}{2}hv^2 + \frac{1}{\mathrm{Fr}^2}h^2, \qquad \eta^S(U) = \left[S(U) + g\frac{h^2}{2\mathrm{Fr}^2}\right]v.$$

The proof of existence follows DIPERNA analysis (1983) (see [80]) and is related to compensated compactness. It corresponds to the manipulation of two kinds of entropy

$$S_{\text{strong}} = \frac{1}{2}v^2 + \frac{1}{\text{Fr}^2}h, \qquad S_{\text{weak}} = \frac{1}{2}hv^2 + \frac{1}{\text{Fr}^2}h^2.$$

If we consider the inviscid shallow-water system without division by h, the velocity field is not defined when h vanishes. Only one entropy is available

$$S_{\text{weak}} = \frac{1}{2}hv^2 + \frac{1}{\text{Fr}^2}h^2$$

and its existence is not easy to prove. Global existence of an entropy solution in one dimension of System (9) is referred to by P.-L. LIONS, B. PERTHAME, P. E. SOUGANIDIS (1993), see [116]. Such existence follows from a regularization and limit process, Re tends to $+\infty$, in the following system

$$\begin{cases} \partial_t h + \partial_x(hv) = \dfrac{1}{\text{Re}}\partial_x^2 h, \\ \partial_t(hv) + \partial_x(hv^2) + \dfrac{1}{2\text{Fr}^2}\partial_x h^2 = \dfrac{1}{\text{Re}}\partial_x^2(hu). \end{cases} \tag{11}$$

Note that such a regularized model has, *a priori* no physical meaning. It could be an interesting open problem to justify a limit process on the following system

$$\begin{cases} \partial_t h + \partial_x(hv) = 0, \\ \partial_t(hv) + \partial_x(hv^2) + \dfrac{1}{2\text{Fr}^2}\partial_x h^2 = \dfrac{4}{\text{Re}}\partial_x(h\partial_x u) + \text{We}\, h\partial_x^3 h \end{cases} \tag{12}$$

when We tends to zero and Re tends to $+\infty$. This system corresponds to the viscous system with surface tension obtained in [96] (see also the nice paper [143]).

(b) *Klein–Gordon and linear shallow-water equations.* Let us consider the following inviscid shallow-water system linearized around the basic state $(H, 0)$ where H is a given constant

$$\begin{cases} \partial_t h + H \text{div}\, v = 0, \\ \partial_t v + \dfrac{1}{\text{Fr}^2}\nabla h + f\dfrac{v^\perp}{\text{Ro}} = 0. \end{cases} \tag{13}$$

Differentiating with respect to time, the mass equation gives

$$\partial_t^2 h = -H \text{div} \partial_t v$$

and therefore

$$\partial_t^2 h = H\left(f\frac{\text{div}\, v^\perp}{\text{Ro}} + \frac{1}{\text{Fr}^2}\Delta h\right).$$

Then we use the fact that div $v^\perp = \operatorname{rot} v = \omega$ to get

$$\partial_t^2 h = Hf\frac{\omega}{\mathrm{Ro}} + \frac{1}{\mathrm{Fr}^2} H\Delta h.$$

Since

$$\partial_t \omega = \frac{f}{H}\partial_t h$$

then

$$\partial_t^2(\partial_t h) = \frac{1}{\mathrm{Fr}^2} H\Delta(\partial_t h) - \frac{\mathrm{f}^2}{\mathrm{Ro}}\partial_t h.$$

Thus $\partial_t h$ satisfies a Klein–Gordon type equation. Using this information, it could be interesting to see if it is possible to get optimal lower bound dependency with respect to Coriolis and Froude numbers for the existence time corresponding to the inviscid rotating shallow-water equations. See, for instance, related works by J-M Delort [74] on Klein–Gordon equations.

(c) *Primitive equations and shallow-water system.* This part is written in order to show why shallow-water systems could be of particular interest to oceanographers. The primitive equations of the ocean (PEs) are one of the basic set of equations in oceanography, see for instance [119,153]. They are obtained using the hydrostatic approximation and Boussinesq approximations and scaling analysis in the 3D Navier–Stokes equations. The linear primitive equations are

$$\begin{cases} \partial_t v + f\dfrac{v^\perp}{\mathrm{Ro}} + \nabla_x p = 0, \\ \partial_z p = \theta, \\ \operatorname{div}_{x,y} v + \partial_z w = 0, \\ \partial_t \theta + N^2 w = 0, \end{cases} \tag{14}$$

where $u = (v, w)$ is the flow velocity with $v = (v_1, v_2)$, p the pressure and θ the buoyancy (sometimes referred to as the temperature), N is the Brunt-Väisälä frequency. Let us separate the variables as follows

$$(u, p) = (u(x, y, t), p(x, y, t))M(z). \tag{15}$$

Therefore

$$\theta = p(x, y, t)M'(z), \qquad w = -N^{-2}\partial_t p M'.$$

We insert in the incompressible condition $(14)_3$, it gives

$$\partial_x v_1 + \partial_y v_2 = \partial_t p N^{-2} M^{-1} M''.$$

Due to (14)$_1$ and Expression (15), the right-hand side in the previous equation has to be independent of z, then there exists a constant c_1, such that

$$N^{-2}M^{-1}M'' = c_1.$$

Assuming an initial condition $c_1 < 0$, then there exists m such that $c_1 = -m^2$ and we can rewrite the velocity equations

$$\partial_t v_1 - \frac{f v_2}{\mathrm{Ro}} = -\partial_x p, \qquad \partial_t v_2 + \frac{f v_1}{\mathrm{Ro}} = -\partial_y p,$$

and

$$\partial_t p + \frac{1}{m^2}(\partial_x v_1 + \partial_y v_2) = 0.$$

For $m \neq 0$, we get the shallow-water equation linearized around $H = 1/m^2$. Note that such decomposition is helpful for numerical purposes when considering the Primitive equations.

3.3. LERAY *solutions*

In this subsection, we will assume the domain Ω to be a periodic box in a two or three dimensional space (results may be obtained in the whole space or bounded domain cases). In his famous 1934 *Acta Mathematica paper* "Sur le mouvement d'un fluide visqueux emplissant l'espace", see [109], J. LERAY introduces the notion of a weak solution (nowadays also referred to as a LERAY solution) for incompressible flows and proves the global in time existence of such solutions in dimension $d = 2$ or 3, and global regularity and uniqueness in dimension $d = 2$. A somewhat simplified compressible case was considered by P-L LIONS in the 1990's, decoupling the internal energy equation from the mass and conservation of momentum equations, see [115] and references cited therein. Assuming barotropic pressure laws (temperature independent equations of state) and constant viscosities, $\mu \equiv \mathrm{cst}$, $\lambda \equiv \mathrm{cst}$, he proved the global existence of weak solutions in the spirit of Jean Leray's work on the following system

$$\begin{cases} \partial_t \rho + \mathrm{div}(\rho u) = 0, \\ \partial_t(\rho u) + \mathrm{div}(\rho u \otimes u) - \nabla p(\rho) + \mu \Delta u + (\lambda + \mu)\nabla \mathrm{div}\, u = 0. \end{cases} \tag{16}$$

Roughly speaking, the construction of weak solutions is based upon estimates induced by the physical energy of the system, supplemented by more subtle bounds on the effective flux, $F = p - (\lambda + 2\mu)\mathrm{div}\, u$, that allow control of density oscillations. This result mainly concerns the case $p(\rho) = a\rho^\gamma$ with

$$\gamma \geq \gamma_0, \quad \text{where } \gamma_0 = \frac{3}{2} \text{ if } d = 2, \qquad \gamma_0 = \frac{9}{5} \text{ if } d = 3.$$

Note that this result has been extended by E. FEIREISL et al. [87] to the case

$$\gamma > \frac{d}{2}$$

in [87] using oscillating defect measures on density sequences associated with suitable approximate solutions. Such an idea was first introduced in the spherical symmetric case by S. JIANG and P. ZHANG, see [105]. Let us remark here that System (1) is not of the form (16) since it concerns density dependent viscosities, λ and μ, and the mathematical tools developed by P.-L. LIONS and E. FEIREISL do not seem to be efficient to cover this case. The first mathematical results have been developed in [37] by using a new mathematical entropy that was introduced by the authors in [43]. It gives an answer related to the following shallow-water type system

$$\begin{cases} \partial_t h + \mathrm{div}(hv) = 0, \\ \partial_t (hv) + \mathrm{div}(hv \otimes v) = -h\dfrac{\nabla(h-b)}{\mathrm{Fr}^2} - h\dfrac{f v^\perp}{\mathrm{Ro}} \\ \quad + \dfrac{2}{\mathrm{Re}}\mathrm{div}(hD(v)) - \mathrm{We}\, h\nabla\Delta h + \mathcal{D} \end{cases} \tag{17}$$

which include some drag terms $\mathcal{D}$. The following section gives a generalization of the mathematical entropy discovered for the first time with $\mu(\rho) = \rho$ and $\lambda(\rho) = 0$ in [43].

3.3.1. A new mathematical entropy: The BD entropy We introduce in this section a new mathematical entropy that has been recently discovered, in its general form, in [36] for more general compressible flows, namely the Korteweg system. The domain Ω being considered is either a periodic domain or the whole space. If the capillarity coefficient σ is taken equal to 0 in [36], namely the Weber number is equal to 0, the considered system may be written in the following form

$$\begin{cases} \partial_t \rho + \mathrm{div}(\rho u) = 0, \\ \partial_t (\rho u) + \mathrm{div}(\rho u \otimes u) \\ \quad = -\nabla p(\rho) + 2\mathrm{div}(\mu(\rho)D(u)) + \nabla(\lambda(\rho)\mathrm{div}\, u). \end{cases} \tag{18}$$

The energy identity for such a system is

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega (\rho|u|^2 + 2\pi(\rho)) + \int_\Omega \mu(\rho)|D(u)|^2 + \lambda(\rho)|\mathrm{div}\, u|^2 = 0$$

where π is the potential given by $\pi(\rho) = \rho\int_{\overline{\rho}}^{\rho} p(s)/s^2 \mathrm{d}s$ with $\overline{\rho}$ a constant reference density.

In [36], a new mathematical entropy has been discovered that helps to get a lot of mathematical results about compressible flows with density dependent viscosities. Namely if $\lambda(\rho) = 2(\mu'(\rho)\rho - \mu(\rho))$, then the following equality holds

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega (\rho|u + 2\nabla\varphi(\rho)|^2 + 2\pi(\rho)) \\ + \int_\Omega \frac{p'(\rho)\mu'(\rho)}{\rho}|\nabla\rho|^2 + \int_\Omega \rho|A(u)|^2 = 0$$

where $A(u) = (\nabla u - {}^t\,\nabla u)/2$ and $\rho\varphi'(\rho) = \mu'(\rho)$. For the reader's convenience, we recall here the alternate proof given in [33].

Proof of the new mathematical entropy identity. Using the mass equation we know that for all smooth function $\xi(\cdot)$

$$\partial_t\nabla\xi(\rho) + u\cdot\nabla\nabla\xi(\rho) + \nabla u\cdot\nabla\xi(\rho) + \nabla(\rho\xi'(\rho)\mathrm{div}\, u) = 0.$$

Thus, using once more the mass equation, we see that $v = \nabla\xi(\rho)$ satisfies:

$$\partial_t(\rho v) + \mathrm{div}(\rho u\otimes v) + \rho\nabla u\cdot\nabla\xi(\rho) + \rho\nabla(\rho\xi'(\rho)\mathrm{div}\, u) = 0$$

which gives, using the momentum equation on u,

$$\partial_t(\rho(u+v)) + \mathrm{div}(\rho u\otimes(u+v)) - 2\mathrm{div}(\mu(\rho)D(u)) - \nabla(\lambda(\rho)\mathrm{div}\, u) \\ + \nabla p(\rho) + \rho\nabla u\cdot\nabla\xi(\rho) + \rho\nabla(\rho\xi'(\rho)\mathrm{div}\, u) = 0.$$

Next we write the diffusion term as follows:

$$-2\mathrm{div}(\mu(\rho)D(u)) = -2\mathrm{div}(\mu A(u)) - 2\nabla u\cdot\nabla\mu \\ - 2\nabla(\mu\mathrm{div}\, u) + 2\nabla\mu\mathrm{div}\, u$$

where $A(u) = (\nabla u - {}^t\,\nabla u)/2$. Therefore, the equation for $u + v$ reads

$$\partial_t(\rho(u+v)) + \mathrm{div}(\rho u\otimes(u+v)) - 2\mathrm{div}(\mu(\rho)A(u)) \\ -2\mu'(\rho)\nabla u\cdot\nabla\rho - 2\nabla(\mu(\rho)\mathrm{div}\, u) + 2\mu'(\rho)\nabla\rho\mathrm{div} u + \nabla p(\rho) \\ - \nabla(\lambda(\rho)\mathrm{div}\, u) + \rho\xi'(\rho)\nabla u\cdot\nabla\rho \\ + \nabla(\rho^2\xi'(\rho)\mathrm{div}\, u) - \rho\xi'(\rho)\nabla\rho\mathrm{div}\, u = 0.$$

This equation may be simplified to give

$$\partial_t(\rho(u+v)) + \mathrm{div}(\rho u\otimes(u+v)) - 2\mathrm{div}(\mu(\rho)A(u)) + \nabla p(\rho) \\ + \nabla((\rho^2\xi'(\rho) - 2\mu(\rho) - \lambda(\rho))\mathrm{div}\, u) + (\rho\xi'(\rho) - 2\mu'(\rho))\nabla u\cdot\nabla\rho \\ + (2\mu'(\rho) - \rho\xi'(\rho))\nabla\rho\,\mathrm{div}\, u = 0.$$

If we choose ξ such that $2\mu'(\rho) = \xi'(\rho)\rho$, then $\lambda(\rho) = \xi'(\rho)\rho^2 - 2\mu(\rho)$ and the last three terms cancel, implying:

$$\partial_t(\rho(u+v)) + \operatorname{div}(\rho u \otimes (u+v)) - \operatorname{div}(\mu(\rho)A(u)) + \nabla p(\rho) = 0.$$

Multiplying this equation by $(u+v)$ and the mass equation by $|u+v|^2/2$ and adding, we easily get the new mathematical entropy equality. We just have to observe that

$$\int_\Omega \operatorname{div}(\mu(\rho)A(u)) \cdot v = 0$$

since v is a gradient. Using the continuity equation and the relation $p(\rho) = \rho\pi'(\rho) - \pi(\rho)$, the term $\nabla p(\rho)$ gives

$$\begin{aligned}\int_\Omega \nabla p(\rho) \cdot (u+v) &= \int_\Omega \rho \nabla \pi'(\rho) \cdot u + \int_\Omega \frac{p'(\rho)\mu'(\rho)}{\rho} |\nabla \rho|^2 \\ &= \frac{\mathrm{d}}{\mathrm{d}t} \int_\Omega \pi(\rho) + \int_\Omega \frac{p'(\rho)\mu'(\rho)}{\rho} |\nabla \rho|^2\end{aligned}$$

where $\pi(\rho) = \rho \int_{\overline{\rho}}^{\rho} p(s)/s^2 ds$ with $\overline{\rho}$ a constant reference density.

We remark that the mathematical entropy estimate gives extra information on ρ, namely

$$\mu'(\rho)\nabla\rho/\sqrt{\rho} \in L^\infty(0,T;L^2(\Omega))$$

assuming $\mu'(\rho_0)\nabla\rho_0/\sqrt{\rho_0} \in L^2(\Omega)$ initially.

Such information is crucial if we want to look at viscous compressible flows with density dependent viscosities and may help in various cases. Recent applications of such information have been given. For instance, in [131], A. MELLET and A. VASSEUR study the stability of isentropic compressible Navier–Stokes equations with the barotropic pressure law $p(\rho) = a\rho^\gamma$ with $\gamma > 1$ and a a given positive constant in dimension $d = 1, 2$ and 3. This result is particularly interesting, recalling that in constant case viscosities, E. FEIREISL's result covers the range $\gamma > d/2$.

An other interesting result concerns the full compressible Navier–Stokes equations which are equations for heat conducting flows. Existence of global weak solutions has been recently obtained in [35] assuming the perfect gas law except close to a vacuum where cold pressure is used to get compactness on the temperature. This completes the result obtained by E. FEIREISL in [85] where the temperature satisfies an inequality instead of an equality in the distribution space $\mathcal{D}'$ and where the perfect gas law does not satisfies the assumption even far from a vacuum. Note the two results do not cover the same range of viscosities. Something more has to be understood.

We remark that the relation between λ and μ

$$\lambda(\rho) = 2(\mu'(\rho)\rho - \mu(\rho)) \tag{19}$$

and the conditions

$$\mu(\rho) \geq c, \qquad \lambda(\rho) + 2\mu(\rho)/d \geq 0$$

can not be fulfilled simultaneously. Indeed, the viscosity μ has to vanish close to a vacuum. In [103], D HOFF and D SERRE prove, in the one-dimensional case, that only viscosities that vanish with density may prevent failure of continuous dependence on initial data for the Navier–Stokes equations of compressible flow. Our relation between λ and μ struggles to achieve such degenerate viscosities. We refer the reader to [131] for some interesting mathematical comments on the relation imposed between λ and μ. We also refer the readers to [30], a Handbook concerning the compressible Navier–Stokes equations: BRESCH-DESJARDINS, DANCHIN, FEIREISL, HOFF, KAZHIKHOV-VAIGANT, LIONS, MATSUMURA-NISHIDA, XIN results will be discussed.
Remark on the viscous shallow-water equations. We stress that, in the equation written at the beginning of this paper, the viscous term does not satisfy the conditions imposed above. Indeed, we have $\mu(\rho) = \rho$ but $\lambda \not\equiv 0$. As a consequence, the usual viscous friction shallow-water equations are far from being solved for weak solutions except in 1D where $-2\nu \mathrm{div}(hD(v)) - 2\nu\nabla(h\mathrm{div}\, v) = -4\nu\partial_x(h\partial_x v)$. Note that in some of the literature 4ν is called the Trouton viscosity, see [96].
Equations with capillarity. Capillarity can be taken into account, as shown in [36], by introducing a term of the form $-\rho\nabla(G'(\rho)\Delta G(\rho))$. It suffices to choose a viscosity equal to $\mu(\rho) = G(\rho)$. This gives the extra mathematical entropy

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega (\rho|u + 2\nabla\varphi(\rho)|^2 + 2\pi(\rho)) + \int_\Omega \frac{p'(\rho)\mu'(\rho)}{\rho}|\nabla\rho|^2 + \mathrm{We}\int_\Omega \mu'(\rho)|\Delta\mu(\rho)|^2 + \int_\Omega \rho|A(u)|^2 = 0$$

where $A(u) = (\nabla u - {}^t\nabla u)/2$ and $\rho\varphi'(\rho) = \mu'(\rho)$. Applications of such mathematical entropy will be provided in [42], to approximations of hydrodynamics. Note that for the shallow-water system, we find the usual surface tension term $h\nabla\Delta h$.
The temperature dependent case. For the full compressible Navier–Stokes equation, the mathematical energy equality reads

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega (\rho|u|^2 + \pi(\rho)) + \int_\Omega \mu(\rho)|D(u)|^2 + \int_\Omega \lambda(\rho)|\mathrm{div}\, u|^2 = \int_\Omega \nabla p \cdot u.$$

and the new mathematical entropy equality reads

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega (\rho|u + v|^2 + \pi(\rho)) + \int_\Omega \mu(\rho)|A(u)|^2 = \int_\Omega \nabla p \cdot (u + v)$$

where $A(u) = (\nabla u - {}^t\nabla u)/2$. To close the estimate, it is sufficient to prove that the extra terms $\int_\Omega \nabla p \cdot u$ and $\int_\Omega \nabla p \cdot (u + v)$ can be controlled by the terms on the left-hand side. This has been done in [34,35], implying the global existence of weak solutions for the full compressible Navier–Stokes equations with pressure laws, including a cold pressure part to control density close to a vacuum. The new mathematical entropy helps to control the density far from a vacuum.

3.3.2. Weak solutions with drag terms When drag terms such as $\mathcal{D} = r_0 v + r_1 h|v|v$ are present, the existence of global weak solutions for System (17) is proved in [37] without the capillarity term (We $= 0$, $r_1 > 0$ and $r_0 > 0$). The initial data are taken in such a way that

$$\begin{aligned} &h_0 \in L^2(\Omega), \qquad \frac{|q_0|^2}{h_0} \in L^1(\Omega), \\ &\nabla\sqrt{h_0} \in (L^2(\Omega))^2, \qquad -r_0 \log_- h_0 \in L^1(\Omega) \end{aligned} \tag{20}$$

where $q_0 = 0$ on $h_0^{-1}(\{0\})$ and $\log_- g = \log\min(g, 1)$.

More precisely, they give the following result

THEOREM 3.1. *Let Ω be a two-dimensional periodic box. Let q_0, h_0 satisfy* (3) and (20) *and assume that $r_1 > 0$ and $r_0 > 0$. Then, there exists a global weak solution of* (5) *and* (17).

Idea of the proof. Drag terms are helpful from a mathematical viewpoint since they give extra information on u and enable the proof of stability results. In [37], the authors use Fourier projector properties to prove the compactness. For the reader's convenience we give here the main idea and in the next section we will give a simpler proof based on the one developed in [131]. More precisely for any $k \in \mathbf{N}$, the kth Fourier projector P_k is defined on $L^2(\Omega)$ as follows: If $\sum_{\ell\in\mathbf{Z}^2} c_\ell \exp(i\ell\cdot x)$ denotes the Fourier decomposition of $f \in L^2(\Omega)$, then $P_k f$ is given as the low frequency part of it, $\sum_{|\ell|\le k} c_\ell \exp(i\ell\cdot x)$. The following classical estimate will be useful in what follows

$$\|f - P_k f\|_{L^2(\Omega)} \le \frac{C_p}{k^{2/p-1}}\|\nabla f\|_{L^p(\Omega)} \quad \text{for all } p \in (1, 2).$$

As a matter of fact, introducing $\beta \in \mathcal{C}^\infty(\mathbf{R})$ such that $0 \le \beta \le 1$, $\beta(s) = 0$ for $s \le 1$ and $\beta(s) = 1$ for $s \ge 2$, we obtain the following estimate, denoting $\beta_\alpha(\cdot) = \beta(\cdot/\alpha)$ for any positive number α

$$\begin{aligned} &\|\sqrt{h_n}u_n - P_k(\sqrt{h_n}u_n)\|_{L^2(0,T;(L^2(\Omega))^2)} \\ &\quad \le c\sqrt{\alpha}\|u_n\|_{L^2(0,T;(L^2(\Omega))^2)} + \frac{C_p}{k^{2/3}}\|\nabla\beta_\alpha(h_n)\sqrt{h_n}u_n\|_{L^2(0,T;(L^{6/5}(\Omega))^2)}. \end{aligned} \tag{21}$$

Note that the gradient can be estimated by

$$
\begin{aligned}
&\|\nabla \beta_\alpha(h_n)\sqrt{h_n}u_n\|_{L^2(0,T;(L^{6/5}(\Omega))^2)} \\
&\quad \le C\|\beta_\alpha(h_n)\sqrt{h_n}\nabla u_n\|_{L^2(0,T;(L^2(\Omega))^4)} \\
&\qquad + \|h_n^{1/3}u_n\|_{L^3(0,T;(L^3(\Omega)^2))}\|\nabla\sqrt{h_n}\|_{L^\infty(0,T;(L^2(\Omega)^2))} \times \|2\beta'_\alpha(h_n)h_n^{2/3} \\
&\qquad + \beta_\alpha(h_n)h_n^{-1/3}\|_{L^\infty(0,T;L^\infty(\Omega))}. \qquad (22)
\end{aligned}
$$

Thus the left-hand side may be estimated by

$$
C\sqrt{\alpha} + \frac{C_\alpha}{k^{2/3}}
$$

where the two above constants do not depend on n. It means that the high frequency of $\sqrt{h_n}u_n$ is arbitrarily small in $L^2(0,T;(L^2(\Omega))^2)$ uniformly in n for a large enough wave number k. It remains to consider the convergence of the product $P_k(\sqrt{h_n}u_n)\cdot\sqrt{h_n}u_n$ for a given $k \in \mathbf{N}$. Again, using a cutoff function such as β_α together with the uniform $L^2(0,T;(L^2(\Omega))^2)$ estimate on u_n, we only have to consider the weak limit of $P_k(\sqrt{h_n}u_n)h_nu_n$. Finally, observing that the momentum equation yields uniform $L^2(0,T;(H^{-s}(\Omega))^2)$ bounds on $\partial_t(h_nu_n)$ for a large enough s, we deduce the strong $L^2(0,T;(L^2(\Omega))^2)$ convergence of $\sqrt{h_n}u_n$ to $\sqrt{h}u$.

REMARK. The authors mentioned that in 1D, r_1 may be taken equal to 0. Indeed, their proof is based on the fact that they are able to neglect the high frequency part of $\sqrt{h_n}u_n$ uniformly in n. In one dimension, we can write $\partial_x(\sqrt{h_n}u_n) = \sqrt{h_n}\partial_x u_n + u_n\partial_x\sqrt{h_n}$ which means that $\partial_x(\sqrt{h_n}u_n)$ is uniformly bounded in $L^2(0,T;(L^2(\Omega))^2)$. Sobolev embeddings in dimension one then allow us to prove this crucial cutoff estimate.

Important Remark. We will see in the next section, a very nice idea due to A. MELLET and A. VASSEUR allowing us to ignore drag terms in the stability process. However, this idea does not seem to be compatible with the approximate solutions constructed in [38]. Without drag terms, building an adequate approximate scheme is an important open problem in dimensions greater than 2. Anyone who succeeds in defining such a regularized system will be the first to find the global existence of a weak solution result for barotropic Navier–Stokes equations with $p(\rho) = a\rho^\gamma$ where $\gamma > 1$ in a dimension space that could be equal to 3.

Approximate scheme. An approximate solutions construction for some degenerate compressible systems has been given in [38]. For the friction shallow-water system, it is based on the following regularized system

$$
\begin{cases}
\partial_t h_{\varepsilon,\eta} + \operatorname{div}(h_{\varepsilon,\eta}u_{\varepsilon,\eta}) = 0, \\
\partial_t(h_{\varepsilon,\eta}u_{\varepsilon,\eta}) + \operatorname{div}(h_{\varepsilon,\eta}u_{\varepsilon,\eta}\otimes u_{\varepsilon,\eta}) - \nu\operatorname{div}(h_{\varepsilon,\eta}\nabla u_{\varepsilon,\eta}) + h_{\varepsilon,\eta}\nabla h_{\varepsilon,\eta} \\
\quad = \varepsilon h_{\varepsilon,\eta}\nabla\Delta^{2k+1}h_{\varepsilon,\eta} + \varepsilon\nabla\log h_{\varepsilon,\eta} - \eta\Delta^2 u_{\varepsilon,\eta}.
\end{cases}
$$

Using this system, the height remains far from zero and the velocity may be regular enough. It remains then first to let η go to zero while keeping ε fixed and then let ε go to zero using the stability procedure.

For the reader's convenience, let us provide the corresponding energy equality and BD entropy for such a regularized system in order to see the dependency of the coefficients ε, η:

Energy equality. The energy identity for such a regularized system is

$$\begin{aligned}
&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega (h_{\varepsilon,\eta}|u_{\varepsilon,\eta}|^2 + |h_{\varepsilon,\eta}|^2) + \frac{\varepsilon}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega |\nabla^{2k+1} h_{\varepsilon,\eta}|^2 \\
&\quad + \varepsilon\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \log h_{\varepsilon,\eta} + 2\int_\Omega h_{\varepsilon,\eta}|D(u_{\varepsilon,\eta})|^2 + \int_\Omega \eta|\Delta u_{\varepsilon,\eta}|^2 = 0.
\end{aligned}$$

BD mathematical entropy. The BD mathematical entropy reads

$$\begin{aligned}
&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega (h_{\varepsilon,\eta}|u_{\varepsilon,\eta} + 2\nabla \log h_{\varepsilon,\eta}|^2 \\
&\quad + |h_{\varepsilon,\eta}|^2) + 2\int_\Omega |\nabla h_{\varepsilon,\eta}|^2 + 2\int_\Omega h_{\varepsilon,\eta}|A(u_{\varepsilon,\eta})|^2 \\
&\quad + 2\varepsilon\int_\Omega |\nabla \log h_{\varepsilon,\eta}|^2 + \varepsilon\int_\Omega |\Delta^{k+1} h_{\varepsilon,\eta}|^2 \\
&\qquad = \eta\int_\Omega \Delta u_{\varepsilon,\eta}\cdot\nabla\Delta\log h_{\varepsilon,\eta}
\end{aligned} \tag{23}$$

where $A(u_{\varepsilon,\eta}) = (\nabla u_{\varepsilon,\eta} -{}^t\,\nabla u_{\varepsilon,\eta})/2$. Let us observe that the energy estimate allows us to prove that for given $\varepsilon > 0$, $h_{\varepsilon,\eta}$ is bounded and bounded away from zero uniformly in η. On the other hand

$$\nabla\Delta\log h = \frac{\Delta\nabla h}{h} - \frac{2(\nabla h\cdot\nabla)\nabla h}{h^2} - \frac{\Delta h\nabla h}{h^2} + \frac{2|\nabla h|^2\nabla h}{h^3}.$$

Thus

$$\begin{aligned}
&\eta\left|\int_\Omega \Delta u_{\varepsilon,\eta}\cdot\nabla\Delta\log h_{\varepsilon,\eta}\right| \\
&\quad \le \sqrt{\eta}\|\sqrt{\eta}\Delta u_{\varepsilon,\eta}\|_{L^2}\left(1 + \|h_{\varepsilon,\eta}\|_{H^s}\right)^3\left(1 + \|h^{-1}_{\varepsilon,\eta}\|_{L^\infty}\right)^3
\end{aligned}$$

using the following lemma based on Sobolev embedding inequalities

LEMMA 3.2. *Let $n \in \mathbf{N}$ and $\sigma > n + d/2$. There exists $C > 0$ such that for all $h \in H^\sigma(\Omega)$ such that $h^{-1} \in L^{n+1}(\Omega)$, one has*

$$\|\nabla^n h^{-1}\|_{L^2(\Omega)} \le C\left(1 + \|h^{-1}\|_{L^{n+1}(\Omega)}\right)^{n+1}\left(1 + \|h\|_{H^\sigma(\Omega)}\right)^n.$$

This result allows us to control the right-hand side in (23) uniformly, with respect to η, when ε is fixed.

3.3.3. Forgetting drag terms – Stability A. MELLET and A. VASSEUR, in the beautiful paper [131], show how to ignore the drag terms for the shallow-water equations without capillary terms. This is useful for the mathematical analysis to get stability results. Indeed, controlling $\nabla\sqrt{\rho}$ implies further information on u assuming this regularity initially. Note that their analysis works for the general barotropic compressible system (18) with the viscosities relation (15). It concerns the whole space $\Omega = \mathbf{R}^d$ and periodic box cases $\Omega = \mathbf{T}^d$ ($d = 1, 2, 3$). More precisely, multiplying the momentum equation $(18)_2$ by $(1 + \ln(1 + |u|^2))u$, using the mass equation $(18)_1$ and assuming that

$$\int \rho_0 \frac{1 + |u_0|^2}{2} \ln(1 + |u_0|^2)\,\mathrm{d}x < C,$$

one proves that $\rho(1 + |u|^2)\ln(1 + |u|^2) \in L^\infty(0, T; L^1(\Omega))$ using the estimates given by the new mathematical entropy. This new piece of information is sufficient to pass to the limit in the nonlinear term without the help of drag terms. Note that such a technique holds for general μ and λ and satisfies the relation needed for the new mathematical entropy.

For the reader's convenience, let us give in the following the nice estimate due to A. MELLET and A. VASSEUR and the strong convergence of $\sqrt{\rho_n}u_n$: key tools in the stability proof.

A new estimate based on the BD entropy. Multiply the momentum equation $(18)_2$ by $(1 + \ln(1 + |u|^2))u$ and the mass equation $(18)_1$ by $\ln(1 + |u|^2)(1 + |u|^2)/2$, sum the results and integrate over Ω. If $d\lambda + 2\mu \geq \nu\mu$ where d is the space dimension, we easily get the following estimate

$$\begin{aligned}
&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \rho(1 + |u|^2)\ln(1 + |u|^2) + \nu\int_\Omega \mu(\rho\ln(1 + |u|^2))|D(u)|^2 \\
&\quad \leq -\int_\Omega (1 + \ln(1 + |u|^2))u \cdot \nabla\rho^\gamma + c\int_\Omega \mu(\rho)|\nabla u|^2.
\end{aligned}$$

The first term in right-hand side is controlled; integrating by parts, using the fact that

$$(\operatorname{div} u)^2 \leq d|D(u)|^2$$

and by the Cauchy–Schwartz inequality, we get

$$\begin{aligned}
&\int_\Omega (1 + \ln(1 + |u|^2))u \cdot \nabla\rho^\gamma \leq \int_\Omega \mu(\rho)|\nabla u|^2 \\
&\quad + \frac{\nu}{2}\int_\Omega \mu(\rho)(1 + \ln(1 + |u|^2))|D(u)|^2 + c\int_\Omega (2 + \ln(1 + |u|^2))\frac{\rho^{2\gamma}}{\mu(\rho)}.
\end{aligned}$$

Then we remark that

$$\int_\Omega (2+\ln(1+|u|^2))\frac{\rho^{2\gamma}}{\mu(\rho)}$$
$$\le c\left(\int_\Omega\left(\frac{\rho^{2\gamma-\delta/2}}{\mu(\rho)}\right)^{2/(2-\delta)}\right)^{(2-\delta)/2}\left(\int_\Omega \rho(2+\ln(1+|u|^2))^{2/\delta}\right)^{\delta/2}.$$

We then finally deduce the following inequality

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \rho(1+|u|^2)\ln(1+|u|^2)+\frac{\nu}{2}\int_\Omega \mu(\rho(1+\ln(1+|u|^2)))|D(u)|^2$$
$$\le c\int_\Omega \mu(\rho)|\nabla u|^2 + c\left(\int_\Omega\left(\frac{\rho^{2\gamma-\delta/2}}{\mu(\rho)}\right)^{2/(2-\delta)}\right)^{(2-\delta)/2}$$
$$\times\left(\int_\Omega \rho(2+\ln(1+|u|^2))^{2/\delta}\right)^{\delta/2}. \tag{24}$$

Since $\mu(\rho)|\nabla u|^2$ is controlled using the BD entropy, $\rho|u|^2$ and ρ are bounded in $L^\infty(0,T;L^1(\Omega))$, this controlled the first and third terms in the right-hand side. Now using the hypothesis on $\mu(\rho)$ and information on ρ obtained through the BD entropy, the second term is controlled when δ is small enough and there is some restriction on γ in 3D: $(2\gamma-1<5\gamma/3)$.

Idea of the stability proof. The stability proof is based on the strong convergence of $\sqrt{\rho}_n u_n$ in $L^2(0,T;(L^2(\Omega))^d)$. Thus, we look at

$$\int_0^T\int_\Omega |\sqrt{\rho_n}u_n-\sqrt{\rho}u|^2 \le \int_0^T\int_\Omega |\sqrt{\rho_n}u_n 1_{|u_n|\le M}-\sqrt{\rho}u 1_{|u|\le M}|^2$$
$$+2\int_0^T\int_\Omega |\sqrt{\rho_n}|u_n 1_{|u_n|\ge M}|^2$$
$$+2\int_0^T\int_\Omega |\sqrt{\rho}u 1_{|u|\ge M}|^2.$$

It is obvious that $\sqrt{\rho_n}u_n 1_{|u_n|\le M}$ is bounded uniformly in $L^\infty(0,T;L^2(\Omega))$, so

$$\sqrt{\rho_n}u_n 1_{|u_n|\le M}\to\sqrt{\rho}u 1_{|u|\le M}\ \text{almost everywhere.}$$

gives the convergence of the first integral

$$\int_0^T\int_\Omega |\sqrt{\rho_n}u_n 1_{|u_n|\le M}-\sqrt{\rho}u 1_{|u|\le M}|^2\,\mathrm{d}x\,\mathrm{d}t\to 0.$$

Finally, we write

$$\int_0^T \int_\Omega |\sqrt{\rho_n}|u_n 1_{|u_n|\geq M}|^2 \, dx \, dt \leq \frac{1}{\ln(1+M^2)} \int_0^T \int_\Omega \rho_n u_n^2 \ln(1+|u_n|^2) \, dx \, dt$$

and

$$\int_0^T \int_\Omega |\sqrt{\rho}|u 1_{|u|\geq M}|^2 \, dx \, dt \leq \frac{1}{\ln(1+M^2)} \int_0^T \int_\Omega \rho u^2 \ln(1+|u|^2) \, dx \, dt.$$

Putting together the three previous estimates, we deduce

$$\lim \sup_{n\to+\infty} \int_0^T \int_\Omega |\sqrt{\rho_n} u_n - \sqrt{\rho} u|^2 \, dx \, dt \leq \frac{C}{\ln(1+M^2)}$$

for all $M > 0$ and the lemma follows by taking $M \to +\infty$.

REMARK. The problem is now to build regular approximate solutions that satisfy the energy estimate, the new mathematical entropy and the estimate on $\rho(1+|u|^2)\ln(1+|u|^2)$. This is an open problem in dimension space greater than 2 without symmetry assumptions. In dimension space equal to one or assuming some space symmetry, construction of approximate solutions has been realized in [132,113,100].

REMARK. We note that the addition of the drag $h|u|u$ gives the bound $h_n|u_n|^3$ which belongs to $L^1(0,T;L^1(\Omega))$, therefore the strong convergence of $\sqrt{h_n}u_n$ in $L^2(0,T;(L^2(\Omega))^2)$ may be obtained, as previously, by replacing the extra regularity established in Mellet-Vasseur's paper by this regularity.

REMARK. Navier–Stokes systems for compressible fluids with density dependent viscosities are considered in [113]. As mentioned in this Handbook, these equations, in particular, include the shallow-water system formally derived from the Navier–Stokes system with free surface and friction bottom conditions (friction shallow-water model). In this paper, it is shown that for any global entropy weak solution, any (possible existing) vacuum state must vanish within finite time. The velocity (even if regular enough and well defined) blows up in finite time as the vacuum states vanish. Furthermore, after the vanishing of vacuum states, the global entropy weak solution becomes a strong solution and tends to the non-vacuum equilibrium state exponentially in time. This last result allows, by the use of the BD entropy equality, to recover the asymptotic stability of steady states established in [124] using the associated Green function.

3.3.4. Bounded domains In [39], the authors have considered the bounded domain case in two or three dimensions in space and barotropic compressible Navier–Stokes equations with turbulent terms. Namely, the following system

$$\begin{cases} \partial_t \rho + \operatorname{div}(\rho u) = 0, \\ \partial_t(\rho u) + \operatorname{div}(\rho u \otimes u) \\ \quad = -\nabla p(\rho) + 2\operatorname{div}(\mu(\rho)D(u)) + \nabla(\lambda(\rho)\operatorname{div} u) - r_1\rho|u|u \end{cases} \tag{25}$$

with the viscosities relation (19). For the sake of simplicity, we present here the case $\mu(\rho) = \rho$ and $\lambda(\rho) = 0$. In this case, the boundary conditions are either the Dirichlet boundary condition

$$\rho u = 0 \quad \text{on } \partial\Omega,$$

on the momentum, or Navier's condition of the type

$$\rho u \cdot n = 0, \qquad \rho(D(u)n)_\tau = -\alpha\rho u_\tau, \quad \text{on } \partial\Omega$$

where α is a positive constant and n the outward unit normal. We recall that the previous condition is a friction condition, showing that the shear force at the boundary is proportional to the tangential velocity. It is widely used in the simulation of geophysical flows, and similar wall laws may be justified in the framework of rough boundaries. In order to preserve the BD entropy, an additional boundary condition on the density is involved, namely that

$$\nabla\rho \times n = 0$$

where $\cdot \times \cdot$ is the vectorial product and n is the exterior unit normal. This condition indicates that the density should be constant on each connected component of the boundary. Note that System (25) is degenerate which means the viscosity vanishes when the density tends to 0. Boundary conditions on ρu and $\nabla\rho$ are well defined using that

$$\begin{aligned}
&\rho u = \sqrt{\rho}(\sqrt{\rho}u) \in L^\infty(0, T; (L^1(\Omega))^2),\\
&\nabla(\rho u) = \sqrt{\rho}(\sqrt{\rho}\nabla u) + \nabla\sqrt{\rho} \otimes (\sqrt{\rho}u) \in L^2(0, T; (L^1(\Omega))^4),\\
&\rho\nabla \log\rho \in L^\infty(0, T; (L^1(\Omega))^2), \qquad \mathrm{curl}(\rho\nabla\log\rho) = 0.
\end{aligned}$$

Concerning the Navier-type condition, we use the following proposition.

PROPOSITION 3.3. *Let $u \in \mathcal{C}^\infty(\overline{\Omega})^d$ with $d = 2, 3$, satisfying $u \cdot n = 0$. Then*

$$(D(u)n)_\tau = \frac{1}{2}\left(\frac{\partial u}{\partial n}\right)_\tau - \frac{1}{2}\kappa(x)u_\tau$$

where κ is the scalar curvature for $d = 2$, the Weingarten endomorphism of $\partial\Omega$ for $d = 3$.

Next, we can derive the two estimates needed in the stability proof:

- The classical energy estimate on $\pi(\rho)$ and $\rho|u|^2$
- The BD entropy estimate on $\rho|u + \nabla\log\rho|^2$.

In the Navier-type boundary case, a classical Gronwall lemma helps to get the appropriate control on $\rho|u + \nabla\log\rho|^2$.

3.4. *Strong solutions*

There are several results proving the local existence of a strong solution for shallow-water equations written as follows

$$\begin{cases} \partial_t h + \operatorname{div}(hv) = 0, \\ \partial_t (hv) + \operatorname{div}(hv \otimes v) - \nu \operatorname{div}(h\nabla v) + h\nabla h = 0. \end{cases} \tag{26}$$

Following the energy method of A. MATSUMURA and T. NISHIDA, see [130], it is natural to show the global (in time) existence of classical solutions to viscous shallow-water equations. The external force field and the initial data are assumed to be small in a suitable space. Such a result has been proved in [108], [59]. In [151] and [152], a global existence and uniqueness theorem of strong solutions for the initial-value problem for the viscous shallow-water equations is established for small initial data with no forcing. Polynomial L^2 and decay rates are established and the solution is shown to be classical for $t > 0$.

More recently, in [160], the Cauchy problem for viscous shallow-water equations is studied. The authors' work in the Sobolev spaces of index $s > 2$ to obtain local solutions for any initial data, and global solutions for small initial data. To this purpose, they made use of the Littlewood–Paley decomposition. The result reads

THEOREM 3.4. *Let $s > 0$, u_0, $h_0 - \overline{h_0} \in H^{s+2}(R^2)$, $\|h_0 - \overline{h_0}\|_{H^{s+2}} \ll \overline{h_0}$ where (h_0, u_0) are the initial data and $\overline{h_0}$ the equilibrium height. Then there exists a positive time T, a unique solution (u, h) of the Cauchy problem (26), such that*

$$u, h - \overline{h_0} \in L^\infty(0, T; H^{2+s}(\mathbf{R}^2)), \qquad \nabla u \in L^2(0, T; H^{2+s}(\mathbf{R}^2)).$$

Furthermore, there exists a constant c such that if $\|h_0 - \overline{h_0}\|_{H^{s+2}(R^2)} + \|u_0\|_{H^{s+2}(R^2)} \le c$ then we can choose $T = +\infty$.

Nothing has been done, yet, using the new mathematical entropy to obtain better results such as the local existence of a strong solution with initial data including a vacuum. Note that such a solution is important from a physical point of view: a burst dam situation for instance. And also note that recently, B. HASPOT in [102], has proved a result in critical spaces which improves the previous theorem.

3.5. *Other viscous terms in the literature*

In [115], different diffusive terms are proposed, namely $-2\operatorname{div}(hD(v))/\mathrm{Re}$, or $-h\Delta v/\mathrm{Re}$ or else $-\Delta(hv)/\mathrm{Re}$. The reader is referred to [20] for the study of such diffusive terms in the low Reynolds approximation. We now give some comments on the last two propositions.

Diffusive term equals to $-h\Delta v$. It is shown by P GENT in [93], that this form, which is frequently used for the viscous adiabatic shallow-water equations, is energetically inconsistent compared to the primitive equations. An energetically consistent form of the shallow-water equations is then given and justified in terms of isopycnal coordinates.

This energetical form is exactly the form considered previously namely $-2\mathrm{div}(hD(v))$. Examples are given of the energetically inconsistent shallow-water equations used in low-order dynamical systems and simplified coupled models of tropical air sea interaction and the El Ninó-southern oscillation phenomena.

From a mathematical point of view, it is easy to see this energetical inconsistency with the diffusion term is equal to $-h\Delta v$. Consider the following system

$$\begin{cases} \partial_t h + \mathrm{div}(hv) = 0, \\ \partial_t(hv) + \mathrm{div}(hv \otimes v) = -h\dfrac{\nabla h}{\mathrm{Fr}^2} + \dfrac{1}{\mathrm{Re}} h\Delta v + hf. \end{cases}$$

Multiplying the momentum equation by v/h, we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega |v|^2 + \frac{1}{\mathrm{Re}}\int_\Omega |\nabla v|^2 + \frac{1}{2}\int_\Omega v\cdot\nabla|v|^2 + \frac{1}{\mathrm{Fr}^2}\int_\Omega v\cdot\nabla h = \int_\Omega f\cdot v.$$

The last term on the left-hand side may be written

$$\int_\Omega v\cdot\nabla h = -\int_\Omega \log h\,\mathrm{div}(hv) = \int_\Omega \log h\,\partial_t h = \frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega (h\log h - h).$$

The third term may be estimated as follows

$$\left|\int_\Omega |v|^2 \mathrm{div}\, v\right| \le \|v\|^2_{L^4(\Omega)}\|v\|_{H^1(\Omega)} \le C\|v\|^2_{H^1(\Omega)}\|v\|_{L^2(\Omega)}.$$

Thus, if we want to get dissipation, we have to look at solutions such that

$$\|v\|_{L^2(\Omega)} < 1/(C\,\mathrm{Re}).$$

That means sufficiently small solutions. Such an analysis has been performed in [142] by looking at solutions such that v is bounded in $L^2(0,T;H^1(\Omega))$, $h \in L^\infty(0,T;L^1(\Omega))$, $h\log h \in L^\infty(0,T;L^1(\Omega))$. We also refer to [65] for some smoothness and uniqueness results. Let us assume the following hypothesis

$$\begin{cases} 2/\mathrm{Re} > \lambda > 0, \qquad B = 2/\mathrm{Re} - \lambda, \quad 0 < \theta < 1, \\ h_0 \ge 0, \qquad h_0 \ln h_0 \in L^1(\Omega), \qquad f \in L^1(0,T;H^{-1}(\Omega)), \\ \|v_0\|_{L^2(\Omega)} + \dfrac{2}{\mathrm{Fr}^2}\|h_0 \ln h_0\|_{L^1(\Omega)} \\ \quad + \dfrac{1}{\lambda}\|f\|_{L^2(0,T;H^{-1}(\Omega))} \\ \quad + \dfrac{2}{e\,\mathrm{Fr}^2}\mathrm{meas}(\Omega) < \theta^2\dfrac{B^2}{C^2}, \end{cases} \tag{27}$$

where e is defined using the convexity of $s \mapsto s\log s$, for $s \ge 0$, namely using that

$h \log h > -1/e$. Thus, under Hypothesis (27), P.ORENGA in [142], gets the following result

THEOREM 3.5. *Let* $v_0 \in H_0^1$, $h^0 \in L^1(\Omega)$, f, θ, λ *satisfying Hypothesis* (27). *Then there exists a global weak solution such that*

$$(v, h) \in \Big(L^2(0, T; H^1(\Omega)) \cap L^\infty(0, T; L^2(\Omega))\Big) \times L^\infty(0, T; L^1(\Omega))$$

and

$$h \log h \in L^\infty(0, T; L^1(\Omega)).$$

Main difficulty. The mathematical problem, in the existence process, is to prove strong convergence of sequences $\{h_n v_n\}_{n\in N}$ in the stability process. For this, the author emphasizes the fact that h_n and $h_n \log h_n$ are uniformly bounded in $L^\infty(0, T; L^1(\Omega))$. Using the Dunford-Pettis theorem and the Trudinger-Moser inequality, he concludes with a compactness argument. Namely from Dunford-Pettis, h_n belongs to a L^1 weak compact. We then use the fact that

$$\begin{aligned} Y = \{hv : v \in L^2(0, T; H_0^1(\Omega)), \qquad & h \geq 0, \\ h \in L^\infty(0, T; L^1(\Omega)), \qquad & h \log h \in L^\infty(0, T; L^1(\Omega))\} \end{aligned} \tag{28}$$

is weakly compact in L^1.

Note that, using this diffusive term, several papers have been devoted to the low Reynolds approximation namely the following system

$$\partial_t h + \operatorname{div}(hv) = 0, \quad \partial_t v - \Delta v + \nabla h^\alpha = f$$

with $\alpha \geq 1$. The most recent, [112], deals with blow up phenomena if the initial density contains a vacuum, using uniform bounds with respect to time of the L^∞ norm on the height. It could be interesting to understand what happens without simplification by h in the momentum equation allowing the height to vanish. Note that such a simplification has been also done in [125] to study the high rotation and low Froude number limit of inviscid shallow-water equations.

Diffusive term equal to $-\nu\Delta(hv)$. Use of this diffusion term gives also energetical inconsistency. Only results concerning the existence of global weak solutions for small data have been obtained.

3.6. *Low Froude number limits*

3.6.1. The quasi-geostrophic model The well-known quasi-geostrophic system for zero Rossby and Froude number flows has been used extensively in oceanography and meteorology for modeling and forecasting mid-latitude oceanic and atmospheric circulation. The quasi-geostrophic equation expresses conservation of the zero-order potential vorticity of the flow. In two dimensions that means neglecting the stratification, a model of this type may be obtained from the shallow-water equations. It gives the following bi-dimensional velocity form system

$$\begin{cases} \partial_t v + \operatorname{div}(v \otimes v) = -\mathcal{D}_{\text{lim}} - \nabla p + \dfrac{1}{\text{Re}} \Delta v - \partial_t \Delta^{-1} v, \\ \operatorname{div} v = 0. \end{cases} \tag{29}$$

with some drag terms $\mathcal{D}_{\text{lim}}$ which could be $\mathcal{D}_{\text{lim}} = r_0 v + r_1 |v| v$. We also note the presence of a new term $\partial_t \Delta^{-1} v$ coming from the free surface, which one cannot derive from the standard rotating Navier–Stokes equations in a fixed domain. To the best knowledge of the authors, there exists one mathematical paper on the derivation of such a model from the degenerate viscous shallow-water equations, namely with vacuum state. It concerns global weak solutions of (17) with well-prepared data in a two-dimensional periodic box Ω with the drag term of the form $\mathcal{D} = r_0 v + r_1 h |v| v$, see [37]. The main theorem reads

THEOREM 3.6. *Assume* $\text{Fr} = \text{Ro} = \varepsilon$ *(*Re *and* We *being fixed) and initial conditions in the form* $(v_0^\varepsilon, h_0^\varepsilon) = \sum_{k \geq 0} \varepsilon^k (v_0^k, h_0^k)$ *with* $h_0 \equiv h_0^0 = 1$, $v_0 \equiv v_0^0 = \nabla^\perp h_0^1 \equiv \nabla^\perp \Psi_0$ *and bounds on initial data for shallow-water satisfied uniformly (energy and* BD *entropy). Assume* $v_0 \in (H^2(\Omega))^2$ *and*

$$(v_0^\varepsilon, h_0^\varepsilon) \to (v_0, 1) \quad \text{in } (L^2(\Omega))^3, \qquad (h_0^\varepsilon - 1)/\varepsilon \to \Psi_0 \quad \text{in } L^2(\Omega),$$
$$\sqrt{\text{We}} \nabla h_0^\varepsilon \to 0 \quad \text{in } L^2(\Omega).$$

Then let $(v^\varepsilon, h^\varepsilon)$ *be global weak solutions of* (17), *where*

$$\begin{aligned} v^\varepsilon &\to v \quad \text{in } L^\infty(0, T; (L^2(\Omega))^2) \\ (h^\varepsilon - 1)/\varepsilon &\to \Psi \quad \text{in } L^\infty(0, T; L^2(\Omega)), \\ \nabla h^\varepsilon &\to 0 \quad \text{in } L^2(0, T; (L^2(\Omega))^2), \\ \sqrt{\text{We}} \nabla h^\varepsilon &\to 0 \quad \text{in } L^\infty(0, T; (L^2(\Omega))^2) \end{aligned}$$

when $\varepsilon \to 0$, *and* $v = \nabla^\perp \Psi$ *is the global strong solution of the quasi-geostrophic equation* (29) *with initial data* v_0.

The reader is also referred to [92] for a recent derivation of equatorial wave propagation with a latitude dependent Coriolis term from a non-degenerate viscous shallow-water model. See also papers [81] and [82] for inviscid models.

REMARK. Mathematical derivation of the quasi-geostrophic equations from the free surface Navier–Stokes or primitive equations is an interesting open problem. Many results exist with a rigid-lid hypothesis, see for instance [66] and references cited therein.

More realistic domains. We now present three simple studies that have been realized in order to better understand what could be new features in more realistic domains. These works concern the Sverdrup relation which is the keystone of the theory of wind-driven oceanic circulation, see [145].

(1) *Islands.* In some papers, it has been shown that addition of an island within an oceanic basin introduces new elements to the problem of the general circulation of the fluid

contained within the basin. In particular, and with the problem of the circulation of the abyssal ocean with mid-ocean ridges, the presence of a long thin meridionally oriented island barrier may induce a recirculation in the sub-basin to the east of the ridge. In [48], the authors have studied, in the stationary case, the effect of an island on the Sverdrup relation mathematical derivation. Assuming Ω to be a bounded domain in two-dimensional space of class $\mathcal{C}^2$ with an island Ω_2 that means $\Omega = \Omega_1 \setminus \overline{\Omega_2}$ with Ω_i simply connected to $\Omega_2 \subset\subset \Omega_1$ and $\Gamma_i = \partial\Omega_i$, they have proved that a corrector has to be added to the Sverdrup solution. In addition, in the island case, the stationary quasi-geostrophic system reads

$$\begin{cases} E\Delta^2\Psi - \mu\Delta\Psi + \varepsilon\nabla^\perp\Psi\cdot\nabla\Delta\Psi + a\cdot\nabla\Psi = \nabla^\perp\cdot f \\ \Psi|_{\Gamma_1} = 0, \qquad \Psi|_{\Gamma_2} = c_{E,\mu,\varepsilon} & \text{on } \partial\Omega, \\ \nabla\Psi\cdot n = 0 & \text{on } \partial\Omega, \end{cases}$$

with the following compatibility condition

$$\int_{\Gamma_2} (E\nabla\Delta\Psi + f^\perp)\cdot n = 0$$

where E, μ, ε are some small positive constants, $\nabla^\perp = (-\partial_y, \partial_x)$, $a = (-1, 0)$ and $f^\perp = (-f_2, f_1)$. The reader interested by compatibility conditions in fluid mechanics problems is referred to [98] and the references cited therein.

Note that in [48], the domain Ω_1 is smooth with horizontal north and south boundaries. The sub-domains Ω^I (respectively Ω^{II}) are defined such that horizontal lines coming from the right external part of the boundary (respectively the left boundary of the island) cross the left external part of the boundary and the right boundary of the island (respectively part of the left external boundary). The following result then holds

THEOREM 3.7. *Let $f \in (H^1(\Omega))^2$ such that $\nabla^\perp\cdot f \in H^2(\Omega)$. Let Ψ be a solution in $H^4(\Omega)$ of (30). Assume that $\varepsilon/\mu^{5/4}E^{7/4} \to 0$. Then, for all neighborhood $V = V_- \cup V_{I,II}$ of $\Gamma^- \cup \Gamma_{I,II}$ where $\Gamma^- = \{x \in \partial\Omega : n_x \leq 0\}$ and $\Gamma_{I,II} = \partial\Omega^I \cap \partial\Omega^{II}$, we have*

$$\Psi \to \overline{\Psi} + c1_{\Omega^{II}} \quad \text{weakly in } L^2(\Omega) \text{ and strongly in } L^2(\Omega\setminus V),$$
$$\partial_x\Psi \rightharpoonup \partial_x\overline{\Psi} \quad \text{weakly in } L^2(\Omega\setminus V_-),$$

where $\overline{\Psi}$ is the solution in $L^2(\Omega)$ of the Sverdrup relation

$$-\partial_x\overline{\Psi} = \nabla^\perp\cdot f \quad \text{in } \Omega,$$
$$\overline{\Psi} = 0 \quad \text{on } \Gamma^+, \tag{30}$$

and c is calculated using the equality

$$c = \frac{-\int_{\Gamma_2^-}\overline{\Psi} + \int_{\Gamma_2} f^\perp\cdot n}{\int_{\Gamma_2^+} n_x}$$

with $\Gamma_2^+ = \{x \in \Gamma_2 : n_x > 0\}$, $\Gamma_2^- = \{x \in \Gamma_2 : n_x < 0\}$ *and* $1_{\Omega^{II}}$ *is the characteristic function corresponding to the subdomain* Ω^{II}.

REMARK. If there is no tangential force applied to the island's boundary that means $f^\perp \cdot n = 0$, then the constant c is the meridional mean value of the trace of $\overline{\Psi}$ on this coast. More precisely

$$c = \frac{1}{(y_n - y_s)} \int_{y_s}^{y_n} \Psi(x_+(y), y)\, \mathrm{d}y$$

where x_+ denotes the graph of the eastern part of the island, y_s and y_n are, respectively, the minimal meridional coordinate (South) and the maximum meridional coordinate (North).

REMARK. Note that it could be really interesting to work on mathematical extensions to the ridges and islands effects on boundary layer stability, stratification for instance.

(2) *Variable depth vanishing on the shore.* Let us point out that several papers have been written by the author concerning variable depth vanishing on the shore. For instance, in [55] the use of the explicit formulation between the velocity field v and the stream function Ψ provides a Stommel equation with variable coefficients. More precisely, as mentioned in [145], a vertical-geostrophic system may be used to describe a small depth velocity close to the surface. Let

$$\omega = \{(x, z) : x \in \Omega, -h(x) < z < 0\},$$

where $\Omega \subset \mathbf{R}^2$ is a simply connected domain and h is the vertical depth. This system, in ω (see [50]), reads

$$\begin{cases} -\partial_z^2 v + K v^\perp + \nabla_x P = 0, \qquad \partial_z P = 0,\\ \operatorname{div}_x v + \partial_z w = 0,\\ (v, w)|_{z=-h} = 0, \quad \partial_z v|_{z=0} = f, \quad w|_{z=0} = 0, \quad \displaystyle\int_{z=-h}^0 v \cdot n_{\partial\Omega} = 0, \end{cases}$$

where K is a coefficient corresponding to the Coriolis force. Such a system has been obtained from the three-dimensional Navier–Stokes equation by assuming the aspect ratio small enough. Assuming the β plan approximation, see [145], $K = e(1 + \beta y)$ with two positive constants e and β linked to the mean latitude coordinate. This system is integrable with respect to z and we can prove that the stream function Ψ associated with the vertical mean value $\overline{v}$ of v, where $h(x, y)$ is the corresponding depth, satisfies the following equation

$$\begin{cases} \operatorname{div}(a_Q \nabla\Psi + b_Q \nabla^\perp\Psi + c_Q f + d_Q f^\perp) = 0,\\ \Psi|_{\partial\Omega} = 0, \end{cases}$$

where $\nabla^{\perp} = (-\partial_y, \partial_x)$, $f^{\perp} = (-f_2, f_1)$ and

$$
\begin{aligned}
&a_Q = \frac{a_P}{a_P^2 + b_P^2}, \qquad b_Q = \frac{-b_P}{a_P^2 + b_P^2},\\
&c_Q = \frac{-a_P d_P + b_P c_P}{a_P^2 + b_P^2}, \qquad d_Q = \frac{a_p c_P + b_P d_P}{a_P^2 + b_P^2},\\
&a_P = \frac{SC - sc}{4\kappa^3 M}, \qquad b_P = -\frac{H}{2\kappa^2} + \frac{SC + sc}{4\kappa^3 M},\\
&c_P = -\frac{Ss}{2\kappa^2 M}, \qquad d_P = \frac{1}{2\kappa^2}\left(1 - \frac{Cc}{M}\right),\\
&S = \sinh(\kappa h), \qquad C = \cosh(\kappa h), \qquad s = \sin(\kappa h), \qquad c = \cos(\kappa h),\\
&M = S^2 + c^2, \qquad \kappa = \sqrt{e(1 + \beta y)/2}.
\end{aligned}
$$

These equations may be written in the following form:

$$
\begin{cases}
a_Q \Delta\Psi + (\nabla a_Q + \nabla^{\perp} b_Q) \cdot \nabla\Psi = -\mathrm{div}(c_Q f + d_Q f^{\perp}),\\
\Psi|_{\partial\Omega} = 0,
\end{cases} \tag{31}
$$

where

$$
a_Q = \frac{4\kappa^3 M(SC - sc)}{(SC - sc)^2 + (SC + sc - 2\kappa h M)^2}, \tag{32}
$$

$$
b_Q = \frac{4\kappa^3 M(2\kappa h M - (SC + sc))}{(SC - sc)^2 + (SC + sc - 2\kappa h M)^2}, \tag{33}
$$

$$
c_Q = \frac{2\kappa Ss(2\kappa h M - (SC + sc)) - (SC - sc)(2\kappa M - 2\kappa Cc)}{(SC - sc)^2 + (SC + sc - 2\kappa h M)^2}, \tag{34}
$$

$$
d_Q = \frac{2\kappa Ss(sc - SC) + (2\kappa M - 2\kappa Cc)(SC + sc - 2\kappa h M)}{(SC - sc)^2 + (SC + sc - 2\kappa h M)^2}. \tag{35}
$$

If we look at the orders of the various coefficients assuming $h = 1$, an external force $f(x, y) = (f_1(y), 0)$ and a domain $\Omega = (0, 1)^2$. This proves that the coefficients do not depend on x and the equation may be simplified as follows

$$
\begin{cases}
-a_Q \Delta\Psi - \partial_y a_Q \partial_y \Psi - \partial_y b_Q \partial_x \Psi = -d_Q \partial_y f_1 + \partial_y d_Q f_1,\\
\Psi|_{\partial\Omega} = 0.
\end{cases} \tag{36}
$$

Since $e >> 1$ and $\beta e >> 1$, then we have

$$
\begin{aligned}
&a_Q \approx \sqrt{e}, \qquad \partial_y a_Q \approx \sqrt{e}\beta,\\
&\partial_y b_Q \approx e^{3/2}\beta, \qquad d_Q \approx 1, \qquad \partial_y d_Q \approx \beta.
\end{aligned}
$$

Thus the equation may be seen as the simple form

$$\begin{cases} -\theta\Delta\Psi + a\cdot\nabla\Psi = g & \text{in } \Omega, \\ \Psi|_{\partial\Omega} = 0 \end{cases} \tag{37}$$

with $a = (-\partial_y a_Q/a_Q, -\partial_y b_Q/a_Q)$ and $\theta = 1/e\beta$. Asymptotically $a \approx \overline{a} = (0, \overline{a}_2)$, where $\overline{a}_2$ does not depend on y and is negative. The characteristic lines of $\overline{a}$ are horizontal lines which all cross $\partial\Omega$. We show that as $\theta \to 0$, Ψ converges to the solution of

$$\begin{cases} \overline{a}\cdot\nabla\overline{\Psi} = g & \text{in } \Omega, \\ \overline{\Psi}|_{\partial\Omega_-} = 0. \end{cases} \tag{38}$$

Assume that Ω is a two-dimensional bounded domain of class $\mathcal{C}^1$ and $\overline{a} \in \mathcal{C}^1(\overline{\Omega})$ with $\operatorname{div}\overline{a} = 0$ and all its characteristic lines cross $\partial\Omega$. Assume $g \in H^1(\Omega)$. Then we get the following theorem

THEOREM 3.8. *Let $\theta = 1/e\beta$. Assume that there exists $s > 2$ such that*

$$\|a - \overline{a}\|_{(L^s(\Omega))^2} \le c(\theta)\theta^{\frac{1}{2}+\frac{1}{s}},$$

where $c(\theta) \to 0$ when $\theta \to 0$. Assume, for θ small enough, that there exists a solution Ψ of system (37). Then Ψ is unique and, when $\theta \to 0$, for all neighborhood V of $\partial\Omega_+ = \{(x,y) \in \partial\Omega : \overline{a}\cdot n > 0\}$ with n the outward unit normal, we have

$$\Psi \to \overline{\Psi} \quad \text{in } L^2(\Omega \setminus V),$$

where $\overline{\Psi}$ is a solution of (38).

The proof relies on integration along the characteristics of $\overline{a}$ and a suitable choice of test functions. This simple application shows the possibility of playing with the measure of the part where the gradient of the topography varies considerably in order to get asymptotic mathematical justification. For instance, it could be interesting to look at the asymptotic quasi-geostrophic limit from Navier–Stokes equations with a rigid lid assumption and the presence of ridges.

REMARK. The vertical-geostrophic equations have been mathematically justified from the Navier–Stokes equations with a vanishing depth close to the shore in [50]. Assuming the depth satisfies some constraints linked to the distance function to the shore, a density lemma equivalent to the usual $\overline{\mathcal{V}}^{\|\cdot\|_{H^1}} = V$ relation for Navier–Stokes in a Lipschitz domain has been established. Recall that, in the Navier–Stokes framework (in space dimension d)

$$\mathcal{V} = \{u \in (\mathcal{D}(\Omega))^d : \operatorname{div} u = 0\}, \qquad V = \{u \in (H_0^1(\Omega))^d : \operatorname{div} u = 0\}.$$

REMARK. These simple calculations show that it should be possible to consider vanishing depth in various geophysical models to deal with the mathematical justification of western intensification of currents. All depends on the ratio between the bathymetry profile and Coriolis force.

(3) *Various boundary conditions.* A simplified stationary model describing homogeneous wind-driven oceanic circulation may be given in a simply connected bounded domain $\Omega \subset \mathbf{R}^2$ by

$$E\Delta^2\Psi - \mu\Delta\Psi - \varepsilon J(\Psi, \Delta\Psi) - \partial_x\Psi = f, \tag{39}$$

where J denotes the Jacobian defined by

$$J(\Psi, \Delta\Psi) = \nabla^\perp\Psi \cdot \nabla\Delta\Psi,$$

and, see [145] page 271,

$$\varepsilon = \left(\frac{\delta_I}{L}\right)^2, \qquad \mu = \frac{\delta_S}{L}, \qquad E = \left(\frac{\delta_M}{L}\right)^3$$

with L the scale of the motion, δ_I, δ_M and δ_S either inertia, horizontal friction, or bottom friction lengths. The function Ψ denotes the stream function such that the vertical mean value of the horizontal component of the velocity is given by $v = (-\partial_y\Psi, \partial_x\Psi)$. This equation is called the quasi-geostrophic equation written for the stream function. The dimensionless constants ε, μ, E are linked to the ratio between nonlinearity, friction on the bottom and lateral diffusion. The external force f is given by $f = -k \cdot \mathrm{curl}\tau$, where $k = (0, 0, 1)$ and $\tau = (\tau_x, \tau_y, \tau_z)$ is the Cauchy tensor due to the wind.

We have to choose boundary conditions for such simple models. On the boundary of the basin, the normal velocity has to be prescribed. If the basin is closed and if we do not consider inflow and outflow, the flux corresponding to the horizontal velocity v has to be null on the boundary. This means

$$v \cdot n = \nabla^\perp\Psi \cdot n = 0,$$

where n denotes the outward external normal. This means that Ψ is constant on each connected boundary. If we assume no island, we can choose

$$\Psi = 0 \quad \text{on } \partial\Omega.$$

The presence of the diffusion term $E\Delta^2\Psi$ implies that we have to choose an other boundary conditions. J. PEDLOSKY in [145] proposes considering one of the following boundary conditions on $\partial\Omega$

$$\nabla\Psi \cdot n = 0, \qquad \Delta\Psi = 0, \qquad \nabla\Delta\Psi \cdot n + \varepsilon^{-1} j \cdot n = 0,$$

where j denotes the unit vector oriented along the y axis. What kind of boundary conditions do we have to choose? If the dissipation is small, i.e. $E << 1$, does the choice have some influence on the global circulation? This choice is not so clear, as indicated by J. PEDLOSKY in [146]. For a first step, we have to study the influence of the boundary conditions on convergence when some parameters tend to zero. In [146], it is noted that if E, ε, μ are small coefficients, the equilibrium equation is reduced to the Sverdrup relation

$$-\partial_x \Psi = f$$

which corresponds to the relation proposed in oceanography in 1947 by Sverdrup. So for the first step, it is possible to consider a classical method exposed, for instance, in J.-L. LIONS [114] and used by T. COLIN in [69], but with mixed boundary-conditions of the type

$$\Delta\Psi = 0 \quad \text{on } \Gamma_1, \qquad \nabla\Psi \cdot n = 0 \quad \text{on } \Gamma_2$$

with $\overline{\Gamma_1} \cup \overline{\Gamma_2} = \partial\Omega$, in a smooth domain with horizontal north and south parts and with an East and West coasts defined by two graphs g_{East} and g_{West} of class $\mathcal{C}^2$ on $(0, 1)$. These boundary conditions are usually used, see for instance J. DENG [75] pages 2183–2184. In [32] the following extra boundary condition

$$\nabla\Delta\Psi = 0 \quad \text{on } \partial\Omega$$

has been studied in a domain of the form $\Omega = (0, 1) \times S^1$, where S^1 denotes a circle in one dimensional space (i.e. a periodic condition with respect to y). The study of small parameters in the stationary quasi-geostrophic equation has been realized, for instance, by V. BARCILON, P. CONSTANTIN, E. TITI [15] with ε and μ fixed and E tending to zero, with the extra boundary condition $\Delta\Psi = 0$ on $\partial\Omega$ and $E \le \mu^3/8$. The Sverdrup relation has been justified in T. COLIN [69] with the extra boundary condition $\nabla\Psi = 0$ on $\partial\Omega$ and assuming μ of same order as ε and assuming some relation between E and ε. The study in [69] emphasizes that $\nabla\Psi \cdot n = 0$ on $\partial\Omega$ and the fact that μ is of same order as ε. We prove in [32] that μ may tend to zero without there being a relationship between E and ε, and eventually $\mu = 0$. By this means, we give an answer to a question posed by J.-L. LIONS on page 403 of [114], which is the study of (39) with $\mu = 0$ for ε, $E \to 0$ with $\Delta\Psi = 0$ on Γ_{East}, Γ_{North}, Γ_{South} and $\nabla\Psi \cdot n = 0$ on Γ_{West}. We find this kind of equation and boundary conditions in G. F. CARRRIER [63], P. CESSI [68], where physical analysis is performed, and in E. BLAYO, J. VERRON [22], K. BRYAN [29], J. PEDLOSKY [145] pages 310–311 where numerical simulations are given in a square domain.

The method used in [114] for the equation $\varepsilon\Delta^2\Psi - \partial_x\Psi = f$ has been used by [69] for the equation (39) with the boundary condition $\nabla\Psi \cdot n = 0$. The method, in order to prove convergence up to the boundary such that $n_x > 0$ (n_x being the horizontal component of the outward unit normal), seems inadequate if we consider the boundary condition $\Delta\Psi$ on this boundary. We have changed the proof of [69] in order to consider different boundary conditions in a domain representing the North Atlantic. We assume that the external force f is more regular than $f \in L^2(\Omega)$. This is the case in all the examples considered in oceanography. We refer the reader to [145] pages 31 and 65 where $\Omega =]0, L_1[^2$ with $f = -W_0 \sin(\pi y/L_1)$ and $f = -W_0 \sin(\pi x/L_1)\sin(\pi y/L_1)$ is considered. In this way, we get a larger range of coefficients than in [69]. We prove, by the theorem in part iii, the convergence for $\Delta\Psi = 0$ on $\partial\Omega$ without the hypothesis $E \leq \mu^3/8$ as done in [15]. The nonlinearity is dominated by the friction coefficient and the lateral diffusion term is independent of μ and ε.

The first result, proved in [32], is the following.

THEOREM 3.9. *Let Ω be defined with g_{East} and g_{West} of class $\mathcal{C}^2$.*

(i) *Let μ, ε, E tend to 0 with $\varepsilon \ll E$ and $f \in L^2(\Omega)$. Let us consider $(\Psi)_{\varepsilon,\mu,E}$ to be solutions in $H^3(\Omega)$ of (39). Then there exists a subsequence of $(\Psi)_{\varepsilon,\mu,E}$ weakly converging to Ψ_0 in $L^2(\Omega)$ with Ψ_0 satisfying, in the distribution sense, the equation*

$$-\partial_x \Psi_0 = f.$$

Moreover, we have

$$\int_\Omega |\Psi|^2 + E\int_\Omega |\Delta\Psi|^2 + \mu\int_\Omega |\nabla\Psi|^2 \leq C,$$

where C is a constant which does not depend on μ, E and ε.

(ii) *Let μ, ε, E tend to 0 with $\varepsilon \ll E$, $\Delta\Psi = 0$ on Γ_{East} and $f \in H^2(\Omega)$. Then $\eta\Psi$ converges to $\eta\Psi_I$ in $L^2(\Omega)$ with Ψ_I the solution of*

$$-\partial_x \Psi_I = f, \qquad \Psi_I = 0 \quad \text{on } \{x \in \partial\Omega n_x > 0\}$$

for all $\eta \in \mathcal{C}^2(\overline{\Omega})$ with $\eta \equiv 0$ in a neighborhood of $\{n_x \leq 0\}$, where n_x is the horizontal component of the outward unit normal of $\partial\Omega$.

(iii) *Let μ, ε, E converge to 0 and $f \in L^2(\Omega)$. Let $(\Psi)_{\varepsilon,\mu,E}$ be solutions in $H^3(\Omega)$ of (39) with $\Gamma_2 = \emptyset$. Then there exists c_Ω such that if $\varepsilon \ll c_\Omega \mu^3$, there exists a subsequence $(\Psi)_{\varepsilon,\mu,E}$ converging to Ψ_0 weakly in $L^2(\Omega)$ with μ_0 solution of the Servdrup relation.*

Finally, considering the problem with the extra condition $\nabla\Delta\Psi \cdot n = 0$ on the boundary, in the linear case that means $\varepsilon = 0$, a more precise estimate than before has been given in [32]. They have proved the following result which gives some uniform estimates on the gradient of Ψ in weighted spaces.

THEOREM 3.10. *Let $\Omega = (0,1) \times S^1$ and $\varepsilon = 0$. Let μ, E tend to 0 and $f \in H^1(\Omega)$. Assume that $(\Psi)_{\mu,E}$ is a solution in $H^4(\Omega)$.*

(a) *If $\int_0^1 f(\xi, y)\,\mathrm{d}\xi = 0$, there exists c_Ω such that $E \ll c_\Omega \mu^3$ then*

$$\Psi_{\mu,E} \rightharpoonup \Psi_I \quad \text{in } H^1(\Omega) \text{ weak},$$

where Ψ_I is given by $\Psi_I = \int_0^x f(\xi, y)\,\mathrm{d}y$.

(b) *If $\int_0^1 f(\xi, y)\,\mathrm{d}\xi \neq 0$, there exists c_ω such that if $E \ll c_\Omega \mu^{7/2}$ then*

$$\Psi_{\mu,E} \rightharpoonup \Psi_2 \quad \text{in } L^2(\Omega) \text{ weak}$$
$$x^{1/2}\partial_x(\Psi_{\mu,E}) \rightharpoonup x^{1/2}\partial_x \Psi_2 \quad \text{in } L^2(\Omega) \text{ weak},$$

where Ψ_2 *is the solution of*

$$\partial_x \Psi_2 = f \quad in\ \Omega, \qquad \Psi_2 = 0 \quad on\ \{1\} \times S^1.$$

The proof is based on a suitable choice of test functions: Ψ, $\Psi \exp x$, $(\Psi - \Psi_I)\Phi$ where $\Phi = \int_{g_{\text{west}}(y)}^{x} \eta(x_\star, y) \mathrm{d}x_\star$ with η a positive regular function which vanishes in the neighborhood of Γ_-. The proof of the first theorem is based on test functions of the form $\int_0^x \Delta\Psi(\xi, y)\,\mathrm{d}\xi$, $\Delta\Psi$, $\Psi \exp x$ and $x\partial_x \Psi$.

REMARK. This part shows the influence of boundary conditions on singular perturbation problems. It could be interesting to describe the effects on more complicated systems such as those involving stratification.

3.6.2. The lake equations The so-called lake equations arise as the shallow-water limit of the rigid lid equations – three dimensional Euler or Navier–Stokes equations with a rigid lid upper boundary condition – in a horizontal basin with bottom topography. They could also be seen as a low Froude approximation of the shallow-water equations. More precisely, let us consider the shallow-water system

$$\begin{cases} \partial_t(hv) + \operatorname{div}(hv \otimes v) \\ \quad = -h\dfrac{\nabla(h-b)}{\mathrm{Fr}^2} + \dfrac{2}{\mathrm{Re}}\operatorname{div}(hD(v)) + \mathcal{D}, \\ \partial_t h + \operatorname{div}(hv) = 0 \end{cases} \tag{40}$$

with a drag term $\mathcal{D}$ that will be given later on, initial conditions (2) and boundary conditions which depend on the domain Ω being considered. Let the Reynolds number be fixed, the Froude number go to zero and let us denote $\mathrm{Fr} = \varepsilon$ for simplicity. We use the following Taylor expansion with respect to ε

$$h = h^0 + \varepsilon^2 h^1 + \cdots$$
$$v = v^0 + \varepsilon^2 v^1 + \varepsilon^2 v^2 + \cdots$$

We plug these expansions in the equation and identify with respect to powers of ε. The momentum equation $(40)_1$, divided by h, gives at order $1/\varepsilon^2$

$$\nabla(h^0 - b) = 0$$

which implies that

$$h^0(t, x) = b(x) + c(t).$$

Using the mass equation $(40)_2$ (assuming periodic boundary conditions for simplicity) and integrating it with respect to the space variables, it gives that $c(t) = 0$ if initially $h|_{t=0} = b$. Now by using the mass equation $(40)_2$, we find the constraint $\operatorname{div}(bv^0) = 0$ at order ε^0.

It remains to look at the next order in the momentum equation (order ε^0), where we get

$$\partial_t(bv^0) + \operatorname{div}(bv^0 \otimes v^0) = -b\nabla h^1 + \frac{2}{\mathrm{Re}}\operatorname{div}(bD(v^0)) + \frac{2}{\mathrm{Re}}\nabla(b\operatorname{div} v^0) + \mathcal{D}_{\mathrm{lim}}.$$

In summary, we get:

$$\begin{cases} \partial_t(bv^0) + \operatorname{div}(bv^0 \otimes v^0) = -b\nabla h^1 + \dfrac{2}{\mathrm{Re}}\operatorname{div}(bD(v^0)) + \mathcal{D}_{\mathrm{lim}}, \\ \operatorname{div}(bv^0) = 0, \end{cases} \tag{41}$$

where b is the basin depth and $\mathcal{D}_{\mathrm{lim}}$ depends on $\mathcal{D}$ and will be defined later on. Remark that system (41) is similar to the model obtained in [128].

The viscous case. The viscous lake model has been formally derived and studied by D. LEVERMORE and B. SAMARTINO, see [111]. In this paper, assuming that the depth b is positive and smooth up to the boundary of Ω, they prove that the system is globally well posed. Note that this model has been used to simulate the currents in Lake Erie, see [148]. Concerning the boundary conditions, they consider a bounded domain with non-slip boundary conditions

$$v^0 \cdot n = 0, \qquad \tau \cdot ((\nabla v^0 +^t \nabla v^0)n) = -\alpha\, v^0 \cdot \tau.$$

where n and τ are the outward unit normal and unit tangent to the boundary and α is a positive constant coefficient. They assume $\alpha \geq \kappa$, where κ is the curvature of the boundary. Note that the asymptotic process from the viscous shallow-water system to the viscous lake equations has been mathematically justified in [41] in a well-prepared case, which assumes a sufficiently smooth basin depth such that $b(x) \geq c > 0$ for all $x \in \Omega$ and considers weak solutions for the shallow-water model (40) with drag terms $\mathcal{D} = r_0 u + r_1 h|u|u$. The result is the following

THEOREM 3.11. *Let Ω be a periodic box in two-dimensional space. Let $v_0^0 \in (H^3(\Omega))^2$ and*

$$(m_0^\varepsilon / h_0^\varepsilon, h_0^\varepsilon) \to (v_0^0, b) \quad \text{in } (L^2(\Omega))^2, \qquad (h_0^\varepsilon - b)/\varepsilon \to 0 \quad \text{in } L^2(\Omega)$$

when $\varepsilon \to 0$. Denote by $(h^\varepsilon, v^\varepsilon)$ a sequence of weak solutions of the degenerate viscous shallow-water system (40) with $\mathcal{D} = r_0 u + r_1 h|u|u$ and initial data $h^\varepsilon|_{t=0} = h_0^\varepsilon$, $(hv)|_{t=0} = m_0^\varepsilon$. Then, as ε tends to 0,

$$\begin{aligned} &v^\varepsilon \to v^0 \quad \text{in } L^\infty(0, T; (L^2(\Omega))^2), \\ &(h^\varepsilon - b)/\varepsilon \to 0 \quad \text{in } L^\infty(0, T; L^2(\Omega)), \\ &\nabla(h^\varepsilon / b) \to 0 \quad \text{in } L^2(0, T; L^2(\Omega))^2 \end{aligned}$$

where v^0 is the global strong solution of the viscous lake equations (41) *with v_0^0 as initial data and drag term $\mathcal{D}_{\lim} = r_0 u + r_1 b|u|u$.*

Such an asymptotics is based on the usual weak-strong process using the BD entropy inequality. More precisely, the author establishes a Gronwall type inequality

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \left(h\left|v+\nu\nabla\left(\frac{h}{b}\right)\right|^2 + \left|\frac{h-b}{\varepsilon}\right|^2\right) - \nu r_0 \frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \ln\frac{h}{b} + r_0\int_\Omega |v|^2$$
$$+\nu\int_\Omega b^2\left|\frac{\nabla(h/b)}{\varepsilon}\right|^2 + \int_\Omega r_1 h|v|^3 \le c\left\|\frac{h-b}{\varepsilon}\right\|^2_{L^2(\Omega)}.$$

This inequality is used effectively in the limit process. This asymptotics is similar to what is called the inelastic limit in meteorology. This approximation has been mathematically studied recently in [86] and [128]. The system being considered concerns the usual compressible Navier–Stokes equations but with constant viscosities. In these papers, the authors consider ill-prepared data and three dimensional flows. The first mentioned paper follows the lines in [73] proving that waves are damped close to the boundary and the second one adapts the local method introduced in [118]. Note that all these studies consider reference densities with no vacuum.

The inviscid case. Let us assume formally that $\mathrm{Re} \to \infty$ in (41) to model an inviscid flow and consider a two-dimensional simply-connected bounded domain Ω and adequate boundary conditions. Assuming no drag term and denoting the velocity v instead of v^0 (for simplicity), we get the following system

$$\begin{cases}\partial_t(bv) + \operatorname{div}(bv\otimes v) + b\nabla p = 0,\\ \operatorname{div}(bv) = 0, \qquad bv\cdot n|_{\partial\Omega} = 0.\end{cases} \tag{42}$$

That means a generalization of the standard two-dimensional incompressible Euler equation is obtained if $b \equiv 1$.

A strictly positive bottom function. YUDOVICH's method may be applied using the fact that the relative vorticity ω/b is transported by the flow. More precisely, the inviscid lake equation may be written using a stream-relative vorticity formulation of the following form

$$\begin{cases}\partial_t\omega_P + v\cdot\nabla\omega_P = 0, \quad \omega_P = \omega/b,\\ -\operatorname{div}(\nabla\Psi/b) = \omega, \qquad \omega = \operatorname{curl} v, \qquad \Psi|_{\partial\Omega} = 0.\end{cases} \tag{43}$$

Assuming $b \ge c > 0$ smooth enough, L^p regularity on the stream function Ψ remains true, that is

$$\|\Psi\|_{W^{2,p}(\Omega)} \le Cp\|\omega\|_{L^p(\Omega)}, \tag{44}$$

where C does not depend on p. Such elliptic estimates with non-degenerate coefficients come from [1]–[2]. This result allows D. LEVERMORE, M. OLIVER, E. TITI, in [110], to conclude the existence and uniqueness of global strong solutions.

A degenerate bottom function. This is the case when b vanishes on the boundary (the shore) $\partial\Omega$. In this case, the inviscid lake equation and corresponding boundary conditions read (42). Suppose that φ is a function equivalent to the distance to the boundary, that is $\varphi \in \mathcal{C}^\infty(\overline{\Omega})$, $\Omega = \{\varphi > 0\}$ and $\nabla\varphi \neq 0$ on $\partial\Omega$. Assuming that

$$b = \varphi^a, \tag{45}$$

where $a > 0$ then the problem $(43)_2$ concerning the stream function Ψ may be written in the form

$$-\varphi\Delta\Psi + a\nabla\varphi \cdot \nabla\Psi = \varphi^{a+1}\omega \quad \text{in } \Omega, \qquad \Psi|_{\partial\Omega} = 0. \tag{46}$$

The case when $a = 1$, that is $b = \varphi$, is physically the most natural. This equation belongs to degenerate elliptic equation class.

In [52], the authors prove that, for such degenerate equations, the L^p regularity estimate (44) remains true. The analysis is based on Schauder's estimates of solutions to (46) (see [99,23]) and on a careful analysis of the associated Green function which depends on the degenerate function b.

Using such estimates, the authors are able to follow the lines of the proof given by V. I. YUDOVICH to get the existence and uniqueness of a global strong solution to (42). Moreover, as a corollary, they prove that the boundary condition $v \cdot n|_{\partial\Omega} = 0$ holds for the velocity.

More precisely, they make the following usual definition.

DEFINITION 3.12. *Given $\omega_0/b \in L^\infty(\Omega)$, (v, ω_P) is a weak solution to the vorticity-stream formulation of the two dimensional lake equation with initial data ω_0, provided*

(i) $\omega_P \in L^\infty([0,T] \times \Omega)$ *and* $b\omega_P \in \mathcal{C}^0([0,T]; L^\infty(\Omega) * \text{weak})$,
(ii) $bv = \nabla^\perp K(b\omega_P) \in \mathcal{C}^0([0,T]; L^2(\Omega))$,
(iii) *for all* $\varphi \in \mathcal{C}^1([0,T] \times \overline{\Omega})$ *and* $t_1 \in [0,T]$:

$$\int_\Omega b\varphi(t_1,x)\omega_P(t_1,x)\,\mathrm{d}x - \int_\Omega \varphi(0,x)\omega_0(x)\,\mathrm{d}x = \int_0^{t_1}\int_\Omega (b\partial_t\varphi + bv\cdot\nabla\varphi)\omega_P\,\mathrm{d}x\mathrm{d}t.$$

The main result follows

THEOREM 3.13. (i) *(Regularity) Assume $\omega_P \in L^\infty((0,T) \times \Omega)$ and v with $bv \in L^\infty([0,T]; L^2(\Omega))$ such that* $\operatorname{div}(bv) = 0$ *satisfying the weak formulation. Then $\omega_P \in \mathcal{C}^0([0,T]; L^r(\Omega))$ and $v \in \mathcal{C}^0([0,T]; W^{1,r}(\Omega))$ for all $r < +\infty$. Moreover, there is C such that for all $p \geq 3$,*

$$\|\nabla v\|_{L^p(\Omega)} \leq Cp\|b\omega_P\|_{L^p(\Omega)}.$$

In addition, the following boundary condition on v *follows*

$$v \cdot n = 0 \quad \text{on } \partial\Omega.$$

(ii) *(Existence) For all* $\omega_0/b \in L^\infty(\Omega)$*, there exists a global weak solution* (v, ω_P) *to the vorticity-stream formulation of* (43).

(iii) *(Uniqueness) The weak vorticity-stream solution is unique.*

This result follows from the Yudovich's procedure in constructing the solution as the inviscid limit of solutions of a system with artificial viscosity, which is the analog of the Navier–Stokes with respect to the Euler equations. The key point of the proof is a regularity result for a degenerate system: the depth vanishing on the shore. More precisely, the main part of the paper concerns the following main result. Consider the following system:

$$\begin{cases} \operatorname{div}(bv) = 0 & \text{in } \Omega, \quad (bv)\cdot n_{|\partial\Omega} = 0, \\ \operatorname{curl} v = f & \text{in } \Omega, \end{cases} \tag{47}$$

where

$$\begin{cases} bv \in L^2(\Omega), \\ f \in L^p(\Omega). \end{cases} \tag{48}$$

Using the definition of b given in (45) and the assumptions on Ω given in the introduction, the authors prove the following result on which the existence and uniqueness result is based.

THEOREM 3.14. *If* (v, f) *satisfy* (47) *and* (48) *with* $p \in]2, \infty[$*, then*

$$v \in C^{1-n/p}(\overline{\Omega}), \quad \nabla v \in L^p(\Omega). \tag{49}$$

There is a constant C_p *independent of* (u, ω) *such that*

$$\|v\|_{C^{1-n/p}(\overline{\Omega})} \le C_p \left(\|f\|_{L^p} + \|bv\|_{L^2}\right). \tag{50}$$

In addition

$$v \cdot n = 0 \quad \text{on } \partial\Omega. \tag{51}$$

Moreover, for all $p_0 > 2$*, there is a constant* C *independent of* (v, f) *and* $p \in [p_0, \infty[$ *such that*

$$\frac{1}{p}\|\nabla v\|_{L^p} \le C \left(\|f\|_{L^p} + \|bv\|_{L^2}\right). \tag{52}$$

In this statement, for $\mu \in]0, 1[$, $C^\mu(\overline{\Omega})$ is the usual space of continuous functions on $\overline{\Omega}$ which satisfy the Hölder condition of order μ. In particular, (50) implies that

$$\|v\|_{L^\infty(\Omega)} \le C \left(\|f\|_{L^\infty} + \|bv\|_{L^2}\right). \tag{53}$$

Idea of the proof. Using local coordinates, Eq. (46) on Ψ reads

$$\widetilde{L}\psi = -x_n \widetilde{P}_2 \Psi + a\widetilde{P}_1 \Psi = x_n^{a+1} f$$

with $\widetilde{P}_2$ and $\widetilde{P}_1$, respectively, a uniform second-order elliptic operator and a first order operator with smooth coefficients. Based respectively, on papers [99,23], they define

$$\Psi = \varphi^{a+1}\Phi, \qquad u = \varphi^{-a}\nabla^{\perp}\Psi = \varphi\nabla^{\perp}\Phi + (a+1)\Phi\nabla^{\perp}\varphi.$$

Using parametrics and a suitable change of variable, see [52] for details, the main operator to be studied is

$$\widetilde{\mathcal{L}} = -\widetilde{x}_n \Delta_{\widetilde{x}} - (a+2)\partial_{\widetilde{x}_n}. \tag{54}$$

Note that, according to [99, Lemma 1], the fundamental solution of $\widetilde{\mathcal{L}} = -\widetilde{x}_n \Delta_{\widetilde{x}} - (a + 2)\partial_{\widetilde{x}_n}$ is

$$\widetilde{E}(\widetilde{x}, \widetilde{y}) = \int_0^1 \widetilde{F}(\widetilde{x}, \widetilde{y}, \theta)\, \mathrm{d}\theta$$

with

$$F(\widetilde{x}, \widetilde{y}, \theta) = \gamma(\widetilde{y}_n)^{a+1} A^{-(a+n)}(\theta(1-\theta))^{a/2},$$
$$A^2(\widetilde{x}, \widetilde{y}, \theta) = \theta D^2 + (1-\theta)\overline{D}^2$$

with γ being some constant depending on a and

$$D^2 = |\widetilde{x}_n - \widetilde{y}_n|^2 + |\widetilde{x}' - \widetilde{y}'|^2, \qquad \overline{D}^2 = |\widetilde{x}_n + \widetilde{y}_n|^2 + |\widetilde{x}' - \widetilde{y}'|^2.$$

Hölder's regularity on Φ is proved following P. BOLLEY, J. CAMUS AND G. MÉTIVIER, see [23]. Then the estimate of L^p follows from the properties of Calderon-Zygmund type operators: operator acting on bounded functions with compact support with kernel $T(x, y)$, locally integrable away from the diagonal $\{x = y\}$. We recall for reader's convenience the classical properties that are used:

PROPOSITION 3.15. *Assume that the kernels $K(x, y)$ satisfy on $\boldsymbol{R}^n_+ \times \boldsymbol{R}^n_+$:*

$$|K(x, y)| \le \frac{C}{|x-y|^{n-1}}, \qquad |\partial_x K(x, y)| \le \frac{C}{|x-y|^n}.$$

Then the operator

$$Tf(x) = \int K(x, y) f(y)\, \mathrm{d}y$$

acts as $L^p_{\mathrm{comp}}(\overline{\boldsymbol{R}}_n^+)$ to $\mathcal{C}^\mu(\overline{\boldsymbol{R}}_n^+)$ for all $\mu < 1 - n/p$.

PROPOSITION 3.16. *Assume that T is a bounded operator in $L^2(\boldsymbol{R}^n_+)$ with kernel $K(x, y)$ satisfying $x \neq y$*

$$|K(x,y)| \leq \frac{C}{|x-y|^n}, \qquad |\partial_x K(x,y)| \leq \frac{C}{|x-y|^{n+1}}.$$

Then T maps $L^p(\boldsymbol{R}^n_+)$ to $L^p(\boldsymbol{R}^n_+)$ with norm $O(p)$ for all $p \in [2, +\infty[$.

To use such propositions, we must use parametrics to get a simpler operator similar to (54), in order to apply the Green function and propositions. It remains then to study the influences of the parametrics to conclude the complete operation. The interested reader is referred to [52] for details.

REMARK. To the author's knowledge, there exists only one paper dealing with the derivation of the lake equations from the Euler equations with free surface, locally in time, see [141].

REMARK. It could be very interesting to look at the effect of b on properties that are well known for the two-dimensional Euler equations, see [126]. For instance the effects on the regularity of the boundary of vortex patches, on concentrations and weak solutions with vortex-sheet initial data, or Log-Lipschitz regularity of the velocity field.

Great lake equations. Note that great-lake equations that model the long-time effects of slowly varying bottom topography and weak hydrostatic imbalance on the vertically averaged horizontal velocity of an incompressible fluid possessing a free surface and moving under the force of gravity, have been derived in [61,62]. To get this asymptotic model, R. CAMASSA, D. R. HOLM AND D. LEVERMORE consider the regime, where the Froude number Fr is much smaller than the aspect ratio δ of the shallow domain. The new equations are obtained from the Froude limit (Fr tends to 0) of the Euler equations with free surface (the rigid-lid approximation) at the first order of an asymptotic expansion in δ^2. These equations have local conservation laws of energy and vorticity and are stated as

$$\begin{cases} \partial_t v - u^\perp \mathrm{curl} v + \nabla\left(b - \frac{1}{2}|u|^2 + u\cdot v\right) = 0, \\ v = u + \delta^2\Big((u\cdot\nabla b)\nabla b \\ \quad + \frac{1}{2}b(\operatorname{div} u)\nabla b - \frac{1}{2}b^{-1}\nabla(b^2 u\cdot\nabla b) - \frac{1}{3}b^{-1}\nabla(b^3 \operatorname{div} u)\Big), \\ \operatorname{div}(bu) = 0, \\ (bu)\cdot n = 0, \qquad u|_{t=0} = u_0. \end{cases} \tag{55}$$

To the author's knowledge no mathematical justification has been given for (55). The reader is referred to [8] and [149] for a second order approximation dealing with shallow-water equations. It would be interesting to understand the differences between these systems.

3.7. *An interesting open problem: Open sea boundary conditions*

Let us mention here an open problem which has received a lot of attention from applied mathematicians, especially A. KAZHIKHOV and V. YOUDOVITCH. Consider the Euler equations, or more generally the inviscid lake equations, formulated in a two-dimensional bounded domain. When the boundary is of inflow type, all the velocity components are prescribed. Along an impermeable boundary, the normal component of the velocity vanishes everywhere. This is known as a slip condition. Along a boundary of outflow type, the normal component of the velocity is prescribed for all points of the boundary.

A local existence and uniqueness theorem is proved in a class of smooth solutions by A. KAZHIKHOV in [107]. A global existence theorem is also proved under the assumption that the flow is close to a uniform one and initial data are small. But the question of global existence and uniqueness in the same spirit as the results by V. I. YUDOVICH remains open, see [162]. In this paper, the author shows how the boundary conditions can be augmented in this more general case to obtain a properly posed problem. Under the additional condition $\operatorname{curl} v|_{\Gamma_1} = \boldsymbol{\pi}(x,t)$, where $\boldsymbol{\pi}(x,t)$ is – modulo some necessary restrictions – arbitrary, the author shows the existence, in the two-dimensional case, of a unique solution for all time. The method is constructive, being based on successive approximations, and it brings out clearly the physical basis for the additional conditions. To understand this better, open sea boundary conditions could be helpful for shallow-water equations, for instance with an application to model the strait of Gibraltar. In order to understand the problem more clearly, let us consider the two-dimensional Euler equations

$$\begin{cases} \partial_t u + u \cdot \nabla u + \nabla p = 0, \\ \operatorname{div} u = 0, \end{cases} \tag{56}$$

with the following boundary conditions

$$\begin{cases} u \cdot n|_{\partial\Omega} = 0 & \text{on } \Gamma^0, \\ u = g & \text{on } \Gamma_1 \text{ with } g \cdot n < 0 \quad \text{on } \Gamma^1, \\ u \cdot n = \gamma & \text{on } \Gamma^2, \end{cases} \tag{57}$$

where $\Gamma = \overline{\Gamma^0} \cup \overline{\Gamma^1} \cup \overline{\Gamma^2}$ with compatibility condition

$$\int_{\Gamma^1} g \cdot n \, \mathrm{d}\Gamma^1 + \int_{\Gamma^2} \gamma \, \mathrm{d}\Gamma^2 = 0.$$

In a square domain, using a stream-vorticity formulation, we get

$$\begin{cases} \Delta \Psi = \omega, \\ \partial_t \omega + \partial_y \Psi \partial_x \omega - \partial_x \Psi \partial_y \omega = 0 \end{cases} \tag{58}$$

with the following boundary conditions

$$\begin{cases} \Psi & \text{given on } \Gamma, \\ \partial_x \Psi = \nabla\Psi \cdot n & \text{given on } \Gamma^1. \end{cases} \tag{59}$$

Here, we see two boundary conditions for Ψ on Γ^1. Usually, when the vorticity is given at the beginning, it enables us to conclude a well posed global problem. The main objective is therefore to get some boundary conditions for ω from the equations and the other boundary conditions. Of course this will not be a local boundary condition. Let us explain with a simpler example, why we can hope for some extra boundary condition on the vorticity.

An example to get extra boundary conditions. Let us prove that it is possible, in a particular case, to deduce boundary condition on the vorticity assuming velocity profile boundary conditions. The result is inspired by [51] and concerns the following two-dimensional hydrostatic Navier–Stokes system

$$\begin{cases} \partial_t v + u \cdot \nabla v - \nu \Delta v + \partial_x p = 0, \\ \partial_z p = 0, \qquad \operatorname{div} u = 0, \\ \partial_z v|_{z=0} = 0, \quad w|_{z=0} = 0, \qquad u|_{z=-1} = 0, \end{cases} \tag{60}$$

where $u = (v, w)$ is the velocity field and p the pressure. We consider a periodic boundary condition in the horizontal variable. We denote $\nabla = (\partial_x, \partial_z)$, $\Delta = \partial_x^2 + \partial_z^2$. This system may be written under the following classical stream-vorticity formulation:

$$\begin{cases} \partial_t \omega + u \cdot \nabla\omega - \nu\Delta\omega = 0, \\ \omega|_{z=0} = 0, \end{cases} \tag{61}$$

and

$$\begin{cases} -\Delta\Psi = \omega, \\ \Psi|_{z=-1} = \Psi|_{z=0} = 0, \qquad \partial_z \Psi|_{z=-1} = 0. \end{cases} \tag{62}$$

Note that, in the two-dimension hydrostatic case, the vorticity function is given by $\omega = \partial_z v$. The question concerns the possibility of finding a boundary condition for ω at the bottom, that means for $z = -1$. The answer is positive and we can easily prove that such a boundary condition is

$$\partial_z \omega|_{z=-1} + \omega|_{z=-1} + 2 \int_{-1}^{0} w\omega \, \mathrm{d}z = 0.$$

In the variable bottom case, that is when $z = -h$ instead of $z = -1$, it is given as an oblique boundary condition

$$-h\partial\omega/\partial n + hh'\partial\omega/\partial\tau + (1+(h')^2)^{-1/2} \times\left((hh''+(1+(h')^2))\omega|_b + 2\int_{-h}^{0} w\omega\,\mathrm{d}z\right) = 0.$$

Such boundary conditions then help to prove the existence and uniqueness of a global strong solution for the two-dimensional hydrostatic Navier–Stokes equations in the non-degenerate case, namely assuming $h \geq c > 0$. This also provides the exponential decay in time of the energy. To the author's knowledge, the degenerate case (that means the case where h vanishes on the shore) is an interesting open problem.

3.8. *Multi-level and multi-layers models*

We now give an example where it is important to write multi-level shallow-water equations: it concerns the modeling of the dynamics of water in the Alboran sea and the strait of Gibraltar. In this sea, two layers of water can be distinguished: the surface Atlantic water penetrating into the mediterranean through the strait of Gibraltar and the deeper, denser mediterranean water flowing into the Atlantic. This observation shows that, if we want to use two dimensional models to simulate such phenomena, we have to consider at least two-layers models. The model which is usually used to study this phenomena considers sea water as composed of two immiscible layers of different densities. In this model, waves appear not only at the surface but also at the interface. It is assumed that for the phenomena under consideration, the wavelength is sufficiently large to make the shallow-water approximation in each layer accurate. Therefore the resulting equations form a coupled system of shallow-water equations. Concerning viscous two-layer shallow-water equations, several papers with the diffusion $-h\Delta u$ in each layer have been written, see [134] and [89] the references cited therein. Results in the spirit of [142] (see Section 3.5) are obtained. Note also the recent paper [72], where existence of global weak solutions of a two-layers model with diffusion of the form $-2\mathrm{div}(hD(u))$ (in each layer) has been proved, generalizing BD mathematical entropy. Multilayer Saint-Venant systems have also been studied in [12].

Rigid lid approximation and two-layers. Note that there are very few mathematical studies concerning the propagation of internal-waves in multi-levels geophysical models with Navier–Stokes type system. More precisely, there are no rigid-lid approximation justifications. To the author's knowledge, there exists only one paper written by B. DI MARTINO, P. ORENGA and M. PEYBERNES [79] where they propose a simplified system that allows them to get a global existence and regularity result. This simplification comes from the necessity to get a L^∞ bound on $1/h_i$. To achieve this they replace h_i by $H_i > c > 0$ in the term $\mathrm{div}(h_i u_i)$ in the mass equation. Note also the recent paper [67] concerning interface boundary value problem for the Navier–Stokes equations in thin two-layers domains with a fixed interface.

Two-layers for submarine avalanches. Submarine avalanches or landslides are poorly studied compared to their subaerial counterpart. This is however a key issue in geophysics. Indeed, submarine granular flows driven by gravity participate in the evolution of the sea floor and, in particular, of the continental margins. The first layer is filled with a homogeneous non-viscous fluid with constant density and the second layer is made of a fluidized granular mass. The two fluids are assumed immiscible. One important equation to find the rheological behavior of the fluidized granular mass over a complex topography should be modeled as well as the interaction between the two layers. For instance new Savage-Hutter type models over a general bottom have been proposed in [88]. This model takes into account non-Newtonian behavior.

Two-layers for inviscid irrotational fluids. As for free surface water-waves, see [7], shallow-water type models can be derived from the free surface Euler equations (also called internal wave equations), but many other regimes can be investigated, which also include dispersive effects. In [70], the authors derive formally asymptotic models in one dimensional space. In [24] a systematic derivation of such models has been performed in two horizontal dimensions and the consistency of these models has been rigorously established. Their full justification (in the spirit of [7]) remains however an interesting open problem (to our knowledge the only fully justified model is the Benjamin Ono equation in a one horizontal dimension with presence of surface tension (see [140])).

3.9. *Friction shallow-water equations derivation*

3.9.1. Formal derivation We refer the reader the dissertation [164] for a nice review of free boundary problems for equations of motion of both incompressible and compressible viscous fluids. We also mention two recent papers dedicated to the formal derivation of viscous shallow-water equations from the Navier–Stokes equations with free surface, see [96] for 1D shallow-water equations and see [127] for two dimensional shallow-water equations.

It could be interesting to prove mathematically such formal derivations. Here we make a few remarks concerning hypotheses which have been used to derive formally viscous shallow-water type equations with damping terms.

First hypothesis. The viscosity is of order ε, meaning that the viscosity is of the same order as the depth, and the asymptotic analysis is performed at order 1.

Second hypothesis. The boundary condition at the bottom for the Navier–Stokes equations is given using wall laws. Namely, the boundary conditions are of the form $(\sigma n)_{\mathrm{tang}} = r_0 u$ on the bottom with $u \cdot n = 0$. These boundary conditions can lead to a drag term. (There is no such drag term if $r_0 = 0$).

This part of the Handbook deals with the derivation of, two-dimensional in space, viscous shallow-water equations from the three-dimensional incompressible Navier–Stokes equations with free surface and Navier boundary conditions at the bottom. We take into account the surface tension and we can assume the viscosity at the main order or at order one. We present here another way to derive models previously obtained by J. F. GERBEAU, B. PERTHAME (see [96]) and A. ORON, S. H. DAVIS, S. G. BANKOFF (see [143]). When the viscosity is of order one, this corresponds to applications in hydrology: Rivers, coastal models or Geophysical flows as in oceans. Indeed, the viscosity seems to be negligible at the main order. When the viscosity is at the main order then it corresponds to lubrication models.

I. *Introduction and equations.* The fluid is assumed to be governed by the three-dimensional Navier–Stokes equations for incompressible flows. A free surface is considered on the top boundary, which is expressed as a graph over the flat horizontal bottom. More precisely, the fluid at time t is located in $\{(x,z) \in T^2 \times \mathbf{R}^+ : 0 < z < h(t,x)\}$, where h denotes the depth of the fluid layer.

The fluid velocity u is decomposed into its horizontal and vertical components $v \in \mathbf{R}^2$ and $w \in \mathbf{R}$, respectively, directed along the x and z coordinates. Denoting by p the pressure field, ν the kinematic viscosity, σ the stress tensor, and $D(u) = (\nabla u +^t \nabla u)/2$ the strain tensor, the evolution equations read as

$$\begin{cases} \partial_t v + v \cdot \nabla_x v + w\partial_z v + \nabla_x(p + \Phi) - \nu(\Delta_x + \partial_z^2)v = 0, \\ \partial_t w + v \cdot \nabla_x w + w\partial_z w + \partial_z(p + \Phi) - \nu(\Delta_x + \partial_z^2)w = 0, \\ \mathrm{div}_x v + \partial_z w = 0, \end{cases}$$

where

$$\Phi = \overline{A}(2yh)^{-3} + \overline{g}z$$

denotes the potential associated with bulk forces, see [143]. On the bottom boundary, friction conditions are considered

$$\nu\partial_z v = \overline{\alpha}_b |v|^{\beta_b} v, \quad w = 0$$

and on the upper surface $\{(x, h(t,x)) \ / \ x \in T^2\}$, we require that the stress tensor $\sigma = 2\nu D(u) - pI$ satisfies

$$(\sigma \cdot n)_{\tan} = -\overline{\alpha}_t |u_{\tan}|^{\beta_t} u_{\tan}, \quad (\sigma \cdot n) \cdot n = -2H\overline{\kappa}, \quad \partial_t h + v \cdot \nabla_x h = w,$$

where for any vector $f \in \mathbf{R}^3$, the notation $f_{\tan}$ stands for $f_{\tan} = f - (f \cdot n)n$, n denoting the outer normal to the boundary and H the mean curvature of the free surface. The expressions of n and H can be given in terms of the depth h

$$n = \frac{1}{\sqrt{1 + |\nabla_x h|^2}}(-\nabla_x h, 1)^t,$$

and

$$-2H = \mathrm{div}_x \left(\frac{\nabla_x h}{\sqrt{1 + |\nabla_x h|^2}} \right).$$

Let us rewrite the boundary conditions on the upper surface in a more convenient way. Denoting by J the expression $\sqrt{1+|\nabla_x h|^2}$, we have

$$(\sigma\cdot n)_{\tan} = \frac{\nu}{J^3}\left|\begin{array}{l} 2\partial_z w\nabla_x h + (\nabla_x w + \partial_z v)\, J^2 - 2\left((\nabla_x w + \partial_z v)\cdot\nabla_x h\right)\nabla_x h \\ \quad -2\left(J^2 D_x(v)\cdot\nabla_x h - (\nabla_x h\cdot D_x(v)\cdot\nabla_x h)\nabla_x h\right) \\ 2\partial_z w|\nabla_x h|^2 - (\nabla_x w + \partial_z v)\cdot\nabla_x h\left(|\nabla_x h|^2 - 1\right) \\ \quad -2\nabla_x h\cdot D_x(v)\cdot\nabla_x h, \end{array}\right.$$

where $u_{\tan}$ is given by

$$u_{\tan} = \frac{1}{J^2}\left|\begin{array}{l} vJ^2 - (v\cdot\nabla_x h)\nabla_x h + w\nabla_x h, \\ w|\nabla_x h|^2 + v\cdot\nabla_x h. \end{array}\right.$$

The normal stress conditions are expressed as

$$\begin{aligned}(\sigma\cdot n)\cdot n &= \frac{\nu}{J^2}\left(2\nabla_x h\cdot(D_x v)\cdot\nabla_x h - 2\nabla_x h\cdot(\nabla_x w + \partial_z v) + 2\partial_z w\right) - p \\ &= \overline{\kappa}\,\mathrm{div}_x\left(\frac{\nabla_x h}{J}\right).\end{aligned}$$

II. *Derivation of the SW equations: Viscosity at main order.*
The scaled Navier–Stokes equations. The motion of an incompressible fluid in the basin described above is governed by the incompressible three-dimensional Navier–Stokes equations. We introduce non-dimensional variables in terms of natural units. More precisely, we rescale the coordinates and unknowns as follows:

$$\begin{aligned} &\tau = k^2 t, \quad \xi = kx, \quad \zeta = z, \\ &v(t,x,z) = kV(\tau,\xi,\zeta), \quad w(t,x,z) = k^2 W(\tau,\xi,\zeta), \\ &p(t,x,z) = k^2 P(\tau,\xi,\zeta), \\ &h(t,x) = H(\tau,\xi), \quad \overline{A} = k^2 A, \quad \overline{g} = k^2 g, \\ &\overline{\alpha}_b = k\alpha_b, \quad \overline{\alpha}_t = k\alpha_t, \quad \overline{\kappa} = \kappa, \quad \nu = \nu_0 \approx 1. \end{aligned}$$

In the rescaled variables, the non-dimensional form of the evolution equations become

$$\begin{cases} k^3\left(\partial_\tau V + V\cdot\nabla_\xi V + W\partial_\zeta V\right) + k^3\nabla_\xi(P+\Phi) - k^3\nu(\Delta_\xi + \dfrac{1}{k^2}\partial_\zeta^2)V = 0, \\ k^4\left(\partial_\tau W + V\cdot\nabla_\xi W + W\partial_\zeta W\right) + k^2\partial_\zeta(P+\Phi) - k^4\nu(\Delta_\xi + \dfrac{1}{k^2}\partial_\zeta^2)W = 0, \\ \mathrm{div}_\xi V + \partial_\zeta W = 0. \end{cases}$$

We note that the non-dimensional form of the incompressibility condition contains no small parameters. On the free surface, the condition on the tangential part of the stress tensor is rewritten in a scaled form as

$$
\begin{aligned}
&2k^3 \partial_\zeta W \nabla_\xi H + k^3 \left(\nabla_\xi W + \frac{1}{k^2} \partial_\zeta V \right) \left(1 + k^2 |\nabla_\xi H|^2 \right) \\
&\quad - 2k^5 \left(\left(\nabla_\xi W + \frac{1}{k^2} \partial_\zeta V \right) \cdot \nabla_\xi H \right) \nabla_\xi H \\
&\quad - 2k^3 \left(D_\xi(V) \cdot \nabla_\xi H \left(1 + k^2 |\nabla_\xi H|^2 \right) - k^2 (\nabla_\xi H \cdot D_\xi(V) \cdot \nabla_\xi H) \nabla_\xi H \right) \\
&\quad\quad = - \left(1 + k^2 |\nabla_\xi H|^2 \right)^{3/2} \frac{\alpha_t}{\nu} k |U_{\text{k, tan}}|^{\beta_t} U_{\text{k, tan horiz}},
\end{aligned}
$$

where

$$
U_{\text{k,tan}} = \frac{1}{1 + k^2 |\nabla_\xi H|^2} \left| \begin{array}{l} kV \left(1 + k^2 |\nabla_\xi H|^2 \right) - k^3 (V \cdot \nabla_\xi H) \nabla_\xi H \\ \quad + k^3 W \nabla_\xi H, \\ k^4 W |\nabla_\xi H|^2 + k^2 V \cdot \nabla_\xi H. \end{array} \right.
$$

The normal stress condition is rewritten similarly as

$$
\begin{aligned}
&\frac{\nu k^4}{1 + k^2 |\nabla_\xi H|^2} \left(2 \nabla_\xi H \cdot (D_\xi V) \cdot \nabla_\xi H - 2 \nabla_\xi H \cdot \left(\nabla_\xi W + \frac{1}{k^2} \partial_\zeta V \right) \right. \\
&\quad \left. + \frac{2}{k^2} \partial_\zeta W \right) - k^2 P = \kappa k^2 \text{div}_\xi \left(\frac{\nabla_\xi H}{\sqrt{1 + k^2 |\nabla_\xi H|^2}} \right).
\end{aligned}
$$

Finally the bottom boundary conditions are rewritten as

$$
k \nu \partial_\zeta V = k^2 \alpha_b |V|^{\beta_b} V, \qquad W = 0.
$$

The change of variables. A classical change of variables allows to us rewrite the system in a fixed domain $T_x^2 \times (0, 1)_z$ by using the following change of variables.

$$
\tau = t, \qquad \xi = X, \qquad \zeta = H(t, X) Z.
$$

The velocity and pressure variables are rewritten as

$$
\begin{aligned}
&\widetilde{V}(\tau, X, Z) = V(\tau, X, HZ), \qquad \widetilde{W}(\tau, X, Z) = W(\tau, X, HZ), \\
&\widetilde{P}(\tau, X, Z) = P(\tau, X, HZ).
\end{aligned}
$$

We easily find that for any function f

$$
\begin{aligned}
&(\partial_\tau f) \circ (t, X, HZ) = \partial_\tau \widetilde{f} - \partial_Z \widetilde{f} \frac{\partial_\tau H}{H} Z, \\
&(\nabla_\xi f) \circ (t, X, HZ) = \nabla_X \widetilde{f} - \frac{\nabla_X H \partial_Z \widetilde{f}}{H} Z.
\end{aligned}
$$

Moreover,

$$(\Delta_\xi f)\circ(t,X,HZ)=\Delta_X\widetilde{f}-\frac{2\nabla_X H\cdot\partial_Z\nabla_X\widetilde{f}}{H}Z \\ -\frac{\Delta_X H}{H}Z\partial_Z\widetilde{f}+2\frac{|\nabla_X H|^2}{H^2}Z\partial_Z\widetilde{f}+\frac{\partial_Z^2\widetilde{f}Z^2|\nabla_x H|^2}{H^2}.$$

We also have

$$(\partial_\zeta f)\circ(t,X,HZ)=\frac{\partial_Z\widetilde{f}}{H}$$

and therefore

$$(\partial_\zeta^2 f)\circ(t,X,HZ)=\frac{\partial_Z^2\widetilde{f}}{H^2}.$$

The asymptotic analysis. We assume that there exists an Ansatz in powers of k^2 that means

$$\widetilde{V}=\widetilde{V}_0+k^2\widetilde{V}_1+\cdots,\qquad \widetilde{W}=\widetilde{W}_0+k^2\widetilde{W}_1+\cdots,$$
$$\widetilde{P}=\widetilde{P}_0+k^2\widetilde{P}_1+\cdots,\qquad H=H_0+k^2H_1+\cdots.$$

Leading-order. We obtain at the leading order from the horizontal part of the momentum equation

$$-\nu\partial_Z^2\widetilde{V}_0=0,$$

whereas the boundary conditions yield

$$\partial_Z\widetilde{V}_0=0\quad\text{on }\{Z=0\},\qquad \partial_\zeta\widetilde{V}_0=0\quad\text{on }\{Z=1\}.$$

It means that $\widetilde{V}_0$ is a function of X and τ only, which means that for the leading order horizontal velocity no Z dependence will develop, provided the initial state is ζ–independent. The divergence equation

$$\frac{\partial_Z\widetilde{W}_0}{H_0}+\mathrm{div}_X\widetilde{V}_0=0,$$

combined with the boundary condition on the bottom $\widetilde{W}_0|_{Z=0}=0$ gives the expression of the leading-order vertical velocity

$$\frac{\widetilde{W}_0}{H_0}=-Z\mathrm{div}\,_X\widetilde{V}_0.$$

The kinematic condition on the free surface provides at the leading order

$$\partial_\tau H_0 + \widetilde{V}_0 \cdot \nabla_X H_0 = \widetilde{W}_0|_{Z=1}.$$

With the previous expression of $\widetilde{W}_0$ taken at $Z = 1$, we obtain the well known transport equation of the shallow-water equations

$$\partial_\tau H_0 + \mathrm{div}_X(H_0 \widetilde{V}_0) = 0.$$

The vertical component of the momentum equation at leading order k^2 yields

$$\frac{\nu}{H_0} \partial_Z^2 \widetilde{W}_0 - \partial_Z(\widetilde{P}_0 + \widetilde{\Phi}_0) = 0,$$

from which we deduce

$$\widetilde{P}_0 = -\widetilde{\Phi}_0 + \frac{\nu}{H_0} \partial_Z \widetilde{W}_0 + D(\tau, X),$$

where D is some arbitrary function of τ and X. Substituting $\widetilde{P}_0$ into the leading order of the upper boundary condition on $(\sigma \cdot n) \cdot n$, we obtain

$$-2\nu \frac{\partial_Z \widetilde{V}_0}{H_0} \cdot \nabla_\xi H_0 + 2\nu \frac{\partial_Z \widetilde{W}_0}{H_0} - \widetilde{P}_0 = \kappa \Delta_X H_0.$$

Using the fact that $\widetilde{V}_0$ does not depend on ζ, we deduce

$$D(\tau, \xi) = -\nu \mathrm{div}_X \widetilde{V}_0 + \widetilde{\Phi}_0 - \kappa \Delta_X H_0,$$

hence

$$\widetilde{P}_0 = -2\nu \mathrm{div}_X \widetilde{V}_0 - \kappa \Delta_X H_0.$$

Next order. Coming back to order k^3 in the horizontal component of the momentum equation, we derive

$$\partial_\tau \widetilde{V}_0 + \widetilde{V}_0 \cdot \nabla_X \widetilde{V}_0 = \nu \Delta_X \widetilde{V}_0 + \frac{\nu}{H_0^2} \partial_Z^2 \widetilde{V}_1 - \nabla_X(\widetilde{P}_0 + \widetilde{\Phi}_0).$$

This equation can be integrated from 0 to 1 in the Z variable using the top and bottom boundary conditions

$$\frac{\nu}{H_0} \partial_Z \widetilde{V}_1 = \alpha_b |\widetilde{V}_0|^{\beta_b} \widetilde{V}_0 \quad \text{on } \{Z = 0\},$$

$$-2\nu D_X(\widetilde{V}_0) \cdot \nabla_X H_0 + \nu \nabla_X \widetilde{W}_0 + \frac{\nu}{H_0} \partial_Z \widetilde{V}_1 + \frac{\nu \nabla_X H_0}{H_0} \partial_Z \widetilde{W}_0$$
$$= -\alpha_t |\widetilde{V}_0 - V_{\mathrm{air}}|^{\beta_t} (\widetilde{V}_0 - V_{\mathrm{air}}) \quad \text{on } \{Z = 1\}.$$

The shallow-water equations. The above conditions, together with the expression of D, $(P_0 + \Phi_0)$, and W_0 as functions of V_0, finally give the following shallow-water equations integrating with respect to Z from 0 to 1

$$\begin{aligned}
&\partial_\tau(H_0\widetilde{V}_0) + \mathrm{div}_X(H_0\widetilde{V}_0 \otimes \widetilde{V}_0) + H_0\nabla_X\Phi_0|_{Z=1} \\
&\quad + \alpha_b|\widetilde{V}_0|^{\beta_b}V_0 + \alpha_t|\widetilde{V}_0 - V_{\mathrm{air}}|^{\beta_t}(V_0 - V_{\mathrm{air}}) \\
&\qquad = \kappa H_0\nabla_X\Delta_X H_0 + \mathrm{div}_X\left(2\nu H_0 D_X(\widetilde{V}_0)\right) + \nabla_X\left(2\nu H_0 \mathrm{div}_X\widetilde{V}_0\right).
\end{aligned}$$

The term involving the limit bulk forces is rewritten as

$$H_0\nabla_X\widetilde{\Phi}_0|_{Z=1} = \frac{3}{2}A\nabla_X\frac{1}{H_0^2} + \frac{g}{2}\nabla_X H_0^2.$$

III. *Derivation of the SW equations: Viscosity at order one.* In this section, the viscosity is assumed to scale at order 1 with respect to k, that means $\nu = k\nu_0$. We will find, in another way, the system obtained in [127]. For the sake of simplicity, we assume $\beta_t = \beta_b = 0$, which means we consider a laminar drag on the top and on the bottom.

The scaled Navier–Stokes equations. We rescale the coordinates and unknowns as follows:

$$\begin{aligned}
&\tau = k^2 t, \quad \xi = kx, \quad \zeta = z, \\
&v(t,x,z) = kV(\tau,\xi,\zeta), \quad w(t,x,z) = k^2 W(\tau,\xi,\zeta), \\
&\quad p(t,x,z) = k^3 P(\tau,\xi,\zeta), \\
&h(t,x) = H(\tau,\zeta), \quad \overline{A} = k^3 A, \quad \overline{g} = k^3 g, \\
&\overline{\alpha}_b = k^3\alpha, \quad \overline{\alpha}_t = k^4\alpha_t, \quad \overline{\kappa} = k\kappa, \quad \nu = k\nu_0.
\end{aligned}$$

The equations are rewritten as:

$$\begin{cases}
\partial_\tau V + V\cdot\nabla_\xi V + W\partial_\zeta V + k\nabla_\xi(P+\Phi) \\
\quad - k\nu_0\left(\Delta_\xi + \dfrac{1}{k^2}\partial_\zeta^2\right)V = 0, \\
\partial_\tau W + V\cdot\nabla_\xi W + W\partial_\zeta W + \dfrac{1}{k}\partial_\zeta(P+\Phi) \\
\quad - k\nu_0\left(\Delta_\xi + \dfrac{1}{k^2}\partial_\zeta^2\right)W = 0, \\
\mathrm{div}_\xi V + \partial_\zeta W = 0.
\end{cases} \tag{63}$$

The top boundary conditions on the tangential part of the stress tensor rescale as

$$2k^4\partial_\zeta W\nabla_\xi H + k^4\left(\nabla_\xi W + \frac{1}{k^2}\partial_\zeta V\right)\left(1+k^2|\nabla_\xi H|^2\right)$$
$$- 2k^6\left(\left(\nabla_\xi W + \frac{1}{k^2}\partial_\zeta V\right)\cdot\nabla_\xi H\right)\nabla_\xi H$$
$$- 2k^4\left(D_\xi(V)\cdot\nabla_\xi H\left(1+k^2|\nabla_\xi H|^2\right) - k^2(\nabla_\xi H\cdot D_\xi(V)\cdot\nabla_\xi H)\nabla_\xi H\right)$$
$$= -\left(1+k^2|\nabla_\xi H|^2\right)^{3/2}\frac{\alpha_t}{\nu_0}k^3|U_{\text{k, tan}}|^{\beta_t}U_{\text{k, tan horiz}},$$

where

$$U_{\text{k,tan}} = \frac{1}{1+k^2|\nabla_\xi H|^2}\left|\begin{array}{l} kV\left(1+k^2|\nabla_\xi H|^2\right) \\ \quad - k^3(V\cdot\nabla_\xi H)\nabla_\xi H + k^3W\nabla_\xi H, \\ k^4W|\nabla_\xi H|^2 + k^2V\cdot\nabla_\xi H. \end{array}\right.$$

The normal stress condition is rewritten as

$$\frac{\nu_0 k^5}{1+k^2|\nabla_\xi H|^2}\Bigg(2\nabla_\xi H\cdot(D_\xi V)\cdot\nabla_\xi H - 2\nabla_\xi H$$
$$\cdot\left(\nabla_\xi W + \frac{1}{k^2}\partial_\zeta V\right) + \frac{2}{k^2}\partial_\zeta W\Bigg) - k^3P = \kappa k^3\text{div}_\xi\left(\frac{\nabla_\xi H}{\sqrt{1+k^2|\nabla_\xi H|^2}}\right).$$

Notice that the pressure has been rescaled in such a way that it does not vanish at the leading order.

The change of variables. As usual in the free surface case, we write the system in a fixed domain using the following change of variables.

$$\xi = X, \qquad \zeta = H(t,x)Z.$$

Then the new domain will be $T^2\times(0,1)$. Let us indicate the change of variables thus

$$\widetilde{V}(\tau,X,Z) = V(\tau,X,HZ), \qquad \widetilde{W}(\tau,X,Z) = W(\tau,X,HZ),$$
$$\widetilde{P}(\tau,X,Z) = P(\tau,X,HZ).$$

We easily find that for any function f

$$(\partial_\tau f)\circ(t,X,HZ) = \partial_\tau\widetilde{f} - \partial_Z\widetilde{f}\frac{\partial_\tau H}{H}Z,$$
$$(\nabla_\xi f)\circ(t,X,HZ) = \nabla_X\widetilde{f} - \frac{\nabla_X H\partial_Z\widetilde{f}}{H}Z.$$

Moreover,

$$(\Delta_\xi f) \circ (t, X, HZ) = \Delta_X \widetilde{f} - \frac{2\partial_Z \nabla_X \widetilde{f} \cdot \nabla_X H}{H} Z - \frac{\Delta_X H}{H} \partial_Z \widetilde{f} Z + 2\frac{|\nabla_X H|^2}{H^2} \partial_Z \widetilde{f} Z + \frac{\partial_Z^2 \widetilde{f} Z^2 |\nabla_x H|^2}{H^2}.$$

We also have

$$(\partial_\zeta f) \circ (t, X, HZ) = \frac{\partial_Z \widetilde{f}}{H}$$

and therefore

$$(\partial_\zeta f) \circ (t, X, HZ) = \frac{\partial_Z^2 \widetilde{f}}{H^2}.$$

The asymptotic analysis. Now, using the previous relations and the equations satisfied by (V, W, P), we can develop the system using an Ansatz with respect to k on $(\widetilde{V}, \widetilde{W}, \widetilde{P})$, giving

$$\begin{cases} \widetilde{V} = \widetilde{V}_0 + k\widetilde{V}_1 + \cdots, \\ \widetilde{W} = \widetilde{W}_0 + k\widetilde{W}_1 + \cdots, \\ \widetilde{P} = \widetilde{P}_0 + k\widetilde{P}_1 + \cdots. \end{cases}$$

Leading-order. We obtain at the leading order from the horizontal part of the momentum equation

$$\frac{\nu_0}{H_0^2} \partial_Z^2 \widetilde{V}_0 = 0.$$

whereas the boundary conditions yield

$$\begin{cases} \frac{\nu_0}{H_0} \partial_Z \widetilde{V}_0|_{Z=0} = 0, \\ \frac{\nu_0}{H_0} \partial_Z \widetilde{V}_0|_{Z=1} = 0. \end{cases}$$

As in the previous section, it means that V_0 is a function of τ and X only, since

$$\partial_Z \widetilde{V}_0 = 0. \tag{64}$$

That means for the leading order horizontal velocity that no Z dependence will develop, provided the initial state is Z-independent.

The divergence-free condition at the leading order,

$$\mathrm{div}_X \widetilde{V}_0 + \frac{\partial_Z \widetilde{W}_0}{H_0} = \frac{\partial_Z \widetilde{V}_0 \cdot \nabla_X H_0}{H_0}$$

reads

$$\mathrm{div}_X \widetilde{V}_0 + \frac{\partial_Z \widetilde{W}_0}{H_0} = 0$$

which gives, with the boundary condition on the bottom $\widetilde{W}_0|_{Z=0} = 0$, the expression

$$\widetilde{W}_0 = -Z H_0 \mathrm{div}_X \widetilde{V}_0. \tag{65}$$

The kinematic boundary condition on the free surface provides at the leading order

$$(\partial_\tau H_0 + \widetilde{V}_0 \cdot \nabla_X H_0)|_{Z=1} = \widetilde{W}_0|_{Z=1}. \tag{66}$$

We obtain at the leading order from the vertical part of the momentum equation

$$\frac{\nu_0}{H_0}\partial_Z^2 \widetilde{W}_0 - \partial_Z(\widetilde{P}_0 + \widetilde{\Phi}_0) = 0,$$

with the boundary conditions

$$\begin{cases} \widetilde{W}_0|_{Z=0} = 0, \\ 2\dfrac{\nu_0}{H_0}\partial_Z \widetilde{W}_0 - \widetilde{P}_0|_{Z=1} - \kappa \Delta_X H_0|_{Z=1} = 0. \end{cases}$$

Using the first equation, we get

$$\widetilde{P}_0 + \widetilde{\Phi}_0 = \frac{\nu_0}{H_0}\partial_Z \widetilde{W}_0 + D(X,\tau) = -\nu_0 \mathrm{div}_X \widetilde{V}_0 + D(X,\tau). \tag{67}$$

Now, we use the boundary condition on the top to determine D which only depends on X and τ. We have

$$\widetilde{P}_0|_{Z=1} = -2\nu_0 \mathrm{div}_X \widetilde{V}_0 - \kappa \Delta_X H_0.$$

Thus

$$(\widetilde{P}_0 + \widetilde{\Phi}_0)|_{Z=1} = -2\nu_0 \mathrm{div}_X \widetilde{V}_0 + \widetilde{\Phi}_0|_{Z=1} - \kappa \Delta_X H_0.$$

We get

$$D = -\kappa \Delta_X H_0 + \widetilde{\Phi}_0|_{Z=1} - \nu_0 \mathrm{div}_X \widetilde{V}_0.$$

Therefore we obtain from (67)

$$P_0 + \Phi_0 = -2\nu_0 \mathrm{div}_X \widetilde{V}_0 + \widetilde{\Phi}_0|_{Z=1} - \kappa \Delta_X H_0. \tag{68}$$

Order 1. Using (64) and coming back to order 1, we get from the horizontal component of the momentum equations

$$\partial_\tau \widetilde{V}_0 + \widetilde{V}_0 \cdot \nabla_X \widetilde{V}_0 = \frac{\nu_0}{H_0} \partial_Z^2 \widetilde{V}_1 \tag{69}$$

with the boundary conditions

$$\begin{cases} \dfrac{\nu_0}{H_0} \partial_Z \widetilde{V}_1|_{Z=0} = \alpha \widetilde{V}_0, \\ \dfrac{\nu_0}{H_0} \partial_Z \widetilde{V}_1|_{Z=1} = 0. \end{cases}$$

Therefore integrating with respect to Z from 0 to 1, we get

$$\partial_\tau \widetilde{V}_0 + \widetilde{V}_0 \cdot \nabla_X \widetilde{V}_0 = \frac{\nu_0}{H_0^2} \partial_Z \widetilde{V}_1|_1 - \frac{\nu_0}{H_0^2} \partial_Z \widetilde{V}_1|_0 = -\frac{\alpha}{H_0} \widetilde{V}_0.$$

Thus

$$\partial_\tau \widetilde{V}_0 + \widetilde{V}_0 \cdot \nabla_X \widetilde{V}_0 = -\frac{\alpha}{H_0} \widetilde{V}_0. \tag{70}$$

This gives

$$H_0(\partial_\tau \widetilde{V}_0 + \widetilde{V}_0 \cdot \nabla_X \widetilde{V}_0) + \alpha \widetilde{V}_0 = 0.$$

Moreover, coming back to (69) and using the previous equality, we get

$$\widetilde{V}_1 = \frac{\alpha H_0 \widetilde{V}_0}{\nu_0} Z \left(1 - \frac{Z}{2}\right) + C(\tau, X). \tag{71}$$

Using the divergence-free condition (65), Eq. (66) gives

$$\partial_\tau H_0 + \mathrm{div}_X (H_0 \widetilde{V}_0) = 0. \tag{72}$$

In conclusion (69) and (72) provide the standard hyperbolic Saint-Venant equation with a friction term at the main order. That means the following system

$$\begin{cases} \partial_\tau H_0 + \mathrm{div}_v (H_0 \widetilde{V}_0) = 0, \\ \partial_\tau (H_0 V_0) + \mathrm{div}_x (H_0 \widetilde{V}_0 \otimes \widetilde{V}_0) + \alpha \widetilde{V}_0 = 0. \end{cases}$$

Order k. We have

$$\partial_\tau \widetilde{V}_1 - \partial_Z \widetilde{V}_1 \left(\frac{\partial_t H_0}{H_0} Z + \frac{\widetilde{V}_0 \cdot \nabla_X H_0}{H_0} Z \right) + \widetilde{V}_1 \cdot \nabla_X \widetilde{V}_0$$
$$+ \widetilde{V}_0 \cdot \nabla_X \widetilde{V}_1 + \widetilde{W}_0 \frac{\partial_Z \widetilde{V}_1}{H_0} + \nabla_X (\widetilde{P}_0 + \widetilde{\Phi}_0)$$
$$- \nu_0 \Delta_X \widetilde{V}_0 - \frac{\nu_0}{H_0^2} \partial_Z^2 \widetilde{V}_2 + \frac{2\nu_0 H_1}{H_0^3} \partial_Z^2 \widetilde{V}_1 = 0.$$

The boundary condition for $\widetilde{V}_2$ is given by

$$\begin{cases} \dfrac{\nu_0}{H_0} \partial_Z \widetilde{V}_2|_{Z=0} = \dfrac{\nu_0 H_1}{H_0^2} \partial_Z \widetilde{V}_1 + \alpha_0 \widetilde{V}_1|_{Z=0} = \alpha_0 \widetilde{V}_1|_{Z=0} + \dfrac{\alpha H_1}{H_0} \widetilde{V}_0, \\ \Big(\dfrac{\nu_0}{H_0} \nabla_X H_0 \partial_Z \widetilde{W}_0 + \nu_0 \nabla_X \widetilde{W}_0 + \dfrac{\nu_0}{H_0} \partial_Z \widetilde{V}_2 \\ \quad - \dfrac{\nu_0 H_1}{H_0^2} \partial_Z \widetilde{V}_1 - 2 D_X \widetilde{V}_0 \cdot \nabla_X H_0 \Big) |_{Z=1} = 0. \end{cases}$$

Let us now give the equation satisfied by the average velocity

$$\overline{\widetilde{V}_1} = \int_0^1 \widetilde{V}_1 \mathrm{d}z$$

over the unit depth. We get

$$\partial_\tau \overline{\widetilde{V}_1} - \frac{(\partial_t H_0 + \widetilde{V}_0 \cdot \nabla_X H_0)}{H_0} \int_0^1 \partial_Z \widetilde{V}_1 Z + \mathrm{div_X}(\widetilde{V}_0 \otimes \overline{\widetilde{V}_1}) + \overline{\widetilde{V}_1} \cdot \nabla_X \widetilde{V}_0$$
$$+ \frac{(\widetilde{W}_0 \widetilde{V}_1)|_{Z=1}}{H_0} + \nabla_X (\widetilde{P}_0 + \widetilde{\Phi}_0) - \nu_0 \Delta_X \widetilde{V}_0$$
$$- \frac{\nu_0}{H_0} [\frac{1}{H_0} \partial_Z \widetilde{V}_2|_{Z=1} - \frac{1}{H_0} \partial_Z \widetilde{V}_2|_{Z=0}] + \frac{2\nu_0 H_1}{H_0^3} [\partial_Z \widetilde{V}_1|_{Z=1} - \partial_Z \widetilde{V}_1|_{Z=0}] = 0.$$

Now we use the fact that

$$\int_0^1 \partial_Z \widetilde{V}_1 Z = \widetilde{V}_1|_{Z=1} - \int_0^1 \widetilde{V}_1$$

and

$$\partial_t H_0 + \widetilde{V}_0 \cdot \nabla_X H_0 = \widetilde{W}_0|_{Z=1}$$

to get that

$$\begin{aligned}
&\partial_\tau \overline{\widetilde{V_1}} - \frac{(\partial_t H_0 + \widetilde{V_0}\cdot\nabla_X H_0)}{H_0}\overline{\widetilde{V_1}} + \mathrm{div}_X(\widetilde{V_0}\otimes\overline{\widetilde{V_1}}) + \overline{\widetilde{V_1}}\cdot\nabla_X\widetilde{V_0} \\
&\quad + \nabla_X(\widetilde{P_0}+\widetilde{\Phi}_0) - \nu_0\Delta_X\widetilde{V_0} \\
&\quad - \frac{\nu_0}{H_0^2}\partial_Z\widetilde{V_2}|_{Z=1} + \frac{1}{H_0}\left(\alpha\widetilde{V_1}|_{Z=0} + \frac{\alpha H_1}{H_0}\widetilde{V_0}\right) - \frac{2\alpha H_1}{H_0^2}\widetilde{V_0} = 0.
\end{aligned}$$

Now using the boundary condition on $\partial_Z\widetilde{V_2}$ and on $\partial_Z\widetilde{V_1}$ on the top and the fact that $\widetilde{V_1}|_{Z=0} = C$ by (71), we get

$$\begin{aligned}
&H_0\partial_\tau\overline{\widetilde{V_1}} - (\partial_t H_0 + \widetilde{V_0}\cdot\nabla_X H_0 + H_0\mathrm{div}_X\widetilde{V_0})\overline{\widetilde{V_1}} \\
&\quad + H_0\widetilde{V_0}\nabla_X\overline{\widetilde{V_1}} + H_0\overline{\widetilde{V_1}}\cdot\nabla_X\widetilde{V_0} \\
&\quad + H_0\nabla_X(\widetilde{P_0}+\widetilde{\Phi}_0) - \nu_0 H_0\Delta_X\widetilde{V_0} \\
&\quad - \frac{\nu_0}{H_0}\nabla_X H_0\partial_Z\widetilde{W_0} + \nu_0\nabla_X\widetilde{W_0} - 2D_X\widetilde{V_0}\cdot\nabla_x H_0 + \alpha C.
\end{aligned}$$

Now using the fact that $W_0 = -H_0\mathrm{div}_X\widetilde{V_0}$ and (68), it gives

$$\begin{aligned}
&H_0\partial_\tau\overline{\widetilde{V_1}} + H_0\overline{\widetilde{V_1}}\cdot\nabla_x\widetilde{V_0} + H_0\widetilde{V_0}\cdot\nabla_X\overline{\widetilde{V_1}} \\
&\quad = \kappa H_0\nabla_X\Delta_X H_0 + 2\nu_0\mathrm{div}_X\left(H_0 D_X(\widetilde{V_0})\right) \\
&\qquad + 2\nu_0\nabla_X\left(H_0\mathrm{div}_X\widetilde{V_0}\right) - \alpha_0 C.
\end{aligned} \tag{73}$$

The shallow-water equations at order 2. Let us now derive the shallow-water equations at order 2. We just recall, using (71), that

$$\overline{\widetilde{V_1}} = \int_0^1 \widetilde{V_1} = \frac{\alpha_0 H_0}{3\nu_0}\widetilde{V_0} + C. \tag{74}$$

The main idea is to write the equation satisfied by the mean velocity

$$\mathcal{V}_k = \left(1 + k\frac{\alpha_0 H_0}{3\nu_0}\right)(\widetilde{V_0} + kC)$$

and the height

$$\mathcal{H}_k = H_0 + kH_1.$$

We remark that

$$\mathcal{V}_\parallel = \widetilde{V_0} + k\overline{\widetilde{V_1}} + k^2\frac{\alpha H_0}{3\nu_0}C.$$

Using (70), (73) and (74), we can write that $(\mathcal{V}_k, \mathcal{H}_k)$ satisfies

$$\begin{aligned}
&\partial_\tau(\mathcal{H}_k\mathcal{V}_k) + \mathrm{div}_X(\mathcal{H}_k\mathcal{V}_k \otimes \mathcal{V}_k) \\
&\quad = k\kappa\mathcal{H}_k\nabla_X\Delta_X\mathcal{H}_k + 2k\nu_0\,\mathrm{div}_X\,(\mathcal{H}_k D_X(\mathcal{V}_k)) \\
&\qquad + 2k\nu_0\nabla_X\,(\mathcal{H}_k\mathrm{div}_X\mathcal{V}_k) - \frac{\alpha_0}{1+\frac{k\alpha_0\mathcal{H}_k}{3\nu_0}}\mathcal{V}_k + \mathcal{O}(k^2).
\end{aligned}$$

Let us now prove that $\mathcal{H}_k$ satisfies

$$\partial_\tau\mathcal{H}_k + \mathrm{div}_X(\mathcal{H}_k\mathcal{V}_k) = \mathcal{O}(k^2). \tag{75}$$

At order 1 on the kinematic condition we have

$$\partial_t H_1 + \widetilde{V}_1|_{Z=1}\cdot H_0 + \widetilde{V}_0\cdot\nabla_X H_1 = \widetilde{W}_1|_{Z=1}.$$

Let us now use the fact that $\partial_Z\widetilde{V}_0 = 0$, $\partial_Z\widetilde{W}_0 = -H_0\mathrm{div}_X\widetilde{V}_0$ and the divergence condition at order 1 is

$$\frac{\partial_Z\widetilde{W}_1}{H_0} + \mathrm{div}_x\widetilde{V}_1 - \frac{1}{H_0}\partial_Z\widetilde{W}_0\frac{H_1}{H_0} + \frac{\nabla_x H_0}{H_0}\partial_Z\widetilde{V}_1 Z = 0.$$

We get

$$\partial_t H_1 + \widetilde{V}_0\cdot\nabla_X H_1 + H_0\,\mathrm{div}_X\overline{\widetilde{V}_1} + H_1\mathrm{div}_X\widetilde{V}_0 = 0.$$

This ends the proof since we get (75) using (72). Recalling that $\overline{\kappa} = k\kappa$, $\nu = \nu_0 k$ and $\alpha = k\alpha_0$ and writing the system with respect to ξ, we find the system derived in [127].

REMARK. We remark that the previous model is of order 2 in the asymptotics. We also have checked that the Reynolds tensor is given by

$$-2\nu\mathrm{div}_x(hD(u)) - 2\nu\nabla_x(h\mathrm{div}_x u).$$

In dimension 1, we find $-4\nu\partial_x(h\partial_x u)$, where 4μ is the Trouton viscosity and we comment that in the literature, it is sometimes assumed as the stress tensor under the form

$$\mathcal{D} = -2\nu\mathrm{div}_x(hD_x(u))$$

even in dimension 2. The existence proof given in [37] used this form but seems to be not valid with the term $\nu\nabla(h\mathrm{div}_x u)$. A really interesting problem is to try to obtain a similar global existence result with this stress tensor.

REMARK. There also exist other diffusion terms such as

$$\mathcal{D} = -h\Delta_x u$$

or

$$\mathcal{D} = -\Delta_x(hu).$$

Such diffusion terms are only heuristically proposed.

REMARK. We remark that the equation on $\mathcal{H}_k$ is only approximately satisfied since

$$\partial_\tau \mathcal{H}_k + \mathrm{div}_X(\mathcal{H}_k \mathcal{V}_k) = \mathcal{O}(k^2).$$

This is not the case for H_0, which satisfies exactly

$$\partial_\tau H_0 + \mathrm{div}_X(H_0 \widetilde{V}_0) = 0.$$

REMARK. Let us mention here that Coriolis force has been taken into account in [122]. In this note the author shows that the cosine part of the Coriolis force has to be taken into account if we consider the viscous shallow-water equations derived, for instance, in [96] and [127]. Note that to get such asymptotic systems, the Rossby number is assumed to be fixed and not linked to the aspect ratio. It could be interesting to study the aspect ratio dependent Rossby number.

3.10. *Oscillating topography*

One of the main features in geophysics is the presence of *small parameters*. They appear in dimensionless equations, after appropriate time and space rescaling. They emphasize the relative strength of some physical phenomena, as for instance the rotation of the Earth, stratification or wind stress. See also the Handbook written by N. MASMOUDI [129].

This subsection is devoted to the effect of topography on geophysical flows. More precisely, we will study two different kinds of asymptotics. The first one when the topography term is a high-oscillating function included in the PDEs. The second one where the oscillation is encountered in the domain itself, that means with a high oscillating boundary. Note that the studies we will talk about concern periodic oscillating functions. It would be really interesting to extend these works to random stationary ones in the spirit of [71] for instance.

3.10.1. Oscillating topography inside PDEs We consider two models derived from shallow-water theory: the quasi-geostrophic equation and the lake equation. Small scale variations of topography appear in these models through a periodic function of small wavelength ε. The asymptotic limit as ε goes to zero reveals homogenization problems in which the cell and the averaged equations are both nonlinear. In the spirit of article [117], we derive rigorously the limit systems, through the notion of two-scale convergence.

The inviscid shallow-water equations read, in a bounded domain Ω:

$$\begin{cases} \partial_t h + \mathrm{div}\,(hu) = 0, & t > 0, x = (x_1, x_2) \in \Omega, \\ \partial_t u + u \cdot \nabla u + \dfrac{u^\perp}{\mathrm{Ro}} = \dfrac{1}{\mathrm{Fr}^2} \nabla\,(h + h_b)\,, & t > 0, x \in \Omega, \\ u \cdot n|_{\partial\Omega} = 0, u|_{t=0} = u_0. \end{cases}$$

The unknowns are the height of water $h = h(t, x)$ and the horizontal velocity field $u = u(t, x)$. The function $h_b = h_b(x)$ describes the bottom topography. The positive parameters Ro and Fr are the Rossby and Froude numbers, they penalize the Coriolis force $u^\perp$ and the pressure term $\nabla(h + h_b)$, respectively. We refer to [145] for all details and possible extensions.

In many questions of geophysical concern, at least one of the parameters Ro or Fr is very small, which leads asymptotically to reduced models. A standard one is the "quasi-geostrophic equation", obtained by the scaling $\mathrm{Ro} \approx \mathrm{Fr} \approx \|h_b\| \ll 1$. It reads in its simplest form, (see textbook [125]):

$$\begin{cases} (\partial_t + u \cdot \nabla)(\Delta\psi - \psi + \eta_b) = 0, & t > 0, x \in \Omega, \\ u = \nabla^\perp \psi, \qquad u \cdot n|_{\partial\Omega} = 0, \qquad u|_{t=0} = u_0, \end{cases} \tag{76}$$

where ψ is the stream function associated with the velocity u, and η_b the (rescaled) bottom topography. Another classical one is the "lake equation", corresponding to the asymptotics $\mathrm{Ro} \gg 1$, $\|h_b\| \approx 1$, $Fr \ll 1$. It leads to:

$$\begin{cases} \partial_t u + u \cdot \nabla u + \nabla p = 0, & t > 0, x \in \Omega, \\ \operatorname{div}(\eta_b\, u) = 0, & t > 0, x \in \Omega, \\ u \cdot n|_{\partial\Omega} = 0, \qquad u|_{t=0} = u_0, \end{cases} \tag{77}$$

cf. [110].

Note that for non-varying bottom $\eta_b = 1$, (77) become the incompressible Euler equations.

The main objective is to describe mathematically the impact of variations of relief on geophysical models. We consider here the effect of fast variations of the topography on systems (76) and (77). We assume in both cases that

$$\eta_b(x) = \eta\,(x, x/\varepsilon)\,, \qquad \eta := \eta(x, y) \in \mathcal{C}^0\left(\overline{\Omega} \times \mathbf{T}^2\right), \quad 0 < \varepsilon \leq 1,$$

where $\mathbf{T}^2 := (\mathbf{R}/\mathbf{Z})^2$ models periodic oscillations at the bottom. We also suppose that η is bounded from above and below by positive constants. To unify notations, we denote η_ε instead of η_b. We will limit ourselves to the weakly nonlinear regime for which $\psi = O(\varepsilon)$, $u = O(\varepsilon)$. Through the change of variables $\psi = \varepsilon\Psi^\varepsilon$, $u = \varepsilon v^\varepsilon$, the previous systems become

$$\begin{cases} \left(\partial_t + \varepsilon v^\varepsilon \cdot \nabla\right)\left(\Delta\Psi^\varepsilon - \Psi^\varepsilon\right) + v^\varepsilon \cdot \nabla\eta_\varepsilon = 0, & t > 0, x \in \Omega, \\ v^\varepsilon = \nabla^\perp \Psi^\varepsilon, \quad v^\varepsilon \cdot n|_{\partial\Omega} = 0, \quad v^\varepsilon|_{t=0} = v_0^\varepsilon, \end{cases} \tag{78}$$

respectively

$$\begin{cases} \partial_t v^\varepsilon + \varepsilon v^\varepsilon \cdot \nabla v^\varepsilon + \nabla p^\varepsilon = 0, \quad t > 0, x \in \Omega, \\ \operatorname{div}(\eta_\varepsilon v^\varepsilon) = 0, \quad t > 0, x \in \Omega, \\ v^\varepsilon \cdot n|_{\partial\Omega} = 0, \quad v^\varepsilon|_{t=0} = v_0^\varepsilon. \end{cases} \tag{79}$$

We wish to understand the influence of η_ε at a large scale, as its typical small wavelength ε goes to zero. This yields homogenization problems, that can be tackled at a formal level by double-scale expansions:

$$\Psi^\varepsilon \sim \Psi^0(t,x) + \varepsilon\Psi^1(t,x,x/\varepsilon) + \varepsilon^2\Psi^2(t,x,x/\varepsilon) + \cdots$$
$$v^\varepsilon \sim v^0(t,x,x/\varepsilon) + \varepsilon v^1(t,x,x/\varepsilon) + \cdots$$

In the quasi-geostrophic model, this expansion easily yields:

$$\begin{cases} \left(\partial_t + v^0 \cdot \nabla_y\right)\Delta_y \Psi^1 + v^0 \cdot \nabla_y \eta = 0, \\ \partial_t\left(\Delta_x \Psi^0 - \Psi^0\right) + \overline{v^0 \cdot \nabla_x \eta} + \overline{\nabla_x^\perp \Psi^1 \cdot \nabla_y \eta} = 0, \\ v^0 = \nabla_x \Psi^0 + \nabla_y \Psi^1, \quad \overline{v}^0 \cdot n|_{\partial\Omega} = 0, \quad \overline{v}^0|_{t=0} = v_0, \end{cases} \tag{80}$$

with $\overline{f} := \int_{\mathbf{T}^2} f(\cdot, y)\mathrm{d}y$. These cell and averaged equations form a coupled nonlinear system. The linearized version of this system has been studied from a physical viewpoint in [157].

In the lake model, the Ansatz leads to

$$\begin{cases} \partial_t v^0 + v^0 \cdot \nabla_y v^0 + \nabla_x p^0 + \nabla_y p^1 = 0, \\ \operatorname{div}_y(\eta v^0) = 0, \quad \operatorname{div}_x(\overline{\eta v^0}) = 0, \\ \overline{v}^0 \cdot n|_{\partial\Omega} = 0, \quad \overline{v}^0|_{t=0} = v_0. \end{cases} \tag{81}$$

In [45], the convergence of (78) toward (80) and (79) toward (81), respectively, are proved. The convergence result involves the notion of two-scale convergence introduced by G. NGUETSENG [137], and developed by G. ALLAIRE [4]. We give here an enlarged definition, that accounts for variable time and various Lebesgue spaces. Classical properties of two-scale convergence extend to this framework (see [4]).

DEFINITION 3.17. *Let Ω a bounded domain of $\boldsymbol{R}^2$. Let $v^\varepsilon = v^\varepsilon(t,x)$ be a sequence of functions in $L^p(\boldsymbol{R}_+; L^q(\Omega))$, with $1 < p, q \le \infty$, $(p,q) \ne (\infty,\infty)$ (respectively in $L^\infty(\boldsymbol{R}_+ \times \Omega)$). Let $v \in L^p\left(\boldsymbol{R}_+; L^q\left(\Omega \times \boldsymbol{T}^2\right)\right)$ (respectively in $L^\infty\left(\boldsymbol{R}^+ \times \Omega \times \boldsymbol{T}^2\right)$). We say that v^ε two-scale converges to v if, for all $w \in L^{p'}\left(\boldsymbol{R}_+; C^0(\overline{\Omega} \times \boldsymbol{T}^2)\right)$,*

$$\lim_{\varepsilon\to 0} \int_{\boldsymbol{R}_+}\int_\Omega v^\varepsilon(t,x) w(t,x,x/\varepsilon)\,\mathrm{d}x\mathrm{d}t$$
$$= \int_{\boldsymbol{R}_+}\int_{\Omega\times\boldsymbol{T}^2} v(t,x,y) w(t,x,y)\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t.$$

For functions independent of t, $p = \infty$ and $q = 2$, one recovers the two-scale convergence on Ω. We state the main result:

THEOREM 3.18. *Let $q_0 > 2$, v_0^ε bounded in $L^{q_0}(\Omega)$, satisfying* $\operatorname{div} v_0^\varepsilon = 0$, $v_0^\varepsilon \cdot n|_{\partial\Omega} = 0$, $\varepsilon \operatorname{curl} v_0^\varepsilon$ *bounded in $L^\infty(\Omega)$. Assume, moreover, that v_0^ε two-scale converges to v_0 on Ω, with*

$$\|v_0^\varepsilon\|_{L^2(\Omega)} \xrightarrow[\varepsilon\to 0]{} \|v_0\|_{L^2(\Omega\times \boldsymbol{T}^2)}.$$

Then, to extract a subsequence, the solution v^ε of (78), resp. (79) two-scale converges to a solution v^0 of (80), resp. (81).

Idea of the proof. Note that the proof relies on the characterization of the defect measure of the velocity field. It follows the strategy used in [117], devoted to Euler equations in a porous medium. The quasi-geostrophic equation is reformulated in an Eulerian framework. We introduce the two-scale defect measure α, resp. β such that $|v^\varepsilon|^2$ two-scale converges to $|v^0|^2 + \alpha$ and $v^\varepsilon \otimes v^\varepsilon$ two-scale converges to $v^0 \otimes v^0 + \beta$. Denoting $\gamma = \overline{\eta\alpha}$, we prove that

$$\partial_t \gamma(t,x) \le C(x) p \gamma(t,x)^{1-1/p}$$

for all $p < +\infty$ and we conclude, assuming $\gamma|_{t=0}(x) = 0$ for almost every x, that $\gamma = 0$, as in YUDOVICH's procedure. Thus $\alpha = 0$ and therefore using the fact that $|\beta| \le C\alpha$, we get $\beta = 0$.

To get the fundamental inequality on γ, we need to use a standard procedure: limit/scalar product − scalar product/limit. More precisely, we pass to the limit on the momentum equation

$$\eta_\varepsilon \partial_t v^\varepsilon + \varepsilon \eta_\varepsilon v^\varepsilon \cdot \nabla v^\varepsilon + \eta_\varepsilon \nabla p^\varepsilon = 0$$

and then take the scalar product with v^0. Then we rewrite the momentum equation as

$$\partial_t v^\varepsilon + \varepsilon \operatorname{curl} v^\varepsilon (v^\varepsilon)^\perp + \nabla(p^\varepsilon + \varepsilon |v^\varepsilon|^2/2) = 0,$$

take the scalar product with v^ε and then pass to the limit. We subtract the results and get

$$\partial_t \overline{\eta\alpha} + 2\overline{\eta\beta : \nabla_y v^0} = 0,$$

where $\cdot : \cdot$ denotes the scalar product for matrices. It suffices now to use the properties of $\nabla_y v$ deduced from the properties of $\varepsilon\omega^\varepsilon$. Note that $\omega^\varepsilon = \operatorname{curl} v^\varepsilon / \eta^\varepsilon$ is transported along the flow. More precisely

$$\partial_t \omega^\varepsilon + \varepsilon v^\varepsilon \cdot \nabla \omega^\varepsilon = 0.$$

To be able to pass to the limit, we emphasize the two-scale definition and some compactness properties. We refer the interested reader to [45] for details and also to [117] for original studies on the homogenization problem.
Bounds on v^ε, ω^ε. We use the following Helmholtz–Hodge decomposition of the velocity

$$v^\varepsilon = Pv^\varepsilon + \nabla\phi^\varepsilon, \qquad \int_\Omega \phi^\varepsilon = 0,$$

where the operator P denotes the LERAY projector on divergence free space. Together with $\mathrm{div}(\eta^\varepsilon v^\varepsilon) = 0$, this decomposition yields

$$\mathrm{div}(\eta^\varepsilon \nabla\phi^\varepsilon) = -\mathrm{div}(\eta^\varepsilon P v^\varepsilon), \qquad \int_\Omega \phi^\varepsilon = 0, \qquad \partial_n \phi^\varepsilon|_{\partial\Omega} = 0.$$

We rewrite the equation

$$\partial_t v^\varepsilon + \varepsilon \mathrm{curl} v^\varepsilon (v^\varepsilon)^\perp + \nabla(p^\varepsilon + \varepsilon|v^\varepsilon|^2/2) = 0$$

and project it in divergence free space to get

$$\partial_t P v^\varepsilon + P\left(\varepsilon \mathrm{curl} v^\varepsilon (v^\varepsilon)^\perp\right) = 0.$$

We now use the following *Meyers elliptic theory.* Let Ω be a bounded domain of $\mathbf{R}^2$, $a \in L^\infty(\Omega)$ and $f \in L^{q_0}(\Omega)$ for some $q_0 > 2$, $\int_{\partial\Omega} f \cdot n = 0$. Let $\phi \in H^1(\Omega)$ be the solution of

$$\mathrm{div}(a\nabla\phi) = \mathrm{div}\, f \quad \text{in } \Omega, \qquad \int_\Omega \phi = 0, \qquad \partial_n\phi|_{\partial\Omega} = 0.$$

Then there exists $2 < q_m = q_m(\|a\|_{L^\infty}, \Omega) \leq q_0$, such that for all $2 \leq q < q_m$, $\phi \in W^{1,q}(\Omega)$ such that

$$\|\phi\|_{W^{1,q}} \leq C\|f\|_{L^q}, \qquad C = C(q, \|a\|_{L^\infty}, \Omega).$$

This gives the following control

$$\|v^\varepsilon\|_{L^p} \leq C\|Pv^\varepsilon\|_{L^p}$$

and therefore the $L^\infty(0, T; L^p(\Omega))$ uniform bound on v^ε, using the equation obtained on Pv^ε and bound on $\varepsilon\omega^\varepsilon$. Using bounds on $\varepsilon\omega^\varepsilon$ and v^ε and the momentum equation, we get the necessary bounds to pass to the limit in the quantities. The reader interested in this procedure is referred to [45].
Asymptotics related to shallow-water equations. Let us now give one formal asymptotics that has been obtained in [123] and which is actually, with others, under consideration for a mathematical justification in collaboration with R. KLEIN in [49]. We also

refer readers interested by multiscale analysis and compressible flows to [150]. We consider the non-dimensional weakly nonlinear inviscid shallow-water equations in a bounded domain Ω:

$$\begin{cases} \partial_t h + \varepsilon \mathrm{div}\,(hu) = 0, \quad t > 0, \quad x = (x_1, x_2) \in \Omega, \\ \partial_t u + \varepsilon u \cdot \nabla u + \dfrac{1}{\varepsilon^2} \nabla\,(h + \eta_b) = 0, \quad t > 0, \quad x \in \Omega, \\ u \cdot n|_{\partial\Omega} = 0, \quad u|_{t=0} = u_0. \end{cases}$$

Note that the Froude number is chosen to be exactly equal to ε. Then using a two scale-expansion

$$h \sim h^0(t, x, x/\varepsilon) + \varepsilon h^1(t, x, x/\varepsilon) + \varepsilon^2 h^2(t, x, x/\varepsilon) + \cdots$$
$$u \sim u^0(t, x, x/\varepsilon) + \varepsilon u^1(t, x, x/\varepsilon) + \cdots$$

we get the following system

$$\begin{cases} \mathrm{div}_y(\eta_b u^0) = 0, \\ \partial_t u^0 + u^0 \cdot \nabla_y u^0 + \nabla_x h^0 + \nabla_x h^2 = 0, \\ \mathrm{div}_x(\eta_b u^0) + \mathrm{div}_y(\eta_b u^1) = 0, \\ \nabla_y h^2 = 0. \end{cases}$$

Note that this is exactly the same system as the one obtained from the lake equations in the previous section.

3.10.2. High-oscillations through the domain In the interior of the domain, small parameters lead to some reduced behavior. For instance, in highly rotating fluids, the velocity is invariant along the rotation axis: this is the so-called Taylor-Proudman theorem. In domains with boundaries, this reduced behavior is often incompatible with the boundary conditions (typically the no-slip condition at a rigid surface). This leads to *boundary layers*, small zones near the border where the fluid velocity changes rapidly, so as to satisfy the boundary conditions.

Geophysical boundary layers have been the matter of extensive work, and we refer to the classical monograph by J. PEDLOSKY [145] for a physical introduction. On the mathematical side, it has also been the subject of many studies, in the context of singular perturbation problems (see for instance [66] and references therein).

Up to the first mathematical study by D. GÉRARD-VARET, see [94], most of the theoretical papers focused on *smooth boundaries*, i.e. defined by an expression

$$x_\perp = \chi\,(x_t)\,, \quad \mathrm{x} = (x_t, x_\perp), \tag{82}$$

where $x_\perp \in \mathbf{R}$ is the "normal" variable, $x_t \in \mathbf{R}^d$ is the "tangential" variable, and $\chi : \mathbf{R} \mapsto \mathbf{R}^d$ is smooth. Briefly speaking, the problem is then to derive asymptotics of the type:

$$v^{\varepsilon}(t,\mathrm{x}) \sim v^{\mathrm{int}}(t,\mathrm{x}) + v^{bl}\left(t, x_t, \frac{x_\perp - \chi\,(x_t)}{\lambda(\varepsilon)}\right), \tag{83}$$

where

- $\varepsilon \in \mathbf{R}^n$, $n \geq 1$ describes the small parameters of the system,
- v^{ε} is the solution of the system (mostly of Navier–Stokes type),
- v^{int} models the evolution of v^{ε} far from the boundary,
- v^{bl} is the boundary layer corrector, $\lambda(\varepsilon) \ll 1$ being the boundary layer size.

However, in many geophysical systems, the boundary representation (82) (and the corresponding approximation (83)) can be too crude since realistic bottoms or coasts vary over a wide range of scales, the smallest being not resolved in numerical computations. There is need for development of simplified models that describe implicitly the effect on the large scale ocean circulation. These small-scale variations are not modeled in (82), (83), in which a slow dependence on x_t is assumed. Therefore, the traditional "smooth" boundary layer analysis can miss interesting physical phenomena.

The aim of this section is to discuss such roughness-induced effects and is mainly inspired by [46]. More precisely, we are concerned with two different geophysical problems in idealized configurations.

The first one is the evolution of *rotating fluids in rough domains*. Recall that the velocity u and the pressure p of the flow corresponding to an incompressible fluid with constant density in a rotating frame, is governed by the following Navier–Stokes-Coriolis equations:

$$\partial_t u + u \cdot \nabla u + \frac{\mathbf{e} \times u}{\mathrm{Ro}} + \frac{\nabla p}{\mathrm{Ro}} - \frac{E}{\mathrm{Ro}} \Delta u = 0, \tag{84}$$

$$\operatorname{div} u = 0, \tag{85}$$

where $\mathbf{e} = (0,0,1)^t$ is the rotation axis, Ro is the Rossby number, and E is the Ekman number. Note that the latitude is assumed to be constant. Let us assume that

$$\mathrm{Ro} = \varepsilon, \quad E = \varepsilon^2, \quad \varepsilon \ll 1.$$

This is a classical scaling, for instance relevant to the Earth's liquid core, for which $\mathrm{Ro} \sim 10^{-7}$ and $E \sim 10^{-15}$. This kind of scaling occurs in oceanography, see for instance [145]. It leads to equations

$$\partial_t u + u \cdot \nabla u + \frac{\mathbf{e} \times u}{\varepsilon} + \frac{\nabla p}{\varepsilon} - \varepsilon \Delta u = 0, \tag{86}$$

$$\operatorname{div} u = 0, \tag{87}$$

in a 3-dimensional space domain Ω^{ε}, together with the Dirichlet condition

$$u^{\varepsilon}|_{\partial\Omega^{\varepsilon}} = 0.$$

System (86), (87) has been widely studied in the framework of smooth boundaries, notably when $\Omega^\varepsilon = \mathbf{R}^2 \times (0, 1)$ (two horizontal plates). In this case, the basic idea (see for instance [145]) is that for small enough ε, solutions u^ε of (86), (87) can be approximated by

$$u^\varepsilon_{\mathrm{app}}(t, x, y, z) = u(t, x, y) + u_-\left(t, x, y, \frac{z}{\varepsilon}\right) + u_+\left(t, x, y, \frac{1-z}{\varepsilon}\right), \tag{88}$$

where

- u is a two-dimensional interior term (i.e. $u_3 = 0$)
- u_- and u_+ are boundary layer terms called *Ekman layers*

Moreover, this idea can be made mathematically rigorous under suitable stability assumptions. We wish here to present (86), (87) in the framework of rough boundaries, with a roughness of characteristic size ε (both horizontally and vertically). In oceanography, accurate numerical models require that we are able to parametrize the effects of unresolved (sub-grid-scale) bathymetric roughness which have a large range of variability. Thus it is important to develop a physical understanding of oceanic flow along an irregular, sloping topography in order to provide new appropriate large scale models. In both cases, it is also important from a physical point of view to understand how the roughness affects the layer and its stability.

The second problem we are interested in is the *effect of rough shores on the oceanic circulation*, see [47]. We use the classical quasi-geostrophic model, see [40] for some derivation studies, which reads

$$\begin{cases} \left(\partial_t + u_1\partial_x + u_2\partial_y\right)(\Delta\Psi + \beta y + \eta_B) + r\Delta\Psi = \beta \operatorname{curl}\tau + \mathrm{Re}^{-1}\Delta^2\Psi, \\ u = (u_1, u_2)^t = \nabla^\perp\Psi = (-\partial_y\Psi, \partial_x\Psi), \\ \Psi|_{\partial\Omega} = \dfrac{\partial\Psi}{\partial n}|_{\partial\Omega} = 0, \\ \Psi|_{t=0} = 0. \end{cases} \tag{89}$$

Recall that, in these equations

- $\Psi = \Psi(t, \mathrm{x}) \in \mathbf{R}$ is a stream function, associated with the velocity field $u = (u_1(t, \mathrm{x}), u_2(t, \mathrm{x}))^t$. The time variable t lies in $\mathbf{R}^+$, and the space variable $\mathrm{x} = (x, y)$ lies in a two-dimensional domain Ω to be defined later on.
- $\mathrm{d}/\mathrm{d}t = \partial_t + u_1\partial_x + u_2\partial_y$ is the transport operator of the two-dimensional flow.
- $\Delta\Psi$ is the vorticity and $r\Delta\Psi$, $r > 0$, is the Ekman pumping term.
- βy is the second term in the development of the Coriolis force.
- η_B is a bottom topography term.
- $\beta \operatorname{curl}\tau$ is the vorticity created by the wind, where τ is a given stress tensor.
- $\mathrm{Re}^{-1}\Delta^2\Psi$ is the usual viscosity term.

Despite its simplicity, this model catches some of the main features of oceanic circulation. We refer the reader to [145] for an extensive physical discussion. Among the related papers, let us mention the important work of B. DESJARDINS and E. GRENIER [78]. These authors have performed a complete boundary layer analysis, under various asymptotics, in domains

$$\Omega = \{\chi_w(y) \leq x \leq \chi_e(y),\, y_{\min} \leq y \leq y_{\max}\}.$$

They have notably derived the so-called Munk layers, responsible for the western intensification of boundary currents.

There again, we wish to understand roughness-induced effects on (89). We will consider a roughness of characteristic size $\varepsilon = \beta^{-1/3}$, which is the size of the Munk layer. In our view, it is an important step to progress in the understanding of the so-called "Gulf stream separation". This expression refers to the abrupt separation of the Gulf Stream from the North American coastline at Cape Hatteras. This phenomenon, which has been observed for many years, seems to have very little variability: the current leaves the continent on a straight path without any visible deflection at the separation point. However, it is still poorly understood from a physical point of view, see [22] for more physical insight (in particular for a numerical study of the influence of boundary conditions on the separation point). Therefore, it is important to try to obtain new models which reflect the Gulf stream separation, taking into account for instance the effect of rough boundaries.

In this part, we will continue the asymptotic analysis of systems (86)–(87) and (89), at the limit of small ε. We will show how to derive new physical models, that incorporate small-scale effects of the roughness. We will then discuss the qualitative properties of such models. We only develop the formal and qualitative aspects of the analysis. In particular, we do not deal with the related existence, uniqueness and stability problems and only mention for the reader a convergence result at the end of this section. The reader interested in the mathematical justification of the results by D. GÉRARD-VARET is referred to [94] and to ours [45], where (86), (87) and (89) are studied exhaustively from the mathematical viewpoint.

The next section is structured as follows: at first, we describe modeling aspects, modeling of the rough domains, then modeling of the approximate solutions. This leads to reduced models for the interior and boundary layer flows, whose qualitative properties are briefly analyzed. We end up with a convergence result theorem.

I. *Modeling of the rough problem*

The rough domains. We first introduce the domains Ω^ε where equations (86), (87) and system (89) hold. These domains read

$$\Omega^\varepsilon = \Omega^\varepsilon_- \cup \Sigma_- \cup \Omega \cup \Sigma_+ \cup \Omega^\varepsilon_+, \tag{90}$$

where

- Ω models the interior part of the domain.
- $\Omega^\varepsilon_\pm$ are the rough parts. They are obtained by periodic translation of canonical cells of roughness, of characteristic size ε.
- $\Sigma_\pm$ are the interfaces.

Let us now be more specific:

(i) *The rotating fluids system.* In this case

$$\Omega := \mathbf{T}^2 \times (0,1), \quad \Sigma_- := \mathbf{T}^2 \times \{0\}, \qquad \Sigma_+ := \mathbf{T}^2 \times \{1\}.$$

Recall that $\mathbf{T}^n$ is an n-dimensional torus, corresponding to the periodic boundary conditions. Let then $\gamma_\pm = \gamma_\pm(X, Y)$ be two positive Lipschitz functions, 1-periodic in X and Y. We define the rough parts by:

$$\Omega^\varepsilon_- := \left\{(x,y,z), 0 > \frac{z}{\varepsilon} > -\gamma_-\left(\frac{x}{\varepsilon}, \frac{y}{\varepsilon}\right)\right\}, \tag{91}$$

$$\Omega^\varepsilon_+ := \left\{(x,y,z), 0 < \frac{z-1}{\varepsilon} < \gamma_+\left(\frac{x}{\varepsilon}, \frac{y}{\varepsilon}\right)\right\}. \tag{92}$$

Thus, $\Omega^\varepsilon_\pm$ consist of a large number of periodically distributed humps, of characteristic length and amplitude ε.
(ii) *The quasi-geostrophic System* (89). In this case

$$\begin{aligned}&\Omega := \{\chi_-(y) \le x \le \chi_+(y), y \in \mathbf{T}\},\\ &\Sigma_- := \{(\chi_-(y), y), y \in \mathbf{T}\}, \qquad \Sigma_+ := \{(\chi_+(y), y), y \in \mathbf{T}\},\end{aligned}$$

where $\chi_\pm$ are smooth functions. Let then $\gamma_\pm = \gamma_\pm(Y)$ be two smooth, positive and 1-periodic functions. We define,

$$\Omega^\varepsilon_- := \left\{(x,y), 0 > \frac{x-\chi_-(y)}{\varepsilon} > -\gamma_-\left(\frac{y}{\varepsilon}\right)\right\}, \tag{93}$$

$$\Omega^\varepsilon_+ = \left\{(x,y), 0 < \frac{x-\chi_+(y)}{\varepsilon} < \gamma_+\left(\frac{y}{\varepsilon}\right)\right\}. \tag{94}$$

REMARK. By including a few more technicalities, one could consider more general roughness: for instance, one can add a slow dependence, with boundaries of the type $\varepsilon^{-1}z = \gamma(x, y, \varepsilon^{-1}x, \varepsilon^{-1}y)$ (or $\varepsilon^{-1}x = \gamma(y, \varepsilon^{-1}y)$).

The boundary layer domains. Besides these rough domains, we need additional boundary layer domains.They are deduced from Ω^ε by focusing on the rough parts.
More precisely, we introduce:
(i) *The case of rotating fluids:*

$$\omega_- := \left\{\mathrm{X} = (X,Y,Z), X, Y \in \mathbf{T}^2, -\gamma_-(X,Y) < Z < 0\right\},$$

and

$$\omega_+ := \left\{\mathrm{X} = (X,Y,Z), X, Y \in \mathbf{T}^2, \gamma_+(X,Y) > Z > 0\right\}.$$

(ii) *The case of the quasi-geostrophic model:*

$$\omega_- := \{\mathrm{X} = (X, Y), Y \in \mathbf{T}, -\gamma_-(Y) < X < 0\},$$

and

$$\omega_+ := \{\mathrm{X} = (X, Y), Y \in \mathbf{T}, \gamma_+(Y) > X > 0\}.$$

We also denote the boundaries of $\omega_\pm$:

$$\Gamma_\pm = \{Z = \pm\gamma_\pm(X, Y)\}.$$

The Ansatz. Once the domains are defined, one can start the asymptotic analysis of the system. The first step is to construct approximate solutions which fit the rough geometry. Expressions like (83) are not adapted anymore, as we now expect a fast dependence on the tangential variables (due to the small horizontal scale of the roughness). This leads to replacing (83) by an expression of the following type:

$$v^\varepsilon(t, \mathrm{x}) = v^{\mathrm{int}}(t, \mathrm{x}) + v^{bl}\left(t, x_t, \frac{x_t}{\lambda_t(\varepsilon)}, \frac{x_\perp - \chi(x_t)}{\lambda_\perp(\varepsilon)}\right),$$

where $\lambda_t(\varepsilon)$ and $\lambda_\perp(\varepsilon)$ are the tangential and normal length scales of the roughness.

Profiles. In the case of System (86), (87), we look for approximate solutions of the type:

$$u^\varepsilon_{\mathrm{app}}(t, x, y, z) = \sum \varepsilon^i \left(u^i(t, x, y, z) + u^i_-\left(t, x, y, \frac{x}{\varepsilon}, \frac{y}{\varepsilon}, \frac{z}{\varepsilon}\right)\right.$$
$$\left. + u^i_+\left(t, x, y, \frac{x}{\varepsilon}, \frac{y}{\varepsilon}, \frac{z-1}{\varepsilon}\right)\right),$$

$$p^\varepsilon_{\mathrm{app}}(t, x, y, z) = \sum \varepsilon^i \left(p^i(t, x, y, z) + p^i_-\left(t, x, y, \frac{x}{\varepsilon}, \frac{y}{\varepsilon}, \frac{z}{\varepsilon}\right)\right.$$
$$\left. + p^i_+\left(t, x, y, \frac{x}{\varepsilon}, \frac{y}{\varepsilon}, \frac{z-1}{\varepsilon}\right)\right).$$

In the case of System (89), it yields:

$$\Psi^\varepsilon_{\mathrm{app}}(t, x, y) = \sum \varepsilon^i \left(\Psi^i(t, x, y) + \Psi^i_-\left(t, y, \frac{x - \chi_-(y)}{\varepsilon}, \frac{y}{\varepsilon}\right)\right.$$
$$\left. + \Psi^i_+\left(t, y, \frac{x - \chi_+(y)}{\varepsilon}, \frac{y}{\varepsilon}\right)\right).$$

In these expressions:

- the u^i, p^i (resp. Ψ^i) are *interior profiles*, defined on $\mathbf{R}^+ \times \Omega^\varepsilon$

- the $u^i_\pm = u^i_\pm(t,x,y,X,Y,Z)$, $p^i_\pm = u^i_\pm(t,x,y,X,Y,Z)$ (resp. $\Psi^i_\pm = \Psi^i_\pm(t,y,X,Y)$) are *boundary layer profiles*, defined for $t > 0$ and X in $\omega^\pm$

Note the difference between the Ansatz case (88) and the previous ones. The fast variables x/ε and y/ε are chosen in order to take into account the variability of the boundary with respect to the space variable x/ε, y/ε.

Boundary conditions. The interior and layer profiles satisfy the boundary conditions. We expect the boundary layer terms not to play any role far from the boundaries, which leads to the conditions

$$u^i_\pm \xrightarrow[Z\to+\infty]{} 0, \quad \Psi^i_\pm \xrightarrow[Z\to+\infty]{} 0. \tag{95}$$

We then want our approximate solutions to satisfy the boundary conditions at $\partial\Omega^\varepsilon$, i.e.

$$\begin{aligned} &u^\varepsilon_{\text{app}}|_{\partial\Omega^\varepsilon} = 0, \\ &\Psi^\varepsilon_{\text{app}}|_{\partial\Omega^\varepsilon} = 0, \qquad \frac{\partial \Psi^\varepsilon_{\text{app}}}{\partial n}|_{\partial\Omega^\varepsilon} = 0. \end{aligned}$$

Using the expressions of the approximations, this reads

$$\sum \varepsilon^i \left[u^i\left(t,x,y,\delta_\pm \pm \varepsilon\gamma_\pm\left(\frac{x}{\varepsilon},\frac{y}{\varepsilon}\right)\right) + u^i_\pm\left(t,x,y,\pm\gamma_\pm\left(\frac{x}{\varepsilon},\frac{y}{\varepsilon}\right)\right)\right] = 0,$$

$$\sum \varepsilon^i \left[\Psi^i\left(t,\chi_\pm(y) \pm \varepsilon\gamma_\pm\left(\frac{y}{\varepsilon}\right),y\right) + \Psi^i_\pm\left(t,y,\pm\gamma_\pm\left(\frac{y}{\varepsilon}\right),\frac{y}{\varepsilon}\right)\right] = 0$$

and

$$\begin{aligned} &\sum \varepsilon^i n^\varepsilon_\pm\left(y,\frac{y}{\varepsilon}\right)\cdot\left[(\nabla\Psi^i)\left(t,\chi_\pm(y)\pm\varepsilon\gamma_\pm\left(\frac{y}{\varepsilon}\right),y\right)\right. \\ &\quad \left. + \left(\frac{1}{\varepsilon}\nabla^y_\pm\Psi^i_\pm + \left(\begin{pmatrix}0\\ \partial_y\end{pmatrix}\right)\Psi^i_\pm\right)\left(t,y,\pm\gamma_\pm\left(\frac{y}{\varepsilon}\right),\frac{y}{\varepsilon}\right)\right] = 0 \end{aligned}$$

with $\delta_+ = 1$, $\delta_- = 0$,

$$n^\varepsilon_\pm(y,Y) := (-1, \partial_y\chi_\pm(y) \pm \partial_Y\gamma_\pm(Y))^t, \qquad \nabla^y_\pm := \left(\partial_X, \partial_Y - \chi'_\pm\partial_X\right)^t.$$

Such conditions are fulfilled as soon as

$$\sum \varepsilon^i \left[u^i\,(t,x,y,\delta_\pm \pm \varepsilon\gamma_\pm\,(X,Y)) + u^i_\pm\,(t,x,y,\pm\gamma_\pm\,(X,Y))\right] = 0,$$

$$\sum \varepsilon^i \left[\Psi^i\,(t,\chi_\pm(y) \pm \varepsilon\gamma_\pm\,(Y)\,,y) + \Psi^i_\pm\,(t,y,\pm\gamma_\pm\,(Y)\,,Y)\right] = 0$$

and

$$\sum \varepsilon^i n_\pm(y,Y)\cdot\Big[(\nabla\Psi^i)\,(t,\chi_\pm(y)\pm\varepsilon\gamma_\pm\,(Y)\,,y)$$
$$+\left(\frac{1}{\varepsilon}\nabla^y_\pm\Psi^i_\pm+\left(\begin{pmatrix}0\\ \partial_y\end{pmatrix}\right)\Psi^i_\pm\right)(t,y,\pm\gamma_\pm\,(Y)\,,Y)\Big]=0.$$

We can express these last equalities as Taylor expansions in powers of ε. Setting each coefficient equal to zero leads to the desired boundary conditions. They read, for $i=0$:

$$u^0_\pm(t,x,y,X,Y,\pm\gamma_\pm(X,Y))=-u^0(t,x,y,0), \tag{96}$$

$$\Psi^0_\pm(t,y,\pm\gamma(Y),Y)=-\Psi^0(t,\chi_\pm(y),y) \tag{97}$$

and

$$n^\pm(y,Y)\Psi^0_\pm(t,y,\pm\gamma(Y),Y)=0. \tag{98}$$

The next terms satisfy

$$u^i_\pm(t,x,y,X,Y,\pm\gamma_\pm(X,Y))=-u^i(t,x,y,\delta_\pm)+f^i_\pm(t,x,y), \tag{99}$$

$$\Psi^i_\pm(t,y,\pm\gamma(Y),Y)=-\Psi^i(t,\chi_\pm(y),y)+g^i_\pm(t,y) \tag{100}$$

and

$$n^\pm(y,Y)\Psi^0_\pm(t,y,\pm\gamma(Y),Y)=h^i_\pm(t,y), \tag{101}$$

where $f^i_\pm$, $g^i_\pm$ and $h^i_\pm$ depend on the u^k, $u^k_\pm$, respectively, Ψ^k, $\Psi^k_\pm$, for $k\le i-1$.

II. *Formal Derivation.* In this section, we will derive the leading terms of the approximate solutions. This derivation will yield new physical models, including the impact of small-scale irregularities. As mentioned in the introduction, we restrict ourselves to the formal part, all mathematical aspects can be found in [94,45].

(i) *Rotating Fluids.* We plug the Ansatz into Equations (86), (87). The resulting equations are ordered according to powers of ε, and the coefficients of the different powers of ε are set equal to zero. It leads to a collection of equations on the profiles.

Geostrophic balance: At order ε^{-2} in the boundary layers, we get

$$\nabla_X\tilde{p}^0_\pm=0\quad\text{in }\tilde{\omega}_\pm, \tag{102}$$

which leads to $p^0_\pm=0$. The pressure does not change in the boundary layer, which is classical (see [145]). At order ε^{-1} in the interior, Equation (86) yields the geostrophic balance

$$\mathbf{e} \times u^0 + \nabla p^0 = 0. \tag{103}$$

At order ε^0 in the interior, we get from (87)

$$\operatorname{div} u^0 = 0. \tag{104}$$

Using the last line of (103), we have

$$\partial_z p^0 = 0 \tag{105}$$

and taking the curl of (103) together with (105), we obtain

$$\partial_z u^0 = 0. \tag{106}$$

Thus, p^0 and u^0 are independent of z. This is the well-known Taylor-Proudman theorem. At order ε^{-1} in the boundary layers, we get from (87)

$$\operatorname{div}_{\mathrm{X}} u^0_{\pm} = 0 \quad \text{in } \tilde{\omega}_{\pm}. \tag{107}$$

Use of Green-Ostrogradsky formula and boundary condition (96) leads to

$$\begin{aligned} 0 &= \int_{\omega_\pm} \operatorname{div}_{\mathrm{X}} u^0_{\pm} \\ &= \int_{\Gamma_\pm} -u^0_{\pm} \cdot n_{\pm} \\ &= \int_{\Gamma_\pm} -u^0(t,x,y) \cdot n_{\pm} \\ &= -\sum_{i=1}^{3} u^0_i(t,x,y) \int_{\Gamma_\pm} e_i \cdot n_{\pm} \\ &= u^0_3(t,x,y), \end{aligned} \tag{108}$$

where the last equality comes from Green-Ostrogradsky formula applied to the constant functions $e_i = (\delta_{i,j})^t_{j=1,\dots,3}$. We thus recover a *two-dimensional velocity field*: $u^0 = (v(t,x,y),0)^t$.

Boundary layer equations. We focus on the lower boundary layer, the upper one leading to similar equations. At order ε^{-1} in the boundary layer, we get from (86)

$$\mathbf{e} \times u^0_- + u^0_- \cdot \nabla_{\mathrm{X}} u^0_- + \nabla_{\mathrm{X}} p^1_- - \Delta_{\mathbf{X}} u^0_- = 0. \tag{109}$$

One can see that Eqs. (96), (107) and (109) involve the variables t, x, y as simple parameters. We can omit them from the notations, and at fixed (t,x,y) set

$$u_- = u^0_-(t,x,y,\cdot), \qquad p_- = p^1_-(t,x,y,\cdot), \qquad \mathbf{v} = v(t,x,y).$$

Boundary layer equations can finally be written

$$\begin{aligned}
&\mathbf{e}\times u_- + \nabla p_- + u_-\cdot\nabla u_- - \Delta u_- = \begin{pmatrix} -\mathbf{v}^\perp \\ 0 \end{pmatrix} \quad \text{in } \tilde{\omega}_-,\\
&\operatorname{div} u_- = 0 \quad \text{in } \tilde{\omega}_-,\\
&u_- = -\begin{pmatrix} \mathbf{v} \\ 0 \end{pmatrix} \quad \text{on } \Gamma_-\\
&u_- \text{ 1-periodic in } (X,Y).
\end{aligned} \tag{110}$$

Interior equations. It remains to identify the system satisfied by $u^0 = (v(t,x,y),0)^t$. At order ε^0, we have from (86)

$$\partial_t u^0 + u^0\cdot\nabla u^0 + \mathbf{e}\times u^1 + \nabla p^1 = 0 \tag{111}$$

and at order ε^1 from (87),

$$\operatorname{div} u_1 = 0. \tag{112}$$

As shown in [145], noting $\zeta = \partial_x v_2 - \partial_y v_1$, the curl of v, we get

$$\partial_t\zeta + v\cdot\nabla_{x,y}\zeta = \partial_z u^1_3.$$

Then we integrate for z from 0 to 1, which yields

$$\partial_t\zeta + v\cdot\nabla_{x,y}\zeta = u^1_3(\cdot, z=1) - u^1_3(\cdot, z=0).$$

Computation of the right hand-side: Eq. (87) gives, at order ε^0 in the boundary layer

$$\operatorname{div}_X u^1 + \partial_x u^0_{-,1} + \partial_y u^0_{-,2} = 0.$$

Proceeding as in (108) we obtain

$$\begin{aligned}
-\int_{\omega_-} \partial_x u^0_{-,1} + \partial_y u^0_{-,2} &= \int_{\omega_\pm} \operatorname{div}_X u^1_\pm\\
&= \int_{\Gamma_-} u^1_\pm\cdot n_\pm\\
&= \int_{\Gamma_-} -u^1(t,x,y,0)\cdot n_\pm\\
&= -\sum_{i=1}^3 u^1_i(t,x,y,0)\int_{\Gamma_-} e_i\cdot n_\pm,
\end{aligned} \tag{113}$$

so that

$$u_3^1(\cdot, z = 0) = -\int_{\omega_-} \left(\partial_x u^0_{-,1} + \partial_y u^0_{-,2}\right) \mathrm{dX}.$$

In the same way,

$$u_3^1(\cdot, z = 1) = \int_{\omega_+} \left(\partial_x u^0_{+,1} + \partial_y u^0_{+,2}\right) \mathrm{dX}$$

so that ζ solves

$$\partial_t \zeta + v \cdot \nabla_{x,y} \zeta + \operatorname{curl} \left(\int_{\omega_-} \begin{pmatrix} u^0_{-,2} \\ -u^0_{-,1} \end{pmatrix} \mathrm{dX} + \int_{\omega_+} \begin{pmatrix} u^0_{+,2} \\ -u^0_{+,1} \end{pmatrix} \mathrm{dX} \right) = 0.$$

We define the operator

$$P_-(\mathbf{v}) := \int_{\omega_-} \begin{pmatrix} u_{-,2} \\ -u_{-,1} \end{pmatrix},$$

where u_- is a solution of (110). We define in the same way function P_+ for the upper layer, and set $P := P_- + P_+$. Finally, we get that v *is solution on* $\mathbf{T}^2$ *of*

$$\begin{aligned} &\partial_t \zeta + v \cdot \nabla \zeta + \operatorname{curl} P(v) = 0, \\ &\zeta = \operatorname{curl} v, \qquad v = \nabla^\perp p. \end{aligned} \tag{114}$$

(ii) *The quasi-geostrophic model.* To derive Ψ^0 and $\Psi^0_\pm$, we proceed as above and inject the approximation into (89).

The Sverdrup relation. In the interior, we obtain at the leading order ε^{-3}, the so-called Sverdrup relation:

$$\partial_x \Psi^0 = \operatorname{curl} \tau.$$

We choose Ψ^0 to cancel at Σ_+, so that

$$\Psi^0(t, x, y) = \int_{\chi_+(y)}^x \operatorname{curl} \tau(t, x', y) \mathrm{d}x'.$$

The western layer. We recall that

$$\nabla^y_\pm := \left(\partial_X, \partial_Y - \chi'_\pm \partial_X\right)^t,$$

and we define

$$\Delta_{\pm}^{y} := \nabla_{\pm}^{y} \cdot \nabla_{\pm}^{y}.$$

In the western layer, we find

$$\begin{aligned}&\left(\nabla_{-}^{y,\perp}\Psi_{-}^{0}(t,y,\cdot)\right)\cdot\nabla_{-}^{y}\left(\Delta_{-}^{y}\Psi_{-}^{0}(t,y,\cdot)\right)\\&\quad+\partial_X\Psi_{-}^{0}(t,y,\cdot)-\left(\Delta_{-}^{y}\right)^2\Psi_{-}^{0}(t,y,\cdot)=0.\end{aligned}\tag{115}$$

Note that (t, y) only plays the role of a parameter. We complete Eqs. (115) by boundary conditions (98),

$$\begin{aligned}&\Psi_{-}^{0}(t,y,\cdot)|_{\Gamma_-}=\int_{\chi_+(y)}^{\chi_-(y)}\operatorname{curl}\tau(t,x',y)\mathrm{d}x',\\&\left(n_-(y,Y)\cdot\nabla_X\Psi_{-}^{0}(t,y,\cdot)\right)|_{\Gamma_-}=0,\\&\Psi_{-}^{0}(t,y,\cdot)\text{ 1-periodic in }Y.\end{aligned}\tag{116}$$

The eastern layer. In the eastern layer, we obtain homogeneous equations: for $X \in \omega_+$,

$$\begin{aligned}&\left(\nabla_{+}^{y,\perp}\Psi_{-}^{0}(t,y,\cdot)\right)\cdot\nabla_{+}^{y}\Delta_{+}^{y}\Psi_{+}^{0}(t,y,\cdot)\\&\quad+\partial_X\Psi_{+}^{0}(t,y,\cdot)-\left(\Delta_{+}^{y}\right)^2\Psi_{+}^{0}(t,y,\cdot)=0.\end{aligned}\tag{117}$$

The boundary conditions are also homogeneous, thanks to our definition of Ψ^0:

$$\begin{aligned}&\Psi_{+}^{0}(t,y,\cdot)|_{\Gamma_-}=\left(n_+(y,Y)\cdot\nabla_X\Psi_{+}^{0}(t,y,\cdot)\right)|_{\Gamma_-}=0,\\&\Psi_{+}^{0}(t,y,\cdot)\text{ 1-periodic in }Y.\end{aligned}\tag{118}$$

Consequently, this system has the solution $\Psi_{+}^{0} \equiv 0$.

REMARK. Our choice for the stream function Ψ^0 is the same as in [78]. This is due to the properties of the eastern coast, which cannot bear a large boundary layer. Indeed, as explained in [78], the equation satisfied by the eastern profile Ψ_{+}^{0} in the non-rough case is

$$-\partial_X\Psi_{+}^{0}-(1+(\chi_{+}')^2)^2\partial_X^4\Psi_e^{+}=0.\tag{119}$$

The characteristic equation of this ODE has only one root with a negative real part, so that one cannot impose simultaneously conditions on Ψ_{+}^{0}, $\partial_X\Psi_{+}^{0}$ at the boundary *and* a condition at infinity. The same difficulty appears in the rough case, as Eqs (118) "contain" the differential equation (119). Note that this difficulty is not present in the western layer, as the underlying ODE

$$\partial_X \Psi^0_- - (1 + (\chi'_-)^2)^2 \partial^4_X \Psi^0_- = 0$$

has two characteristic roots with negative real parts.

This ends the formal computations. Starting from the rough domain Ω^ε, we have derived new models, defined on smooth domains. They give, at least formally, the leading dynamics of the flow. Therefore, we will discuss their qualitative properties in the following section. III. *Brief physical insight.* We end this part with a few qualitative elements on the boundary layer systems (110) and (115), (116), as well as on the interior system (114).

Boundary layers. The first point to notice about (110) and (115), (116) is that *they are partial differential equations (PDE's).* This is a big difference from the non-rough case, for which the boundary layers are governed by ordinary differential equations (ODE's).

Indeed, the Ekman layers are governed by:

$$i(u_{\pm,1} + iu_{\pm,2}) - \partial^2_Z(u_{\pm,1} + iu_{\pm,2}) = 0,$$

which is an ODE in variable Z (see [145]). In the same fashion, the classical Munk layer obeys to the equation

$$\partial_X \Psi^0_- - (1 + \chi'_-(y)^2)\partial^4_X \Psi^0_- = 0$$

which again is an ODE in variable X.

The reason for this change is that the small tangential scales of the roughness lead to tangential derivatives of the same order ε^{-1} as the normal ones. Therefore, all derivatives get involved in the boundary layer system. As a physical consequence, the dynamics of the fluid becomes more complex.

As another source of complexity, *the equations become nonlinear*. They include quadratic terms, derived from the advective term of the Navier–Stokes equations. There again, this is very different from the non-rough case, which involves only linear equations.

For all these reasons, the behavior of the flow is less tractable. In particular, no explicit resolution of the system is possible. However, thanks to various (and quite technical) mathematical tools, one can recover some information. For instance, one can show that these boundary layer solutions still have exponential decay. We refer to [94,45] for the proof of such results. We finish this section with a convergence theorem established in [45].

THEOREM 3.19. *Let Ψ^ε be the unique smooth solution of quasi-geostrophic model. There exists C_∞ such that if $\| \int_{\chi_+}^{\chi_-} \mathrm{curl}\tau \|_\infty < C_\infty$ then*

$$\|\Psi^\varepsilon - \Psi^\varepsilon_{\mathrm{app}}\|_{L^\infty(0,T;H^1(\Omega^\varepsilon))} \to 0 \quad \text{as } \varepsilon \to 0,$$

where

$$\Psi^\varepsilon_{\mathrm{app}}(t, x, y) = \Psi^0(t, x) + \Psi^0_-\left(t, y, \frac{x - \chi_-(y)}{\varepsilon}, \frac{y}{\varepsilon}\right)$$

with μ^0 and Ψ^0_- defined as before.

Note that in [45], existence and exponential decay of the boundary layer corrector rely on the implicit function theorem and main lines adapted from [91] to the periodic case, as in [9]. More precisely, we separate the stream-function Ψ into two parts: the main value $\overline{\Psi}$ with respect to the horizontal variables and the remaining part $\widetilde{\Psi}$. We prove that $\|\widetilde{\Psi}\|_{H^m(\omega^R)}$ decreases when $\|\Delta^2\widetilde{\Psi}\|_{L^2(\omega^R)}$ decreases with respect to R where $\omega^R = \omega_-^\infty \cap \{X > R\}$. Then we prove that $\|\Delta^2\widetilde{\Psi}\|_{L^2(\Omega^R)}$ decreases using the standard Gronwall inequality.

REMARK. It would be very interesting to generalize such an approach to a non-periodic roughness distribution and more generally to random spatially distributed roughness boundaries in order to be more accurate.

3.10.3. Wall laws and oceanography The general idea in wall laws is to remove the stiff part from boundary layers, replacing the classical no-slip boundary condition by a more sophisticated relationship between the variables and their derivatives. Depending on the field of applications, (porous media, fluid mechanics, heat transfer, electromagnetism), different names for wall laws are encountered such as SAFFMAN-JOSEPH, NAVIER, ROBIN etc. In the shallow-water theory, drag terms such as CHEZYE, MANNING are encountered. They come from small scale parametrization of bottom roughness and they are far from being justified since they involves highly nonlinear quantities. High order effective macroscopic boundary conditions may also be proposed, depending on the order we cut off the process in the Ansatz, such as WENTZELL boundary conditions (second-order boundary conditions). Anyway, in the steady case, numerical simulations have shown that first or second order macroscopic wall laws provide the same order of approximation. This is due to the complete averaging procedure from which we lose all the microscopic behavior. This is exactly the idea in [53] where new wall laws, including microscopic oscillations and a counter-example, are presented. More precise multi-scale wall laws are under considerations in a paper still in progress, see [31]. Note that we have seen that boundary conditions at the bottom greatly influence the derived shallow-water type systems: A non-slip boundary condition implies a parabolic main profile and a slip boundary condition implies a main profile that does not depend on the depth variable. In fact it would be interesting to look at roughness effects (depending on the difference between the roughness profile amplitude and the free surface aspect ratio) on the shallow-water systems derivation. In this section, we will not enter into details concerning wall laws and will not be exhaustive at all.
Periodic Case. Let us explain in a simple way the idea behind Navier type laws. Consider a stationary flow governed by the following system

$$\begin{cases} -\nu\Delta u^\varepsilon + \nabla p^\varepsilon = f & \text{in } \Omega_\varepsilon, \\ \operatorname{div} u^\varepsilon = 0 & \text{in } \Omega_\varepsilon, \\ u^\varepsilon = 0 & \text{on } \Gamma^\varepsilon, \end{cases} \tag{120}$$

where Γ^ε is a periodic oscillating boundary defined by a function γ and

$$\Omega^\varepsilon = \{(x, z) : x \in \mathbf{R}^2, \varepsilon\gamma(x/\varepsilon) < z\}.$$

Since the domain depends on the small parameter, the main idea is to approximate the system using a model defined on a fixed domain, namely

$$\begin{cases} -\nu\Delta u^1 + \nabla p^1 = f & \text{in } \Omega_0, \\ \operatorname{div} u^1 = 0 & \text{in } \Omega_0, \\ \varepsilon\overline{\beta}\dfrac{\partial u_*^1}{\partial z} + u_*^1 = 0, \qquad u_3^1 = 0 & \text{on } \Gamma^0, \end{cases} \tag{121}$$

where β is a well defined boundary corrector depending on $\xi = x/\varepsilon$ and $u^1 = (u_*^1, u_3^1)$. We note $\overline{\beta} = \frac{1}{|\pi|}\int_\pi \beta(y)\mathrm{d}y$, the mean value of β over the periodic cell. The new boundary condition on the bottom is called the Navier boundary condition. The formal asymptotics have been mathematically justified by A. MIKELIC- W. JAEGER, S. NAZAROV, O. PIRONNEAU et al. where the boundary corrector is described on a periodic cell and has an exponential decay property far from the boundary. It has already been proved that the H^1 error between u^ε and u^0 is $O(\varepsilon^{3/2})$. Note that in [53], we derive exact approximations of u^ε on the rugous wall up to the second order. For these fully oscillatory approximations we prove exponential convergence inside the domain. Then we show that, despite this great rate of convergence, the corresponding macroscopic wall law behaves badly and does not conserve the nice properties of the full boundary layer approximation. Microscopic oscillations have to be taken into account in wall laws. More precise multiscale wall-laws are still in progress in [31].

Generalization of wall laws functions formulation have been obtained recently for curved rough boundaries (see [136], [135] and references therein).

Non-periodic case. In some recent papers by A. BASSON, D. GÉRARD-VARET, see [16] and [95], the analysis, justifying Navier boundary conditions, has been extended to random spatially homogeneous boundaries. The main conclusion (that is approximation by a macroscopic Navier law) remains valid, but the analysis of the auxiliary system is much more difficult, and the asymptotic behaviour far from the boundary is modified, solutions do not decay exponentially fast. The justification of this behavior involves tools like Saint-Venant estimates, the ergodic theorem, or a central limit theorem for weakly correlated random variables. The last paper relies on ideas for homogenization of elliptic systems, see [14] and finds, under some conditions on the random repartition, an H^1 error up to $O(\varepsilon^{3/2}|\log\varepsilon|)$.

We also mention the recent papers [56,58,57], which have proposed a new method based on a mechanical and energetic account of arbitrary asperities in terms of Young and capacitary measures. It concerns arbitrary roughness and the periodic distribution considered in [64]. A mechanical account of the rugosity is developed in terms of parametrized (Young) measures. This allows us to obtain very general situations in which the perfect slip conditions are transformed in complete (or partial) adherence ones. Partial adherence means the effect of ribbed boundaries: perfect adherence in one direction and perfect slip in another. For instance, they prove that under mild assumptions any crystalline boundary produces the rugosity effect, regardless of the distribution of the asperities. On the second hand, an energetic account of the rugosity effect is performed in terms of capacitary measures. At this moment, they are able to estimate from an energetic point of view, how the flow is enabled to slip in a (given) direction of a vector field tangent to the boundary. This estimation is done in terms of a capacitary measure and involves a type of friction law. Very rough asperities lead to perfect adherence and mild asperities lead to a directional friction.

Conclusion. Previous sections employ a multi-scaled approach to derive new nonlinear boundary layers equations, new wall laws or new approximate equations. These results may be seen as parametrization of the effect of small-scale topography on large scale flows in geophysics. Of course a lot of studies have to been done: quasi-periodic framework, random spatially homogeneous framework on curved rough boundaries, non-Newtonian flows, multiphasic flows, multi-scale wall laws.

4. No-slip shallow-water equations

Suppose that one starts with standard Dirichlet boundary conditions instead of a wall law, which by itself is a model of boundary layers close to the bottom. Then, assuming a slope term or a strong enough external force, we get in the shallow-water type system a linear drag term due to the parabolic profile of the velocity, and the usual quadratic term $\partial_x(hu^2)$ is replaced by $6\partial_x(hu^2)/5$: see system (4) for the shallow-water equations derived in [158]. For the sake of completeness, we will give this formal derivation in what follows. Note that a rigorous derivation of such applicable shallow-water equations has been recently written with P. NOBLE, see [54] for a proof in one-dimensional space for shallow-water with a non-zero capillarity coefficient

THEOREM 4.1. *Assume that the Weber number* We *has the form* We $= \frac{\kappa}{\varepsilon^{5/2}}$. *Then there exists* $\varepsilon_1 > 0$ *such that if*

$$\frac{\sin(\theta)}{\sqrt{\varepsilon}\kappa} \le \varepsilon_1, \qquad \frac{|h_0|_\infty}{\varepsilon} + |\partial_x^2 h_0|_0 + \varepsilon^{\frac{1}{4}}|h_0|_{\frac{5}{2}} + \sqrt{\varepsilon}\|(u_0, w_0)\|_{H^2} \le \varepsilon_1, \tag{122}$$

and that $\frac{\sin(\theta)}{\kappa}$ *is sufficiently small, then the function* $\widetilde{\mathcal{R}}$, *given in* (168), *is bounded as follows*

$$\|\widetilde{\mathcal{R}}\|_{L^2(0,\infty;L^2(\boldsymbol{T}))} \le C\sqrt{\varepsilon}$$

with C, a constant which does not depend on ε.

This proves the convergence from the Navier–Stokes equations with free surface to a shallow-water type system with a non-zero surface tension coefficient. To extend this result to the zero surface tension coefficient is a really interesting open problem. Let us mention here for the convenience of the reader some recent works written by [25,27], in order to propose some generalizations of shallow-water equations which take into account first order variation in the slope of the bottom. Looking at the asymptotic with the adherence condition on the bottom and free surface conditions on the surface, he proves that depending on the Ansatz for the horizontal velocity and for the viscosity coefficient, we can formally get various asymptotic inviscid models at the main order.

Idea of the proof. The proof is divided into several steps. In the first part, we derive uniform estimates, with respect to the aspect ratio, of solutions of the Navier–Stokes equations combining similar process to those used in [138] and similar estimates to those in [154]. Reader interested in Navier–Stokes equations with free surface is also referred to [3,17, 18]. See also [155] for a result on Navier–Stokes flows down as inclined plane. We then use the formal derivation to conclude the rigorous derivation of the shallow-water model, proving that the remaining terms can be neglected. Note that our mathematical justification implies a non-zero capillary coefficient assumption. We also get some results concerning lubrication models which are included in the shallow-water mathematical derivation with no surface tension coefficients.

Lubrication models. As a by-product of the derivation of shallow-water equations, we shall obtain a hierarchy of a model with a single equation on the fluid height h. These are lubrication models. First of all let us recall that h and v satisfy the mass conservation equation

$$h_t + (hv)_x = 0. \tag{123}$$

As a first approximation, hv satisfies the expansion

$$hv = 2s\frac{h^3}{3} + \int_0^h (u - u^{(0)}). \tag{124}$$

Substituting Eq. (124) into (123), one obtains the Burgers-type equation

$$h_t + \left(s\frac{h^3}{3}\right)_x = -\partial_x \left(\int_0^h (u - u^{(0)})\right). \tag{125}$$

We then deduce from the analysis of the previous section that

$$-\partial_x \left(\int_0^h (u - u^{(0)})\right) = \mathcal{O}(\sqrt{\varepsilon})$$

in the $L^2\big((0,\infty), L^2(\mathbf{R})\big)$-norm and this justifies the fact that the Burgers-type equation (125) is a first approximation of shallow-water flows. Furthermore, we can obtain more accurate models, using the more precise expansion of hv:

$$(hv)(t,x) = 2s\frac{h^3(t,x)}{3} + \varepsilon \int_0^{h(t,x)} \overline{u}^{(1)}(t,x,\zeta)\mathrm{d}\zeta + \varepsilon \mathrm{Re} \int_0^{h(t,x)} \mathcal{R}_2^{(1)}(t,x,\zeta)\mathrm{d}\zeta, \tag{126}$$

where $\overline{u}^{(1)}$ is the function

$$\overline{u}^{(1)} = -2p^{(0)} - \mathrm{Re}\left(u_t^{(0)} + u^{(0)}u_x^{(0)} + w^{(0)}u_z^{(0)}\right).$$

Then there is a lengthly but straightforward computation to show that h satisfies the lubrication equation

$$\partial_t h + \partial_x \left(\frac{h^3}{3} \left(2s + \varepsilon \left(\mathrm{Re} \frac{8s^2}{5} h^3 - 2c \right) h_x + 2\varepsilon\overline{\kappa} h_{xxx} \right) \right) = \mathcal{R}_{lub} \qquad (127)$$

and the function $\mathcal{R}_{lub}$ satisfies the estimate

$$\|\mathcal{R}_{lub}\|_{L^2(0,\infty;L^2(\mathbf{R}))} \le C\varepsilon^{\frac{3}{2}}.$$

This justifies the lubrication approximation for shallow-water flows

$$\partial_t h + \partial_x \left(\frac{h^3}{3} \left(2s + \varepsilon \left(\mathrm{Re} \frac{8s^2}{5} h^3 - 2c \right) h_x + 2\varepsilon\overline{\kappa} h_{xxx} \right) \right) = 0. \qquad (128)$$

At this stage, one can derive a viscous Burgers equation: let us write h in the form $h = 1 + \varepsilon\overline{h}(\varepsilon(x - 2st), \varepsilon^2\tau)$. This is precisely the diffusive scaling used in [156], to derive a viscous Burgers equation from the full Navier-Stokes system. Then one proves that $\overline{h}$ satisfies, up to zeroth order in ε, the equation

$$\partial_t \overline{h} + \alpha\partial_x(\overline{h}^2) = \beta\partial_{xx}\overline{h}, \quad \alpha, \beta > 0. \qquad (129)$$

As a conclusion, we recover a result similar to that of H. UECKER [156] in the particular case of *periodic* functions in the streamwise variable: here we can also prove that the perturbation $\overline{h}$ decays exponentially, whereas the localized solutions of (129) have a self-similar decay. Reader interested by results around lubrication approximation are referred to [21,84].

REMARK. It is really important to understand that the derivation procedure proposed by J.-P. VILA (reproduced below and generalized in [28] for arbitrary bottom topography) is really general. This means that more complex fluids may be considered, such as power law flow or Bingham fluids. It provides a hierarchy of models and includes, for instance, lubrication models derived by C. ANCEY, N. BALMFORTH, M. WEI and others. Such studies are still in progress by L. CHUPIN, P.NOBLE, J.-.P. VILA. See also [147,10] for a review on avalanche models.

REMARK. Our mathematical justification assumes periodic horizontal boundary conditions. It would be interesting to work on rigorous shallow-water type approximations with contact angle, see [144,97] for a lubrication approximation.

Formal shallow-water equations derivation from no-slip bottom boundary conditions. In what follows, we reproduce the formal derivation, due to J.-P. VILA, of the shallow-water model from the incompressible Navier–Stokes equations with a free surface and no-slip condition on the bottom. The fluid flows downward to an inclined plane under the effect of gravity. The Navier–Stokes system comes with boundary conditions: we assume a no-slip condition at the bottom and continuity of the fluid stress at the free surface, the fluid being submitted here to surface tension forces. Moreover, assuming that the layer of fluid is advected by the fluid velocity, we obtain an evolution equation for the fluid height. We first scale the equations in the shallow-water setting. Then we calculate formally an asymptotic expansion of the flow variables with respect to the film parameter ε. Inserting the expansion in depth-averaged continuity and momentum equation, we obtain a shallow-water model and identify a remainder that formally tends to 0 as $\varepsilon \to 0$.

4.1. *Scaling the incompressible Navier–Stokes system*

We consider a relatively thin layer of fluid flowing down an inclined plane, at an angle θ with respect to the horizontal, under the effect of gravity. The fluid is incompressible and the fluid density is constant and set to 1. The flow is supposed to be 2-dimensional. A coordinate system (x, z) is defined with the x-axis down the slope and the z-axis upwards, normal to the plane bed. The longitudinal and transverse velocity are denoted by (u, w), the pressure by p and the total flow depth by h. The fluid layer Ω_t is the set

$$\Omega_t = \left\{(x, z) \in \mathbf{R}^2 : 0 < z < h(t, x)\right\}.$$

The incompressible Navier-Stokes equations reads, for all $(x, z) \in \Omega_t$:

$$\begin{cases} u_t + uu_x + wu_z + p_x = g\sin\theta + \nu\Delta u, \\ w_t + uw_x + ww_z + p_z = -g\cos\theta + \nu\Delta w, \\ u_x + w_z = 0. \end{cases} \tag{130}$$

The ν constant denotes the fluid viscosity and g is the gravity constant. The Navier-Stokes equation is supplemented by boundary conditions. More precisely, we assume at the bottom a no-slip condition:

$$u|_{z=0} = w|_{z=0} = 0. \tag{131}$$

At the free boundary, we assume that the atmospheric pressure p_{atm} is constant and set to 0. The fluid is submitted to surface tension forces. Hence, the continuity of the fluid stress at the free surface yields the following conditions:

$$\begin{aligned} &p|_{z=h} + \kappa h_{xx}(1 + h_x^2)^{-\frac{3}{2}} = -2\nu\frac{1 + h_x^2}{1 - h_x^2}u_x|_{z=h}, \\ &u_z|_{z=h} + w_x|_{z=h} = -4\frac{h_x}{1 - h_x^2}w_z|_{z=h}, \end{aligned} \tag{132}$$

where κ measures the capillarity. The fluid layer is advected by the speed $\vec{u} = (u, w)$, giving the evolution equation for the fluid height (the usual kinematic condition)

$$h_t + h_x u|_{z=h} = w|_{z=h}. \tag{133}$$

This set of equations possesses a classical steady solution, the so-called Nusselt flow. The height of the fluid $h(x,t) = H$ is constant, the transverse component velocity w is equal to 0 and the longitudinal component speed u has a parabolic dependence with respect to z

$$U(x,z) = \frac{g}{2\nu}(2Hz - z^2)\sin\theta. \tag{134}$$

Moreover, the pressure is hydrostatic:

$$p(x,z) = g(H - z)\cos\theta. \tag{135}$$

Adimensionalization and rescaling. We shall use the Nusselt flow in the sequel to scale the Navier–Stokes equations. Using the notation $U_0 = gH^2/2\nu$, we define the rescaled variables

$$\begin{aligned}
&z = H\overline{z}, \qquad h = H\overline{h}, \qquad x = L\overline{x}, \qquad \varepsilon = \frac{H}{L},\\
&u = U_0\overline{u}, \quad w = \varepsilon U_0\overline{w}, \qquad p = gH\overline{p},\\
&t = \frac{L}{U_0}\overline{t}, \qquad \mathrm{Re} = \frac{HU_0}{\nu}, \qquad \mathrm{We} = \frac{\kappa}{gH^2}.
\end{aligned} \tag{136}$$

In what follows, we denote $c = \cos\theta$, $s = \sin\theta$. Under the shallow-water scaling and dropping the bar over the dot, the Navier–Stokes equations read for all (x,z) in Ω_t:

$$\begin{aligned}
&u_t + u\,u_x + w\,u_z + \frac{2}{\mathrm{Re}}p_x = \frac{2s}{\varepsilon\,\mathrm{Re}} + \frac{1}{\varepsilon\,\mathrm{Re}}(\varepsilon^2\,u_{xx} + u_{zz}),\\
&w_t + u\,w_x + w\,w_z + \frac{2}{\varepsilon^2\,\mathrm{Re}}p_z = -\frac{2c}{\varepsilon^2\,\mathrm{Re}} + \frac{1}{\varepsilon\,\mathrm{Re}}(\varepsilon^2\,w_{xx} + w_{zz}),\\
&u_x + w_z = 0.
\end{aligned} \tag{137}$$

The no-slip condition $u|_{z=0} = w|_{z=0} = 0$ and the evolution equation for h

$$h_t + h_x\,u|_{z=h} = w|_{z=h}, \tag{138}$$

are unchanged, whereas the continuity of the fluid stress reads

$$
\begin{aligned}
&p|_{z=h} + \varepsilon^2 \,\mathrm{We}\, h_{xx}(1+\varepsilon^2 h_x^2)^{-\frac{3}{2}} = -\varepsilon \frac{1+\varepsilon^2 h_x^2}{1-\varepsilon^2 h_x^2} u_x|_{z=h},\\
&u_z|_{z=h} + \varepsilon^2 w_x|_{z=h} = -\frac{4\varepsilon^2 h_x}{1-\varepsilon^2 h_x^2} w_z|_{z=h}.
\end{aligned}
\tag{139}
$$

In fact, we prove that in the asymptotic regime $\varepsilon \to 0$, the solutions of the Navier–Stokes equations (137) with rescaled boundary conditions are close to the stationary Nusselt flow and have an asymptotic expansion with respect ε. This is done in the next section.

4.2. *Asymptotic expansion of solutions*

Under the shallow-water scaling, the flow variables are close to a Nusselt flow. More precisely, we write the Navier–Stokes equations as a differential system in the cross-stream variable z with boundary conditions at $z = 0$ and $z = h$. First, the longitudinal component velocity u satisfies

$$
\begin{aligned}
&u_{zz} + 2\,s = 2\,\varepsilon p_x + \varepsilon \,\mathrm{Re}\,(u_t + u u_x + w\, u_z) - \varepsilon^2 u_{xx}\\
&\qquad\quad = \Psi_u\,(u, w, p)\,(t, x, z),\\
&u|_{z=0} = 0,\\
&u_z|_{z=h} = -\varepsilon^2 w_x|_{z=h} + \frac{4\varepsilon^2}{1-\varepsilon^2 h_x^2} h_x\, u_x|_{z=h}\\
&\qquad\quad = \Pi_u\,(u, w)\,(t, x).
\end{aligned}
\tag{140}
$$

The functions $\Psi_u\,(u, w, p)$, $\Pi_u\,(u, w)$ satisfy formally the estimates

$$
\Psi_u\,(u, w, p) = \mathcal{O}\,(\varepsilon + \varepsilon \mathrm{Re}), \qquad \Pi_u\,(u, w) = \mathcal{O}\left(\varepsilon^2\right). \tag{141}
$$

Integrating (140) with respect to z, we obtain

$$
u(t, x, z) = 2s\left(h(t, x)z - \frac{z^2}{2}\right) + \mathcal{F}_u\,(u, w, p)\,(t, x, z), \tag{142}
$$

where the function $\mathcal{F}_u\,(u, w, p)$ is defined by

$$
\begin{aligned}
&\mathcal{F}_u\,(u, w, p)\,(t, x, z)\\
&\quad = z\Pi_u\,(u, w)\,(t, x) - \int_0^z \int_{\overline{z}}^h \Psi_u\,(u, w, p)\,(t, x, y)\mathrm{d}y\mathrm{d}\overline{z}.
\end{aligned}
\tag{143}
$$

As a consequence, the longitudinal velocity component u expands in the form

$$u(t,x,z) = 2s\left(h(t,x)z - \frac{z^2}{2}\right) + \mathcal{O}\left(\varepsilon + \varepsilon \mathrm{Re}\right). \tag{144}$$

The profile for u is then closed to a Nusselt type flow. We show in a similar way that the pressure is close to a hydrostatic distribution. More precisely, the pressure p satisfies the ordinary differential equation with respect to z:

$$\begin{aligned} p_z + c &= \frac{\varepsilon}{2}(w_{zz} + \varepsilon^2 w_{xx}) - \frac{\varepsilon^2\,\mathrm{Re}}{2}(w_t + uw_x + ww_z) \\ &= \Psi_p(u,w)(t,x,z), \\ p|_{z=h} &= -\overline{\kappa}\frac{\mathrm{Re}}{2}\frac{h_{xx}}{(1+\varepsilon^2\,h_x^2)^{\frac{3}{2}}} - \varepsilon\,u_x|_{z=h}\frac{1+\varepsilon^2\,h_x^2}{1-\varepsilon^2\,h_x^2} \\ &= -\overline{\kappa}\frac{\mathrm{Re}}{2}h_{xx} + \Pi_p\,(u)\,(t,x), \end{aligned} \tag{145}$$

where the constant $\overline{\kappa}$, defined by $\overline{\kappa} = \varepsilon^2\mathrm{We}$, is assumed to be of order $\mathcal{O}(1)$ in order to see the capillarity effects. The functions $\Psi_p(u,w)$, $\Pi_p\,(u)$ satisfy the formal estimates

$$\Psi_p(u,w) = \mathcal{O}\left(\varepsilon\right), \qquad \Pi_p(u) = \mathcal{O}\left(\varepsilon\right). \tag{146}$$

We integrate (145) with respect to z and find

$$p(x,z,t) = -\overline{\kappa}\frac{\mathrm{Re}}{2}h_{xx} + c(h-z) + \mathcal{F}_p(u,w,p)(t,x,z), \tag{147}$$

where the function $\mathcal{F}_p(u,w,p)$ is defined by

$$\mathcal{F}_p(u,w,p) = \Pi_p(u)(x,t) - \int_z^h \Psi_p(u,w)(t,x,y)\mathrm{d}y. \tag{148}$$

The pressure p has, in a first approximation, an hydrostatic distribution and expands to the form

$$p(x,z,t) = -\overline{\kappa}\frac{\mathrm{Re}}{2}h_{xx} + c(h-z) + \mathcal{O}(\varepsilon). \tag{149}$$

The transverse component velocity w is determined, using the free divergence condition $w_z + u_x = 0$ and the no-slip condition $w(x,0) = 0$. More precisely, we find $w(t,x,z) = -\int_0^z u_x(t,x,y)\mathrm{d}y$. We deduce that w has the asymptotic expansion

$$w(t,x,z) = -\int_0^z u_x(t,x,y)\mathrm{d}y = -sh_x z^2 + \mathcal{O}(\varepsilon + \varepsilon\mathrm{Re})\,. \tag{150}$$

In the following, we describe an iterative scheme to compute a formal expansion of Navier-Stokes solutions to *any* order in ε. First, we define the functions $u^{(0)}$ and $p^{(0)}$ as

$$\begin{aligned} u^{(0)}(t,x,z) &= s(2h(t,x)z - z^2), \\ p^{(0)}(t,x,z) &= -\overline{\kappa}\,\mathrm{Re}\,h_{xx}/2 + c\,(h(t,x) - z)\,. \end{aligned} \tag{151}$$

We write the Eqs. (142) and (147) in the form

$$\begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} u^{(0)} \\ p^{(0)} \end{pmatrix} + \mathcal{F}(u,p,\varepsilon), \tag{152}$$

where the function $\mathcal{F}(u,p,\varepsilon)$ is defined by

$$\mathcal{F}(u,p,\varepsilon) = \begin{pmatrix} \mathcal{F}_u(u,w,p) \\ \mathcal{F}_p(u,w,p) \end{pmatrix}, \qquad w(t,x,z) = -\int_0^z u_x(t,x,y)\mathrm{d}y. \tag{153}$$

Then, any solution of the Navier-Stokes equations appears as a fixed point of an operator that is formally $\mathcal{O}(\varepsilon + \varepsilon\mathrm{Re})$-Lipschitz on any bounded set. Following the proof of the fixed point theorem, we define the sequence of functions $\left(u^n, w^{(n)}, p^{(n)}\right)_{n\geq 1}$ such that

$$\begin{pmatrix} u^{(n+1)} \\ p^{(n+1)} \end{pmatrix} = \begin{pmatrix} u^{(0)} \\ p^{(0)} \end{pmatrix} + \mathcal{F}(u^{(n)}, p^{(n)}, \varepsilon) \tag{154}$$

and $w^{(n)}(t,x,z) = -\int_0^z u_x^{(n)}(t,x,y)\mathrm{d}y$.
We assume that all the derivatives of (u,w,p) remain bounded and the sequence $\left(u^n, w^{(n)}, p^{(n)}\right)_{n\geq 1}$ is bounded. We can prove by induction the formal estimate

$$\max\left(|u - u^{(n)}|, |w - w^{(n)}|, |p - p^{(n)}|\right) = \mathcal{O}\left((\varepsilon + \varepsilon\mathrm{Re})^{n+1}\right). \tag{155}$$

As a consequence, we clearly see that the n-th term of the sequence of functions $\left(u^n, w^{(n)}, p^{(n)}\right)_{n\geq 1}$ is a formal approximation of a solution of the full Navier-Stokes equations up to order $\mathcal{O}\left((\varepsilon + \varepsilon\mathrm{Re})^{n+1}\right)$. In what follows, we use that approximation of the solutions to close the depth-average continuity and momentum equations and obtain a shallow-water model.

4.3. *The shallow-water model*

In the following, we write the shallow-water model for the fluid height h and the total discharge rate $hv = \int_0^h u(.,y,.)\mathrm{d}y$. Here v represents the mean velocity along the depth flow.

On the one hand, integrating the divergence free condition $u_x + w_z = 0$ along the fluid height and using the kinematic equation for h namely $h_t + h_x u|_{z=h} = w|_{z=h}$, we find

$$h_t + \left(\int_0^h u(x,z)\,\mathrm{d}z\right)_x = h_t + (hv)_x = 0. \tag{156}$$

Note that the equation is exact and already in a closed form. Let us now write an evolution equation for $hv = \int_0^h u(\cdot,\cdot,\zeta)\mathrm{d}\zeta$. For that purpose, we integrate the evolution equation on the velocity component u along the flow depth and substitute the boundary conditions (139) into the resulting equation:

$$\begin{aligned}\left(\int_0^h u\,\mathrm{d}z\right)_t + \left(\int_0^h u^2\,\mathrm{d}z\right)_x + \frac{2}{\mathrm{Re}}\left(\int_0^h p\,\mathrm{d}z\right)_x + \overline{\kappa}\frac{h_{xx}}{(1+\varepsilon^2 h_x^2)^{\frac{3}{2}}} h_x \\ = \frac{1}{\varepsilon\,\mathrm{Re}}\,(2\,sh - u_z(x,0)) + \frac{\varepsilon}{\mathrm{Re}}\left(\int_0^h 2\,u_x\,\mathrm{d}z\right)_x.\end{aligned} \tag{157}$$

In order to write a momentum equation in a closed form, we shall calculate an expansion of the averaged quantities in (157) with respect to ε, h, v. First, let us note that

$$\begin{aligned}(hv)(t,x) &= \int_0^{h(t,x)} u(t,x,\zeta)\mathrm{d}\zeta = \int_0^{h(t,x)} u^{(0)}(t,x,\zeta)\mathrm{d}\zeta + \int_0^{h(t,x)} \delta\, u(t,x,\zeta)\mathrm{d}\zeta \\ &= 2s\frac{h^3}{3} + \int_0^{h(t,x)} \delta\, u(t,x,\zeta)\mathrm{d}\zeta.\end{aligned}$$

Here, the function δu is simply $\delta u = u - u^{(0)}$. From the previous section, we deduce the formal estimate $\delta u = \mathcal{O}(\varepsilon + \varepsilon\mathrm{Re})$. We compute the average quantities $\int_0^h u^2$ and $\int_0^h p$:

$$\begin{aligned}\int_0^{h(t,x)} u^2(t,x,\zeta)\mathrm{d}\zeta &= \int_0^{h(t,x)} (u^{(0)})^2(t,x,\zeta)\mathrm{d}\zeta \\ &\quad + \int_0^{h(t,x)} \left(u^2 - (u^{(0)})^2\right)(t,x,\zeta)\mathrm{d}\zeta \\ &= \frac{8h(t,x)}{15}(sh^2(t,x))^2 \\ &\quad + \int_0^{h(t,x)} \left(u^2 - (u^{(0)})^2\right)(t,x,\zeta)\mathrm{d}\zeta.\end{aligned} \tag{158}$$

We substitute $2sh^2 = 3v - \frac{1}{h}\int_0^h \delta u$ into (158). The average quantity $\int_0^h u^2$ reads

$$\int_0^{h(t,x)} u^2(t,x,\zeta)\mathrm{d}\zeta = \frac{6}{5}(hv^2)(t,x) + \mathcal{R}_1^{(1)}, \tag{159}$$

where the function $\mathcal{R}_1^{(1)}$ is defined by

$$\mathcal{R}_1^{(1)}(t,x) = \int_0^{h(t,x)} (u^2 - (u^{(0)})^2)(t,x,\zeta)\mathrm{d}\zeta + \frac{12}{15} v \int_0^{h(t,,x)} \delta u(t,x,\zeta)\mathrm{d}\zeta + \frac{2}{15\, h(t,x)} \left(\int_0^{h(t,x)} \delta u(t,x,\zeta)\mathrm{d}\zeta \right)^2 .$$

Inserting the estimate $u - u^{(0)} = \mathcal{O}(\varepsilon + \varepsilon \mathrm{Re})$ into the expression for $\mathcal{R}_1$, we find that, formally, $\mathcal{R}_1 = \mathcal{O}(\varepsilon + \varepsilon \mathrm{Re})$. We easily compute the average quantity $\int_0^h p$:

$$\int_0^h p(\cdot,\cdot,\zeta)\mathrm{d}\zeta = -\overline{\kappa}\frac{\mathrm{Re}}{2} h h_{xx} + c\frac{h^2}{2} + \mathcal{R}_1^{(2)} \tag{160}$$

with

$$\mathcal{R}_1^{(2)}(t,x) = \int_0^{h(t,x)} \left(p - p^{(0)}\right)(t,x,\zeta)\mathrm{d}\zeta.$$

We easily prove that, formally, $\mathcal{R}_1^{(2)}$ satisfies the estimates $\mathcal{R}_1^{(2)} = \mathcal{O}(\varepsilon + \varepsilon \mathrm{Re})$. Substituting (159) and (160) into (157), we obtain

$$(hv)_t + \left(\frac{6}{5} h v^2 + \frac{2}{\mathrm{Re}} c \frac{h^2}{2} \right)_x - \overline{\kappa} h h_{xxx} = \frac{1}{\varepsilon\, \mathrm{Re}} (2sh - u_z(x,0)) + \mathcal{R}_1, \tag{161}$$

where the function $\mathcal{R}_1$ is defined by

$$\mathcal{R}_1 = -\partial_x \mathcal{R}_1^{(1)} - \partial_x \mathcal{R}_1^{(2)} + \frac{\varepsilon}{\mathrm{Re}} \partial_x \left(\int_0^h u_x(\cdot,\cdot,\zeta)\mathrm{d}\zeta \right) + \overline{\kappa} \left(\left(1 - (1 + \varepsilon^2 h_x^2)^{-\frac{3}{2}} \right) h_x h_{xx} \right) \tag{162}$$

and satisfies the formal estimate

$$\mathcal{R}_1 = \mathcal{O}(\varepsilon(\mathrm{Re} + 1 + \mathrm{Re}^{-1})).$$

Dropping the "small" term $\mathcal{R}_1$, we can see that the momentum equation (161) is almost in a closed form and yields a shallow-water model for h, hv. There remains the elimination of $u_z(x,0)$. Due to the presence of the singular factor $1/(\varepsilon \mathrm{Re})$, we cannot directly use the expansion

$$u_z(0) = u_z^{(0)}(0) + \mathcal{O}(\varepsilon + \varepsilon \mathrm{Re}) = \frac{3v}{h} + \mathcal{O}(\varepsilon + \varepsilon \mathrm{Re}).$$

In that case, the remainder would be of order $\mathcal{O}(1)$. As a consequence, we need an asymptotic expansion of $u_z(0)$ and hv with respect to ε and a function of h and its derivatives up to order 1. For that purpose, we use the function $u^{(1)}$ introduced in the previous section. Recall that we have the estimate $u - u^{(1)} = \mathcal{O}\left((\varepsilon + \varepsilon \mathrm{Re})^2\right)$. We aim to compute an expansion of hv and $u_z(.,0)$ up to order 1: let us expand u with respect to ε. We find, using the iterative scheme introduced previously, that

$$u = u^{(0)} - \int_0^z \int_y^h 2\varepsilon\, p^{(0)} + \varepsilon \mathrm{Re}(u_t^{(0)} + u^{(0)} u_x^{(0)} + w^{(0)} u_z^{(0)}) + \varepsilon \mathrm{Re}\, \mathcal{R}_2^{(1)} \quad (163)$$

where the notation $\int_0^z \int_y^h f$ is used to define the function of (t, x, z) as

$$\int_0^z \int_y^h f(t,x) = \int_0^z \int_y^{h(t,x)} f(t,x,\bar{z}) \mathrm{d}\bar{z} \mathrm{d}y.$$

Moreover, the function $\mathcal{R}_2^{(1)}$ is defined as

$$\begin{aligned}
\mathcal{R}_2^{(1)} &= \frac{\varepsilon}{\mathrm{Re}} \int_0^z \int_y^h u_{xx} - u_{xx}^{(0)} - \int_0^z \int_y^h \frac{2}{\mathrm{Re}} (p - p^{(0)})_x \\
&\quad - \frac{\varepsilon z}{\mathrm{Re}} \left((w_x(h) + \frac{w_z(h)}{1 - \varepsilon^2 h_x^2}) - (w_x^{(0)}(h) + \frac{w_z^{(0)}(h)}{1 - \varepsilon^2 h_x^2}) \right) \\
&\quad + \int_0^z \int_y^h (u - u^{(0)})_t + (u u_x - u^{(0)} u_x^{(0)}) + (w u_z - w^{(0)} u_z^{(0)}).
\end{aligned}$$

It is easily seen that $\mathcal{R}_2^{(1)} = \mathcal{O}\left(\varepsilon(1 + \mathbf{Re} + \mathrm{Re}^{-1})\right)$. In order to simplify the notations, we denote by $\overline{u}^{(1)}$ the function

$$\overline{u}^{(1)} = -2\, p^{(0)} - \mathrm{Re}(u_t^{(0)} + u^{(0)} u_x^{(0)} + w^{(0)} u_z^{(0)}).$$

Then the component velocity u expands in the form

$$u = u^{(0)} + \varepsilon\, \overline{u}^{(1)} + \varepsilon \mathrm{Re} \mathcal{R}_2^{(1)}. \quad (164)$$

From (164) we deduce that

$$u_z|_{z=0} = 2sh + \varepsilon\, \partial_z \overline{u}^{(1)}|_{z=0} + \varepsilon \mathrm{Re} \partial_z \mathcal{R}_2^{(1)}|_{z=0},$$

$$(hv)(t,x) = 2s\frac{h^3(t,x)}{3} + \varepsilon\int_0^{h(t,x)} \overline{u}^{(1)}(t,x,\zeta)\mathrm{d}\zeta + \varepsilon\mathrm{Re}\int_0^{h(t,x)} \mathcal{R}_2^{(1)}(t,x,\zeta)\mathrm{d}\zeta.$$

Then eliminating $2sh$ from the expansion of $u_z|_{z=0}$, one obtains the asymptotic expansion

$$u_z|_{z=0} = \frac{3v}{h} + \varepsilon\left(\partial_z\overline{u}^{(1)}|_{z=0} - \frac{3}{h^2}\int_0^h \overline{u}^{(1)}\right) + \varepsilon\mathrm{Re}\left(\partial_z\mathcal{R}_2^{(1)}|_{z=0} - \frac{3}{h^2}\int_0^h \mathcal{R}_2^{(1)}\right). \tag{165}$$

We substitute the expansion (165) into the momentum equation (161)

$$(hv)_t + \left(\frac{6}{5}hv^2 + \frac{2}{\mathrm{Re}}c\frac{h^2}{2}\right)_x - \overline{\kappa}hh_{xxx} = \frac{1}{\varepsilon\,\mathrm{Re}}\left(2sh - \frac{3v}{h}\right) - \tau + \mathcal{R}_1 + \mathcal{R}_2, \tag{166}$$

with $\tau = \frac{7}{120}(2s)^2\,h^4\,h_x + \frac{1}{8}(2s)\,h^2\,h_t$ and $\mathcal{R}_2 = \partial_z\mathcal{R}_2^{(1)}|_{z=0} - \frac{3}{h^2}\int_0^h \mathcal{R}_2^{(1)}$. Dropping the "small" term $\mathcal{R} = \mathcal{R}_1 + \mathcal{R}_2 = \mathcal{O}(\varepsilon(1 + \mathrm{Re} + \mathrm{Re}^{-1}))$, we obtain a shallow-water model in a closed form

$$\begin{cases} h_t + (hv)_x = 0, \\ (hv)_t + \left(\frac{6}{5}hv^2 + \frac{2}{\mathrm{Re}}c\frac{h^2}{2}\right)_x - \overline{\kappa}hh_{xxx} = \frac{1}{\varepsilon\,\mathrm{Re}}\left(2sh - \frac{3v}{h}\right) - \tau. \end{cases} \tag{167}$$

The term τ can be written as the sum of a conservative term and a remainder of order $\mathcal{O}(\varepsilon + \varepsilon\mathrm{Re})$. More precisely, we substitute the expansion

$$hv = 2s\frac{h^3}{3} + \int_0^h (u - u^{(0)})$$

into the conservation law $h_t + (hv)_x = 0$. As a result we find that

$$h_t = -2sh^2\,h_x - \partial_x\left(\int_0^h (u - u^{(0)})\right)$$

and the function τ reads $\tau = -\frac{(2s)^2}{75}(h^5)_x - \frac{sh^2}{4}\partial_x\left(\int_0^h (u - u^{(0)})\right)$. If we introduce the notation $\widetilde{\mathcal{R}} = \mathcal{R} + \frac{sh^2}{4}\partial_x\left(\int_0^h u - u^{(0)}\right)$, the momentum equation (166) reads

$$(hv)_t + \left(\frac{6}{5}hv^2 + \frac{2}{\mathrm{Re}}c\frac{h^2}{2} - \frac{(2s)^2}{75}h^5\right)_x - \overline{\kappa}hh_{xxx} = \frac{1}{\varepsilon\,\mathrm{Re}}\left(2sh - \frac{3v}{h}\right) + \widetilde{\mathcal{R}}. \quad (168)$$

The remainder $\widetilde{\mathcal{R}}$ is formally of order $\mathcal{O}(\varepsilon(1 + \mathrm{Re} + \mathrm{Re}^{-1}))$. Dropping this "small" term, we obtain a conservative form of the shallow-water model (167):

$$\begin{cases} h_t + (hv)_x = 0, \\ (hv)_t + \left(\frac{6}{5}hv^2 + \frac{c\,h^2}{\mathrm{Re}} - \frac{(2s)^2}{75}h^5\right)_x \\ \quad - \overline{\kappa}hh_{xxx} = \frac{1}{\varepsilon\,\mathrm{Re}}\left(2sh - \frac{3v}{h}\right). \end{cases} \quad (169)$$

This concludes the formal derivation of a shallow-water model from the full Navier–Stokes equations with free surface. Formally the remainder term $\mathcal{R}$ (resp. $\widetilde{\mathcal{R}}$) tends to 0 as $\varepsilon \to 0$. To justify the formal derivation presented here, we compute *a priori* estimates in classical Sobolev norms on the solutions (u, w, p) of the full Navier–Stokes system and prove rigorously that the convergence of $\mathcal{R}$ (resp. $\widetilde{\mathcal{R}}$) to 0 holds true in a suitable norm. Finally, we can easily obtain a shallow-water model without capillarity by setting the capillary coefficient $\overline{\kappa}$ to 0 in (169). However, capillarity is important to obtain energy estimates on solutions to Navier–Stokes equations and to justify the asymptotic process between Navier–Stokes and shallow-water.

5. Coupling models — Some comments

5.1. *Multi-fluid models*

This section closely follows the introduction in [33]. Reader interested by modeling at different scales for multiphase flows is referred to [77]. Most of the flows encountered in nature are multi-fluid flows. Such a terminology encompasses flows of non-miscible fluids like air and water, oil and water or gas and water, to cite a few. When two fluids are miscible, they form in general a "new" single fluid with its own rheological properties. A stable emulsion between water and oil is an interesting example: this emulsion is a non-Newtonian fluid while water and oil are Newtonian ones.

The simplest, although non-trivial, multi-fluid flow is certainly small amplitude waves propagating at the interface between air and water, a separated flow in the two-fluid terminology. In this case, as far as modeling is concerned, each fluid obeys to its own model and the coupling occurs through the free surface. One of the mathematical models involved here is the Euler equation with a free surface:

$$\begin{cases} \partial_t \rho + \operatorname{div}(\rho u) = 0, \\ \partial_t(\rho u) + \operatorname{div}(\rho u \otimes u) + \nabla p = \rho g, \end{cases} \tag{170}$$

where g denotes gravity. The volume occupied by the two fluids (air and water) is separated by the unknown free surface $z = \eta(x, y, t)$ and above we have air, $\rho = \rho^-$, while below we have water, $\rho = \rho^+$. These fluids have their own equations of state:

$$\mathcal{F}^{\pm}(\rho^{\pm}, p) = 0. \tag{171}$$

As already stated, the coupling between these two independent Euler equations is done by the free surface and we have the kinematic relation

$$\frac{\partial \eta}{\partial t} + u_x \frac{\partial \eta}{\partial x} + u_y \frac{\partial \eta}{\partial y} = u_z \tag{172}$$

on both side of the surface $z = \eta$ and the pressure is continuous across this surface.

As the amplitude of the wave grows, wave breaking may occur. Then, in the vicinity of the interface between air and water, small droplets of liquid are present in the gas. Bubbles of gas are also present in the liquid. These inclusions might be very small. Collapse and fragmentation occur, making the free surface topologically very complicated and involving a large variety of length scales. We are in one of the situations in which two-fluid models become relevant if not unavoidable. The classical modeling techniques consists in performing a volume average to derive a model without a free surface, a two-fluid model. We denote by $\alpha^+(x, t)$ the volumetric rate of the presence of fluid $+$, the liquid in the case discussed before, and $\alpha^-(x, t)$ that of fluid $-$, the gas. That is in an infinitesimal volume $\mathrm{d}x^3$ around the point x, the volume occupied by the fluid $+$ is $\alpha^+(x, t)\,\mathrm{d}x^3$ while that occupied by fluid $-$ is $\alpha^-(x, t)\,\mathrm{d}x^3$. Hence $\alpha^+ + \alpha^- = 1$. The averaging procedure (see *e.g.* M. ISHII [104]), when it is applied to (170), leads then to:

$$\begin{cases} \partial_t(\alpha^{\pm}\rho^{\pm}) + \operatorname{div}(\alpha^{\pm}\rho^{\pm}u^{\pm}) = 0, \\ \partial_t(\alpha^{\pm}\rho^{\pm}u^{\pm}) + \operatorname{div}(\alpha^{\pm}\rho^{\pm}u^{\pm} \otimes u^{\pm}) + \alpha^{\pm}\nabla p = \alpha^{\pm}\rho^{\pm} g \pm F_D, \end{cases} \tag{173}$$

where the drag force F_D is proportional to $u^+ - u^-$ and is given in terms of the other variables. For example a classical expression is:

$$F_D = \frac{C}{\ell} \frac{\alpha^+\alpha^-\rho^+\rho^-}{\alpha^+\rho^+ + \alpha^-\rho^-} |u^+ - u^-|(u^+ - u^-),$$

where C is a non-dimensional number and ℓ a length scale. The pressure p is assumed to be shared by the two fluids and (173) is closed by the algebraic relations (171). Note that this system is neither hyperbolic nor conservative.

We have discussed a case, in the context of water waves, in which a separated flow can lead to a two-fluid model as far as modeling is concerned. As we have already said, two-fluid flows are very common in nature, but also in many industries like power, nuclear, chemical-process, oil-and-gas, cryogenic, space, bio-medical or micro-technology. Depending on the context, the models used for simulation are very different. However averaged models share the same structure as (173). We generalize this system by introducing viscosity and capillarity effects and write:

$$\begin{cases} \partial_t(\alpha^\pm \rho^\pm) + \operatorname{div}(\alpha^\pm \rho^\pm u^\pm) = 0, \\ \partial_t(\alpha^\pm \rho^\pm u^\pm) + \operatorname{div}(\alpha^\pm \rho^\pm u^\pm \otimes u^\pm) \\ \quad + \alpha^\pm \nabla p = \operatorname{div}(\alpha^\pm \tau^\pm) + \sigma^\pm \alpha^\pm \rho^\pm \nabla \Delta \left(\alpha^\pm \rho^\pm\right). \end{cases} \tag{174}$$

The right hand side of Eq. $(174)_2$ is the sum of

- internal viscous forces, involving the viscous symmetric stress tensor $\tau^\pm$ of each phase;
- internal capillary forces represented by the so called Korteweg model on each phase. It is written as a force involving third order derivatives of $\alpha^\pm \rho^\pm$ and constant coefficients $\sigma^\pm > 0$.

In order to close the system, three additional laws are necessary:
(i) *Viscous stress tensor.* First, the fluids are assumed to be Newtonian, so that there exists shear and bulk viscosity coefficients $\mu^\pm$ and $\lambda^\pm$ such that

$$\tau^\pm = 2\mu^\pm D(u^\pm) + \lambda^\pm \operatorname{div} u^\pm \mathrm{Id}, \tag{175}$$

where $D(u^\pm)$ denotes the strain rate tensor, given as the symmetric part of the velocity gradient $\nabla u^\pm$.
(ii) *Equations of State.* We assume that the two fluids are barotropic, that is:

$$p = p^\pm(\rho^\pm). \tag{176}$$

And sometimes, we shall use:

$$p(\rho^\pm) = A^\pm (\rho^\pm)^{\gamma^\pm} \quad \text{where } \gamma^\pm \text{ are given constants greater than 1.} \tag{177}$$

Note that the pressure law hypothesis (177) may be slightly relaxed.
(iii) *Capillary forces.* The last term in the momentum equations involves capillary effects. It is usually designed in single velocity models to account for phase transition problems where a finite thickness transition zone is assumed to separate pure phases. In the present framework, it may be viewed mathematically as a stabilizing term with respect to density oscillations and physically as internal capillary forces within each phase.

The expression introduced in equation $(174)_2$ may be rewritten in a conservative manner as the divergence of a symmetric capillary stress tensor K:

$$\sigma^\pm R^\pm \nabla \Delta R^\pm = \operatorname{div} K,$$

where

$$K_{ij} = -\sigma^{\pm}\partial_i R^{\pm}\partial_j R^{\pm} + \sigma^{\pm}\delta_{ij}\left(\Delta\frac{R^{\pm 2}}{2} - \frac{1}{2}|\nabla R^{\pm}|^2\right),$$

for $(i, j) \in \{1, 2, 3\}^2$. System (174) is also supplemented by initial conditions

$$(\alpha^{\pm}\rho^{\pm})|_{t=0} = R_0^{\pm}, \qquad (\alpha^{\pm}\rho^{\pm}u^{\pm})|_{t=0} = m_0^{\pm}. \tag{178}$$

The functions ρ_0 and m_0 are assumed to satisfy

$$\begin{aligned} &R_0^{\pm} \geq 0, \quad \alpha_0^{\pm} \in [0, 1] \quad \text{such that} \quad \alpha_0^{+} + \alpha_0^{-} = 1, \\ &\text{and} \quad \frac{|m_0^{\pm}|^2}{R_0^{\pm}} = 0 \quad \text{on } \{x \in \Omega / R_0^{\pm}(x) = 0\}. \end{aligned} \tag{179}$$

The purpose of [33] was to investigate the mathematical properties of (174). More precisely, they address the question of whether available mathematical results in the case of a single fluid governed by the compressible barotropic Navier–Stokes equations may be extended to this generic two-phase model.

Considered from a mathematical viewpoint, the evolution problem of single phase flows is still a big challenge, only partial results are currently available as far as existence and uniqueness of global solutions to the compressible Navier–Stokes equations are concerned. Some of the mathematical difficulties involve for instance:

– The mixed hyperbolic/parabolic property of the partial differential system. As a matter of fact, there is no diffusion on the mass conservation equations, whereas temperature and velocity evolve according to parabolic equations, thanks to thermal conduction and viscosity phenomena.
– The fact that the density may go to zero, see for instance [161]. In that case, singularities may prevent the existence of global solutions.
– The question of the compactness of the pressure gradient and the convection term in the momentum equation is not easy, since strong convergence of both density and velocity is required in order to approach the limit in the momentum conservation equations.
– In the case of density dependent viscosity coefficients, viscosity may vanish with density, so that the only naturally defined variable is not the velocity u alone, but rather the momentum density ρu. Such a degeneracy requires significant mathematical efforts and suitable additional forces such as drag forces or capillary effects.

Since single-phase flow models may be considered as a particular case of two-phase flow models at the limit when one of the two phases volume fraction tends to zero, extending to two-phase models the currently available results for single phase flow models is not an easy task. Additional problems make the study even more difficult. Indeed, the underlying Jacobian matrix is not always hyperbolic, i.e. it may have non-real eigenvalues, which makes a significative difference to the single-phase case. Moreover, the system cannot be expressed in a conservative form, since the pressure gradient in the momentum equation is multiplied by the volume fraction.

Let us now describe the main features of the model considered in [33]. The space domain is assumed to be a box $\Omega = \mathbf{T}^3 = \mathbf{R}^3/(2\pi\mathbf{Z})^3$ with periodic boundary conditions. The case of the whole space $\Omega = \mathbf{R}^3$ and a bounded domain Ω with suitable boundary conditions is postponed to forthcoming works adapting in particular the single phase flow techniques introduced in [39]. The paper [33] presents a well-posed result for the generic two-fluid model. It also provides a linear theory with eigenvalues analysis and give a new result regarding the matter of invariant regions. More precisely, we get the following result

THEOREM 5.1. *Assume (177) holds with $1 < \gamma^\pm < 6$ and that the initial data $(\rho_0^\pm, m_0^\pm)$ satisfy (179) and are taken in such a way that*

$$\int_\Omega \frac{|m_0^\pm|^2}{R_0^\pm}\,\mathrm{d}x < +\infty,$$

that the initial density fraction $R_0^\pm$ satisfies

$$R_0^\pm \in L^1(\Omega) \quad \text{and} \quad \nabla\sqrt{R_0^\pm} \in (L^2(\Omega))^3.$$

Then, there exists a global in time weak solution to (174)–(176) and (178).

Note the restriction $1 < \gamma^\pm < 6$ which comes from the $L^2(0, T; (L^r(\Omega))^3)$ with $r > 1$ bound required on $\alpha_n^\pm \nabla p_n$ to enable the limit of the term $\alpha_n^\pm R_n^\pm \nabla p_n^\pm$ to be reached. Also recall finally that the definition of weak solutions follows the one in [43]. For details and other mathematical results, readers are referred to [33].

5.2. *Pollutant propagation models*

(a) **Transport equation.** Shallow-water equations taking into account the non-constant density of the material are subject to investigation. Only a few results exist in this direction, see for instance [13]. They may be seen as perturbations of the known results for the standard shallow-water equations. For instance let us comment on the model studied recently in [101]. This inviscid model reads

$$\begin{cases} \partial_t(\rho h) + \partial_x(\rho h u) = 0, \\ \partial_t(\rho h u) + \partial_x(\rho h u^2 + \frac{1}{2}\beta(x)\rho h^2) = \rho h g, \end{cases} \tag{180}$$

where $\rho = h^\alpha$, $\alpha \geq 0$ being a constant, $\beta = \beta(x)$ and $g = g(v, x)$ are given functions. Here h stands for height, ρ for density and v for velocity, and the whole models an avalanche down an inclined slope.

Moreover, the author considers the following particular choices for β and g: $\beta(x) = k\cos(\gamma(x))$, $g(v, x) = \sin(\gamma(x)) - \operatorname{sign}(v)\cos(\gamma(x))\tan(\delta_F(x))$, where δ_F and γ are given functions. Here sign is the sign function, with $\text{sign}(0) = [-1, 1]$. After a change of variables, this system can be rewritten in the form

$$\frac{\partial}{\partial t} u + \frac{\partial}{\partial x} F(u) \in \tilde{G}(u, x),$$

with essentially the same structure as the system of isentropic gas dynamics in one-dimensional space (see for instance [116]), except for the fact that there is an inclusion instead of an equality. A precise (natural) definition of entropy solutions of such systems and a long-time existence theorem of such solutions, under the assumption that the initial height is bounded below by a positive constant, which corresponds to avoiding a vacuum, may be performed using well-known tools, without further difficulty.

To propose and study mathematically better shallow-water models taking into account the density variability of the material could be an interesting research area. Let us mention, for instance, a recent paper written by C. MICHOSKI and A. VASSEUR, see [133]. In this work, the authors consider a barotropic system with the flow driven by a pressure p that depends on the density and mass fraction of each chemical/phase component of the system. In a shallow-water framework, it could be written

$$\begin{cases} \partial_t(\rho h) + \partial_x(\rho h u) = 0, \\ \partial_t(h) + \partial_x(hu) = 0, \\ \partial_t(hu) + \partial_x(hu^2) + \partial_x p(h, \rho) - \partial_x(\nu(h, \rho)\partial_x u) = 0. \end{cases} \tag{181}$$

Note that here the viscosity is a function of the pressure p. Recent works by J. MALEK, K. RAJAGOPAL et al. provide results on these types of viscosity laws in an incompressible framework.

(b) **Kazhikhov–Smagulov type models.** A model to describe convective overturning of a fluid layer due to density differences is derived, based on equations of KAZHIKHOV–SMAGULOV, see for instance [90]. Such a system has also been recently used for avalanche simulation, see [83]. From a mathematical view-point, the system is given by

$$\begin{cases} \partial_t \rho + \operatorname{div}(\rho u) = 0, \\ \partial_t(\rho u) + \operatorname{div}(\rho u \otimes u) + \nabla p + \text{Diff} = \rho f, \\ u = v - \alpha \nabla \log \rho, \qquad \operatorname{div} v = 0, \end{cases} \tag{182}$$

where Diff is a diffusion term.

REMARK. More general systems exist, for instance assuming

$$u = v \mp \alpha \nabla \varphi_{\pm}(\rho), \quad \operatorname{div} v = 0,$$

where

$$\rho \nabla \varphi_{\pm}(\rho) = \pm \nabla \Psi(\rho)$$

with $\Psi(\rho) > 0$ and $\varphi_{\pm}$, respectively, for increasing or decreasing functions. Using $\varphi_- = 1/\rho$ provides the low Mach combustion system studied by E. EMBID, A MAJDA. Using $\varphi_+ = \log\rho$ provides the pollutant model studied by A. KAZHIKHOV, H. BEIRAO-DA VEIGA, P. SECCHI and P.-L. LIONS.

Note that such systems may be obtained from the full compressible Navier–Stokes equations for heat-conducting fluids. They correspond to low Mach (low Froude) number limit, see for instance [5,6].

Assuming Diff $= -\mu\Delta u$, periodic boundary conditions and changing the variable $u = v - \alpha\nabla\log\rho$, the α^2 term coming from the non-linear term does not have a gradient form. The new system is given by

$$\begin{cases} \partial_t\rho + \operatorname{div}(\rho v) - \alpha\Delta\rho = 0, \\ \rho\partial_t v + \rho v\cdot\nabla v - \alpha\nabla\rho\cdot\nabla v - \lambda v\cdot\nabla\nabla\rho + \nabla P \\ \quad -\mu\Delta v + \alpha^2\nabla\rho\cdot\nabla\left(\dfrac{\nabla\rho}{\rho}\right) = \rho f, \\ \operatorname{div} v = 0. \end{cases} \tag{183}$$

This implies local existence results and no global existence without data restriction. Even if we neglect the α^2 term, the system reads

$$\begin{cases} \partial_t\rho + \operatorname{div}(\rho v) = \alpha\Delta\rho, \\ \partial_t(\rho v) + \operatorname{div}(\rho v\otimes v) - \mu\Delta v + \nabla p \\ \quad +\alpha\nabla v\cdot\nabla\rho - \alpha v\cdot\nabla\nabla\rho = \rho f, \\ \operatorname{div} v = 0. \end{cases} \tag{184}$$

In this case, let us assume $f \in L^2(0,T;(L^2(\Omega))^d)$ and the initial data ρ_0 belonging to $H^1(\Omega)$ and $v_0 \in (L^2(\Omega))^d$ with for a.a. $x\in\Omega$

$$0 < m \le \rho_0(x) \le M < +\infty.$$

In the book written by S.N. ANTONSEV, A.V. KAHIKHOV and V.N. MONAKOV [11], they get the following result

THEOREM 5.2. *Let us assume data satisfying the above properties. Let α, μ, M and m be such that*

$$\mu - \alpha\frac{M-m}{2} > 0.$$

Then there exists a global weak solution of (184) such that

$$v \in L^\infty(0,T;(L^2(\Omega))^d) \cap L^2(0,T;(H^1(\Omega))^d),$$
$$\rho \in L^\infty(0,T;H^1(\Omega)) \cap L^2(0,T;H^2(\Omega)).$$

Using a diffusive term which is "well" prepared, for example in the pollutant case, Diff $= -\alpha\operatorname{div}(hD(u))$, we get, see [44], a global existence of weak solutions with no restriction on the data. The power two (namely in α^2) bad term enters in the gradient of the pressure and the diffusive term is added to the power, one extra term (namely α terms). More precisely we have to study the following system

$$\begin{cases} \partial_t \rho + \operatorname{div}(\rho v) = \alpha \Delta \rho, \\ \partial_t(\rho v) + \operatorname{div}(\rho v \otimes v) - \alpha \operatorname{div}(\rho D(v)) + \nabla p \\ \quad + \alpha \nabla v \cdot \nabla \rho - \alpha v \cdot \nabla \nabla \rho = \rho f, \\ \operatorname{div} v = 0. \end{cases} \tag{185}$$

We get the following global existence result

THEOREM 5.3. *Let us assume data satisfying the above properties. Then there exists a global weak solution of* (185) *such that*

$$v \in L^\infty(0, T; (L^2(\Omega))^d) \cap L^2(0, T; (H^1(\Omega))^d),$$
$$\rho \in L^\infty(0, T; H^1(\Omega)) \cap L^2(0, T; H^2(\Omega)).$$

This contrasts with the results of H. BEIRAO DA VEIGA, E. EMBID, A. KAZHIKHOV, A. MAJDA, P. SECCHI and P.-L. LIONS, where they try to get results with a general viscosity $\mu = \mu(\rho)$.

REMARK. All the calculations made previously concern the periodic or whole space cases since they are based on a change of variable, comparing u and v. The study of a bounded domain is an interesting open mathematical problem. Note that this concerns also Fick's law type derivations in a bounded domain.

5.3. *Sedimentation*

Phenomena related to sediment transport are of great interest as they affect human life. The analysis of sediment transport is important to predict and prevent natural disasters. For this purpose, many physical and mathematical models are proposed in the literature in order to predict evolution of the bed and the changes in water regime when such unsteady flows occur. Among the mathematical models, that most often used is based on the Saint-Venant-Exner equations. This model couples a hydrodynamic Saint-Venant (shallow-water model) system to a morpho-dynamic bed-load transport sediment equation as follows

$$\begin{cases} \partial_t h + \operatorname{div}(hv) = 0, \\ \partial_t(hv) + \operatorname{div}(hv \otimes v) + h \dfrac{\nabla(h + z_b)}{\mathrm{Fr}^2} = 0, \\ \partial_t z_b + \xi \operatorname{div}(q_b(h, q)) = 0, \end{cases} \tag{186}$$

where z_b is the movable bed thickness, $\xi = 1 + (1 - \psi_0)$ with ψ_0 the porosity of the sediment layer, and q_0 denotes the solid transport flux or sediment discharge. It depends on the height h of the fluid and the water discharge $q = hv$, where v is the velocity. For the solid transport flux q_b, there exist several formulae in the literature: The Grass equation, the Meyer-Peter and Muller equation, or the formulas of NIELSEN, FERNÁNDEZ LUQUE and VAN BEEK. All of them are obtained using empirical methods. The most basic sediment model is the Grass equation, where the sediment movement begins at the same time as the fluid motion. In this case the solid transport is given by

$$q_b(h,q) = A_g|v|^{m_g}v, \qquad 0 \le m_g \le 3,$$

where the constant A_g includes the effects due to the grain and kinematic viscosity. Note that viscous sediment models exist in the mathematical and physical literature. Recently the following system has been studied

$$\begin{cases} \partial_t h + \operatorname{div}(hv) = 0, \\ \partial_t(hv) + \operatorname{div}(hv \otimes v) + h\dfrac{\nabla(h+z_b)}{\mathrm{Fr}^2} - A\operatorname{div}(hD(v)) = 0, \\ \partial_t z_b + A\operatorname{div}(h|v|^k v) - \dfrac{A}{2}\Delta z_b = 0, \end{cases}$$

with the initial conditions

$$h|_{t=0} = h_0 \ge 0, \qquad z_b|_{t=0} = z_{b0}, \qquad hv|_{t=0} = m_0.$$

What is really interesting is the MELLET-VASSEUR's multiplier (namely $|u|^\delta u$ for some δ) which plays an important role in the mathematical study, see [163]. We refer also to the book [60] for more information on sedimentation and to [26] for erosion studies

Acknowledgments

The author thanks referees for providing constructive comments and help in improving the contents of this handbook.

References

[1] S. Agmon, A. Douglis, L. Nirenberg, *Estimates near the boundary for solutions of elliptic partial differential equations I*, Comm. Pures Appl. Math. **12** (1959) 623–727.
[2] S. Agmon, A. Douglis, L. Nirenberg, *Estimates near the boundary for solutions of elliptic partial differential equations II*, Comm. Pures Appl. Math. **17** (1964) 35–92.
[3] G. Allain, *Un problème de Navier–Stokes avec surface libre et tension superficielle*, Ann. Fac. Sci. Toulouse. Sér. 5 **7** (1) (1985) 29–56.
[4] G. Allaire, *Homogenization and two-scale convergence*, SIAM J. Math. Anal. **23** (6) (1992) 1482–1518.
[5] Th. Alazard, *Low Mach number limit of the full Navier–Stokes equations*, Arch. Ration. Mech. Anal. **180** (1) (2006) 1–73.
[6] Th. Alazard, *Low Mach number flows and combustion*, SIAM J. Math. Anal. **38** (4) (2006) 1186–1213.
[7] B. Alvarez-Samaniego, D. Lannes, *Large time existence for* 3*D water-waves and asymptotics*, Invent. Math. **171** (2008) 485–541.
[8] B. Alvarez-Samaniego, D. Lannes, *A Nash–Moser theorem for singular evolution equations. Application to the Serre and Green–Naghdi equations*, Indiana Univ. J. **57** (2008) 97–132.

[9] Y. Amirat, D. Bresch, J. Lemoine, J. Simon, *Effect of rugosity on a flow governed by stationary Navier–Stokes equations*, Quart. Appl. Math. **59** (4) (2001) 769–785.
[10] C. Ancey, *Dynamique des avalanches*, Presses polytechniques et universités romandes, 2006.
[11] S.N. Antonsev, A.V. Kazhikhov, V.N. Monakov, *Boundary Value Problems in Mechanics of Nonhomogeneous Fluids*, North-Holland, 1990.
[12] E. Audusse, *Multilayer Saint–Venant system: Derivation and numerical validation*, DCDS-B, 5 (2) (2005) 189–214 .
[13] E. Audusse, M.-O. Bristeau, *Transport of pollutant in shallow water flows: A two time steps kinetic method*, M2AN, 37 (2) (2003) 389–416 .
[14] M. Avellaneda, F.-H. Lin, *Compactness methods in the theory of homogeneization*, Comm. Pure Appl. Math. **40** (1987) 803–847.
[15] V. Barcilon, P. Constantin, E.S. Titi, *Existence of solutions to the Stommel-Charney model of the gulf stream*, SIAM J. Math. Anal. **19** (1988) 1355–1364.
[16] A. Basson, D. Gérard-Varet, *Wall laws for fluid flows at a boundary with random roughness*, Comm. Pure Appl. Math. (2008).
[17] J.T. Beale, *The initial value problem for the Navier–Stokes equations with a free surface*, Comm. Pure Appl. Math. **34** (3) (1981) 359–392.
[18] J.T. Beale, *Large-time regularity of viscous surface waves*, Arch. Rational. Mech. Anal. **84** (1984) 307–352.
[19] S. Benzoni, D. Serre, *Multi-dimensional Hyperbolic Partial Differential Equations: First-order Systems and Applications*, Oxford University Press, 2007.
[20] C. Bernardi, O. Pironneau, *On the shallow water equations at low Reynolds number*, Comm. Partial Differential Equations **16** (1991) 59–104.
[21] A.L. Bertozzi, M. Pugh, *The lubrication approximation for thin viscous films: Regularity and long time behaviour of weak solutions*, Comm. Pure Appl. Math. **49** (1994) 85–123.
[22] É. Blayo, J. Verron, *The no-slip condition and separation of western boundary currents*, J. Phys. Oceanography **26** (1995) 1938–1951.
[23] P. Bolley, J. Camus, G. Métivier, *Estimations de Schauder et régularité Hölderienne pour une classe de problèmes aux limites singuliers*, Comm. Partial Differential Equations **11** (1986) 1135–1203.
[24] J. Bona, D. Lannes, J.-C. Saut, *Asymptotic models for internal waves*, J. Math. Pures Appl. **9** (2008) 538–566.
[25] F. Bouchut, A. Mangeney, B. Perthame, J.P. Vilotte, *A new model of Saint Venant and Savage-Hutter type for gravity driven shallow water flows*, C.R. Acad. Sci. Paris, Sér. I **336** (6) (2003) 531–536.
[26] F. Bouchut, E.D. Fernández-Nieto, A. Mangeney, P.-Y. Lagrée, *On new erosion models of Savage-Hutter type for avalanches*, Acta Mech. **199** (2008) 181–208.
[27] F. Bouchut, M. Westdickenberg, *Gravity driven shallow water models for arbitrary topography*, Comm. Math. Sci. **2** (3) (2004) 359–389.
[28] M. Boutounet, L. Chupin, P. Noble, J.-P. Vila, *Shallow water flows for arbitrary topography*, Comm. Math. Sci. **6** (1) (2008) 531–536.
[29] K. Bryan, *A numerical investigation of non-linear model of a wind driven ocean*, Ocean. J. Atmos **20** (1963) 594–606.
[30] D. Bresch, *On the compressible Navier–Stokes equations*, in: Morningiside Notes on the Analysis of PDE, World Scientific Press, 2008 (in preparation) .
[31] D. Bresch, D. Bucur, V. Milisic, *Multi-scale wall laws*, 2008 (in progress) .
[32] D. Bresch, T. Colin, *Some remarks on the derivation of the Sverdrup relation*, J. Math. Fluid. Mech. **4** (2) (2002) 95–108.
[33] D. Bresch, B. Desjardins, J.-M. Ghidaglia, E. Grenier, *Mathematical results for a generic two-fluid model*, Arch Rational Mech. Anal. (in press) .
[34] D. Bresch, B. Desjardins, *On the existence of global weak solutions to the Navier–Stokes equations for viscous compressible and heat conducting fluids*, J. Math. Pures Appl. **1** (87) (2007) 57–90.
[35] D. Bresch, B. Desjardins, *Stabilité de solutions faibles pour les équations de Navier–Stokes compressibles avec conductivité de chaleur*, C.R. Acad. Sci. Paris, Section Mathématiques **343** (3) (2006) 219–224.
[36] D. Bresch, B. Desjardins, *Some diffusive capillary models of Korteweg type*, C. R. Acad. Sci., Paris, Section Mécanique **332** (11) (2004) 881–886.
[37] D. Bresch, B. Desjardins, *Existence of global weak solutions for a 2D viscous shallow water equations and convergence to the quasigeostrophic model*, Commun. Math. Phys. **238** (1–2) (2003) 211–223.

[38] D. Bresch, B. Desjardins, *On the construction of approximate solutions for the 2d viscous shallow water model and for compressible Navier–Stokes models*, J. Math. Pures Appl. **4** (86) (2006) 262–268.
[39] D. Bresch, B. Desjardins, D. Gérard-Varet, *On compressible Navier–Stokes equations with density dependent viscosities in bounded domains*, J. Math. Pures Appl. **2** (87) (2007) 227–235.
[40] D. Bresch, B. Desjardins, D. Gérard-Varet, *Rotating fluids in a cylinder*, Discrete Cont. Dynam. Syst. Ser. B **3** (3) (2003) 469–477.
[41] D. Bresch, M. Gisclon, C.K. Lin, *An example of low Mach (Froude) number effects for compressible flows with nonconstant density (height) limit*, M2AN, 39 (3) (2005) 477–486 .
[42] D. Bresch, A. Juengel, H.-L. Li, Z.P. Xin, *Effective viscosity and dispersion (capillarity) approximations to hydrodynamics*, 2008 (in progress) .
[43] D. Bresch, B. Desjardins, C.K. Lin, *On some compressible fluid models: Korteweg, lubrication and shallow water systems*, Comm. Partial Differential Equations **28** (3–4) (2003) 1009–1037.
[44] D. Bresch, E.H. Essoufi, *Effect of density dependent viscosities on multiphasic incompressible fluid models*, J. Math. Fluid Mech. **9** (3) (2007) 377–397.
[45] D. Bresch, D. Gérard-Varet, *On some homogeneization problems from the shallow water theory*, Appl. Math. Lett. **20** (5) (2007) 505–510.
[46] D. Bresch, D. Gérard-Varet, *Roughness Induced Effects in Geophysics*, in: School and Workshop on Oceanography, Lakes and Rivers, Centro international de Matematica 25, 2004.
[47] D. Bresch, D. Gérard-Varet, *About roughness-induced effects on the quasigeostrophic model*, Comm. Math. Phys. **253** (1) (2005) 179–200.
[48] D. Bresch, F. Guillen-Gonzalez, M.A. Rodriguez-Bellido, *A corrector for the Sverdrup solution for a domain with Islands*, Appl. Anal. **83** (3) (2004) 217–230.
[49] D. Bresch, R. Klein, C. Lucas, *Multiscale asymptotics justifications around shallow-water systems*, 2008 (in progress) .
[50] D. Bresch, J. Lemoine, J. Simon, *A geostrophic model with vertical diffusion*, Nonlinear Anal. TMA **43** (2001) 449–470.
[51] D. Bresch, A.V. Kazhikhov, J. Lemoine, *On the two-dimensional hydrostatic Navier–Stokes equations*, SIAM J. Math. Anal. **36** (3) (2004) 796–814.
[52] D. Bresch, G. Métivier, *Global existence and uniqueness for the lake equations with vanishing topography: Elliptic estimates for degenerate equations*, Nonlinearity **19** (3) (2006) 591–610.
[53] D. Bresch, V. Milisic, *High order multiscale wall laws, part I: The periodic case,* Quart. Appl. Math. (2008) (in press) .
[54] D. Bresch, P. Noble, *Mathematical justification of a shallow-water model*, Methods Appl. Anal. **14** (2) (2007) 87–117.
[55] D. Bresch, J. Simon, *Western boundary currents versus vanishing depth*, Discrete Cont. Dynam. Syst. Ser. B **3** (3) (2003) 469–477.
[56] D. Bucur, E. Feireisl, Š. Nečasová, J. Wolf, *On the asymptotic limit of the Navier–Stokes system with rough boundaries*, J. Differential Equations **244** (11) (2008) 2890–2908.
[57] D. Bucur, E. Feireisl, Š. Nečasová, *On the asymptotic limit of flows past a ribbed boundary*, J. Math. Fluid Mech. **10** (4) (2008) 554–568.
[58] D. Bucur, E. Feireisl, *Influence of wall roughness on the slip behavior of viscous fluids*, Proc. Royal Soc. Edinburgh Sec. A. **138** (5) (2008) 957–973.
[59] A.T. Bui, *Existence and uniqueness of a classical solution of an initial boundary value problem of the theory of shallow waters*, SIAM J. Math. Anal. **12** (1981) 229–241.
[60] M.C. Bustos, F. Concha, R. Burger, E. Tory, *Sedimentation and Thickening Phenomenological Foundation and Mathematical Theory*, Kluwer Academic publishers, 1999.
[61] R. Camassa, D.L. Holm, C.D. Levermore, *Long time shallow water equations with a varying bottom*, J. Fluid Mech. (1997) 173–189.
[62] R. Camassa, D.L. Holm, C.D. Levermore, *Long time effects of the bottom topography in shallow water*, Physica D (1996) 258–286.
[63] G.F. Carrier, *Singular perturbation theory and geophysics*, SIAM Rev. **12** (1970) 175–193.
[64] J. Casado-Diaz, E. Fernández-Cara, J. Simon, *Why viscous fluids adhere to rugose walls (a mathematical explanation)*, J. Differential Equations **189** (2003) 526–537.
[65] F.-J. Chatelon, P. Orenga, *Some smoothness and uniqueness for a shallow water problem*, Adv. Diff. Eqs. **3** (1) (1998) 155–176.

[66] J.-Y. Chemin, B. Desjardins, I. Gallagher, E. Grenier, *Basics of Mathematical Geophysics*, Oxford University Press, 2006.
[67] I.D. Chueshov, G. Raugel, A.M. Rekalo, *Interface boundary value problem for the Navier–Stokes equations in thin two-layers domains*, J. Differential Equations **208** (2) (2005) 449–493.
[68] P. Cessi, *Laminar separation of colliding western boundary currents*, J. Marine Res. **49** (1991) 697–717.
[69] T. Colin, *Remarks on a homogeneous model of ocean circulation*, Asymptot. Anal. **12** (1996) 153–168.
[70] W. Craig, P. Guyenne, H. Kalisch, *Hamiltonian long-wave expansions for free surfaces and interfaces*, Comm. Pure Appl. Math. **58** (12) (2005) 1587–1641.
[71] W. Craig, P. Guyenne, D. Nicholls, C. Sulem, *Hamiltonian long wave expansions for water waves over a rough bottom*, Proc. Royal Soc. A **461** (2005) 1–35.
[72] J. De Dieu Zabonsoré, G. Narbona, *Existence of global weak solutions for 2D viscous bi-layer Shallow Water model*, 2008 (in press).
[73] B. Desjardins, E. Grenier, P.-L. Lions, N. Masmoudi, *Incompressible limit for solutions of the isentropic Navier–Stokes equations with Dirichlet boundary conditions*, J. Math. Pures Appl. **78** (9) (1999) 461–471.
[74] J.-M. Delort, *Temps d'existence pour l'équation de Klein–Gordon semi-linaire à données petites faiblement décroissantes,* in: Séminaire: Équations aux Dérivées Partielles, 1999–2000, ExpNo. II, 18 pp., Sémin. équ. Dériv. Partielles, École Polytech., Palaiseau, 2000 .
[75] J. Deng, *The problem of Gulf Stream separation: A barotropic approach*, J. Phys. Oceanogr. **23** (1992) 2182–2200.
[76] De Saint Venant, *Théorie du mouvement non permanent des eaux, avec application aux crues des rivières et l'introduction des marées dans leur lit*, C. R. Acadi. Sci. Paris **LXXIII** (1871) 147–154.
[77] L. Desvillettes, *Some aspects of the modeling at different scales for multiphase flows*, in: A. Ibrahimbegovic, F. Dias, H. Matthies, P. Wriggers (Eds.), ECCOMAS Thematic Conference on Multi-Scale Computational Methids for Solids and Fluids, Cachan, France, November 28–30, 2007 .
[78] B. Desjardins, E. Grenier, *On the homogeneous model of wind-driven ocean circulation*, SIAM J. Appl. Math. **60** (1999) 43–60.
[79] B. Di Martino, P. Orenga, M. Peybernes, *On a bi-layer Shallow Water Problem with the Rigid-Lid Hypothesis*, in: Mathematical Models and Methods in Applied Sciences, vol. 15, 2005, pp. 843–869.
[80] R.J. DiPerna, *Convergence of the viscosity method for isentropic gas dynamics*, Comm. Math. Phys. **91** (1983) 1–30.
[81] A. Dutrifoy, A.J. Majda, *Fast wave averaging for the equatorial shallow-water equations*, Comm. Partial Differential Equations **32** (10) (2007) 1617–1642.
[82] A. Dutrifoy, A.J. Majda, S. Schochet, *A simple justification of the singular limit for equatorial shallow-water dynamics*, Comm. Pure Appl. Math. **62** (3) (2008) 322–333.
[83] J. Etienne, E. Hopfinger, P. Saramito, *Numerical simulations of high density ratio lock-exchange flows*, Phys. Fluids **17** (3) (2005).
[84] T. Erneux, S.H. Davis, *Nonlinear rupture of free films*, Phys. Fluids A, Fluid Dynamics **5** (1993) 1117–1122.
[85] E. Feireisl, *Dynamics of Viscous Compressible Fluids*, Oxford Science Publication, Oxford, 2004.
[86] E. Feireisl, J. Málek, A. Novotny, I. Straškraba, *Anelastic approximation as a singular limit of the compressible Navier–Stokes system*, Comm. Partial Differential Equations **3** (1) (2008) 157–176.
[87] E. Feireisl, A. Novotny, H. Petzeltová, *On the existence of globally defined weak solutions to the Navier–Stokes equations of compressible isentropic fluids*, J. Math. Fluid. Dynam. **3** (2001) 358–392.
[88] E.D. Fernández-Nieto, F. Bouchut, D. Bresch, M. Castro, A. Mangeney, *A New Savage-Hutter model for submarine avalanches and generated tsunamis*, J. Comput. Phys **227** (16) (2008) 7720–7754.
[89] F. Flori, P. Orenga, M. Peybernes, *Sur un problème de shallow water bicouche avec conditions aux limites de Dirichlet*, C. R. Acad. Sci. **340** (10) (2005) 719–724.
[90] F. Franchi, B. Straughan, *A comparison of the Graffi and Kazhikhov–Smagulov models for top heavy pollution instability*, Adv. Water Resources **24** (6) (2001) 585–594.
[91] G.P. Galdi, *An introduction to the mathematical theory of the Navier–Stokes equations.*, in: Vol. II, Nonlinear Steady Problems, in: Springer Tracts in Natural Philosophy, vol. 39, Springer-Verlag, New York, 1994.
[92] I. Gallagher, L. Saint-Raymond, *About the Influence of Earth's Rotation on Geophysical Flows*, in: S. Friedlander, D. Serre (Eds.), Handbook of Mathematical Fluid Dynamics, vol. 4, 2007, pp. 201–329.
[93] P.R. Gent, *The energetically consistent shallow water equations*, J. Atmos. Sci. **50** (1993) 1323–1325.
[94] D. Gérard-Varet, *Highly rotating fluids in rough domains*, J. Math. Pures Appl. (9) **82** (11) (2003) 1453–1498.

[95] D. Gérard-Varet, *The Navier wall law at a boundary with random roughness*, Comm. Math. Phys **286** (2009) 81110.
[96] J.-F. Gerbeau, B. Perthame, *Derivation of viscous saint-venant system for laminar shallow water; numerical validation*, Discrete Contin. Dynam. Syst., Ser. B **1** (1) (2001) 89–102.
[97] L. Giacomelli, F. Otto, *Rigorous lubrication approximation*, Interfaces Free Bound. **5** (2003) 483–529.
[98] V. Girault, A. Raviart, *Finite element approximations of the Navier–Stokes equations*, in: Lecture Notes in Mathematics, Springer-Verlag, New–York, 1981.
[99] C. Goulaouic, N. Shimakura, *Régularité Höldérienne de certains problèmes aux limites dégénérés*, Ann. Scuola Norm. Sup. Pisa **10** (1983) 79–108.
[100] Z. Guo, Q. Jiu, Z.P. Xin, *Spherically symmetric isentropic compressible flows with density-dependent viscosity coefficients*, SIAM J. Math. Anal (2008) 1402–1427.
[101] P. Gwiazda, *An existence result for a model of granular material with non-constant density*, Asymptot. Anal. **30** (2002) 43–80.
[102] B. Haspot, *Cauchy problem for viscous shallow water equations with a term of capillarity*, 2007 (submitted for publication) .
[103] D. Hoff, D. Serre, *The failure of continuous dependence on initial data for the Navier–Stokes equations of compressible flow*, SIAM J. Appl. Math. **51** (4) (1991) 887–898.
[104] M. Ishii, *Thermo-fluid Dynamic Theory of Two-Phase Flows*, Eyrolles, Paris, 1975.
[105] D. Jiang, P. Zhang, *On spherically symetric solutions of the compressible isentropic Navier–Stokes equations*, Comm. Math. Phys. **215** (2001) 559–581.
[106] Q. Jiu, D. Niu, *Vanishing viscous limits for the 2D lake equations with Navier boundary conditions*, J. Math. Anal. Appl. **338** (2) (2008) 1070–1080.
[107] A.V. Kazhikhov, *Initial-boundary value problems for the Euler equations of an ideal incompressible fluid*, Moscow Univ. Math. Bull. **46** (5) (1991) 10–14.
[108] P.E. Kloeden, *Global existence of classical solutions in the dissipative shallow water equations*, SIAM J. Math. Anal. **16** (1985) 301–315.
[109] J. Leray, *Sur le mouvement d'un liquide visqueux emplissant l'espace*, Acta Math. **63** (1934) 193–248.
[110] C.D. Levermore, M. Oliver, E.S. Titi, *Global well-posedness for models of shallow water in a basin with a varying bottom*, Indiana Univ. Math. J. **45** (1996) 479–510.
[111] C.D. Levermore, M. Sammartino, *A shallow water model in a basin with varying bottom topography and eddy viscosity*, Nonlinearity **14** (6) (2001) 1493–1515.
[112] J. Li, Z.P. Xin, *Some uniform estimates and blowup behavior of global strong solutions to the stokes approximation equations for two-dimensional compressible flows*, J. Differential Equations (2005) (in press) .
[113] H.-L. Li, J. Li, Z.P. Xin, *Vanishing of vacuum states and blowup phenomena of the compressible Navier–Stokes equations*, Comm. Math. Phys. **281** (2008) 401–444.
[114] J.-L. Lions, *Perturbations singulières dans les problèmes aux limites et en contrôle optimal. Vol. 323*, Springer-Verlag, Berlin, 1973.
[115] P.-L. Lions, *Mathematical Topics in Fluid Dynamics, Vol. 2*, in: Compressible Models, Oxford Science Publication, Oxford, 1998.
[116] P.-L. Lions, B. Perthame, P.E. Souganidis, *Existence of entropy solutions to isentropic gas dynamics System*, Comm. Pure Appl. Math. **49** (6) (1996) 599–638.
[117] P.-L. Lions, N. Masmoudi, *Homogeneization of the Euler system in a 2D porous medium*, J. Math. Pures Appl. **9** (1) (2005) 1–20.
[118] P.-L. Lions, N. Masmoudi, *Une approche locale de la limite incompressible*, C.R. Acad. Sci. Paris **329** (1999) 387–392.
[119] J.-L. Lions, R. Temam, S. Wang, *New formulations of the primitive equations of atmosphere and applications*, Nonlinearity (5) (1992) 237–288.
[120] R. Liska, L. Margolin, B. Wendroff, *On nonhydrostatic two layer models of incompressible flow*, Comput. Math. Appl. **29** (9) (1995) 25–37.
[121] R. Liska, B. Wendroff, *Anaysis and comparison with stratified fluid models*, J. Comput. Phys. **137** (4) (1997) 212–244.
[122] C. Lucas, *Cosine effect on a viscous Shallow Water model and its quasigeostrophic and lake equations limits*, C.R. Acad. Sci. **345** (6) (2007) 313–318.
[123] C. Lucas, *Effets de petites échelles, du tenseur des contraintes, des conditions au fond et à la surface sur les équations de Saint-Venant*, Ph.D. Thesis, Université Joseph Fourier, 2007 .

[124] C. Mascai, F. Rousset, *Asymptotic stability of steady-states for Saint-Venant equations with real viscosity*, in: C. Calgaro, J.-F. Coulombel, Th. Goudon (Eds.), Analysis and Simulation of Fluid Dynamics, in: Adv. in Math. Fluid Mechanics, 2006, pp. 155–162.

[125] A.J. Majda, *Introduction to PDEs and waves for the atmosphere and ocean*, in: Courant Lecture Notes in Mathematics, 2003.

[126] A.J. Majda, A.L. Bertozzi, *Vorticity and Incompressible Flows*, in: Cambdrige Texts in Applied Mathematics, vol. 27, Cambridge University Press, Cambridge, 2002.

[127] F. Marche, *Derivation of a new two-dimensional shallow water model with varying topography, bottom friction and capillary effects*, Eur. J. Mech. B Fluids **26** (1) (2007) 49–63.

[128] N. Masmoudi, *Rigorous derivation of the anelastic approximation*, J. Math Pures et Appliquées **88** (3) (2007) 230–240.

[129] N. Masmoudi, *Examples of Singular Limits in Hydrodynamics*, in: C.M. Dafermos, E. Feireisl (Eds.), Handbook of Differential Equations. Evolutionary Equations, vol. 3, North-Holland, 2006, pp. 197–269, (Chapter 3) 652 pages.

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

[131] A. Mellet, A. Vasseur, *On the barotropic compressible Navier–Stokes equations*, Comm. Partial Differential Equations **32** (1–3) (2007) 431–452.

[132] A. Mellet, A. Vasseur, *Existence and Uniqueness of global strong solutions for one-dimensional compressible Navier–Stokes equations*, SIAM J. Math. Anal. **39** (4) (2007–2008) 1344–1365.

[133] C. Michoski, A. Vasseur, *Existence and uniqueness of strong solutions for a compressible multiphase Navier–Stokes miscible fluid-flow problem in dimensionn* $= 1$, Math. Models Methods Appl. Sci. (in press) .

[134] M.L. Munoz-Ruiz, F.-J. Chatelon, P. Orenga, *On a bi-layer shallow-water problem*, Nonlinear Anal. Real World Appl. **4** (1) (2003) 139–171.

[135] S. Nazarov, *Dirichlet problem in an angular domain with rapidly oscillating boundary: Modelin and solution asymptotics*, S. Petersburgh Math. J. **19** (2007) 1–35.

[136] N. Neuss, M. Neuss-Radu, A. Mikelić, *Effective laws for the Poisson equation on domains with curved oscillating boundaries*, Appl. Anal. **85** (5) (2006) 479–502.

[137] G. Nguetseng, *A general convergence result for a functional related to the theory of homogenization*, SIAM J. Math. Anal. **20** (3) (1989) 608–623.

[138] T. Nishida, Y. Teramoto, H.A. Win, *Navier–Stokes flows down an inclined plane: Downward periodic motion*, J. Math. Kyoto Univ. **33-3** (1993) 787–801.

[139] A. Novotny, I. Straškraba, *Introduction to the mathematical theory of compressible flow*, in: Oxford lecture series in Mathematics and its applications, 2004.

[140] K. Ohi, T. Iguchi, *A two-phase problem for capillary-gravity waves and the Benjamin-Ono equation*, Discrete Contin. Dynam. Syst. **23** (2009) 1201–1236.

[141] M. Oliver, *Justification of the shallow water limit for a rigid lid flow with bottom topography*, Theoret. Comput. Fluid Dynam. **9** (1997) 311–324.

[142] P. Orenga, *Un théorème d'existence de solutions d'un problème de shallow water*, Arch. Rational Mech. Anal. **130** (1995) 183–204.

[143] A. Oron, S.H. Davis, S.G. Bankoff, *Long-scale evolution of thin liquid films*, Rev. Modern Phys. **9** (3) (1997).

[144] F. Otto, *Lubrication approximation with prescribed non-zero contact angle*, Comm. Partial Differential Equations **23** (1998) 2077–2164.

[145] J. Pedlosky, *Geophysical Fluid Dynamics*, Springer-Verlag, 1979.

[146] J. Pedlosky, *Ocean Circulation Theory*, Springer-Verlag, 1996.

[147] S.P. Pudasaini, K. Hutter, *Avalanche dynamics*, in: Dynamics of Rapid Flows of Dense Granular Avalanches, Springer-Verlag, 2007.

[148] D. Ropp, *A numerical study of shallow water models with variable topography*, Ph.D. Thesis, University of Arizona, 2000 .

[149] J. Sainte-Marie, M.-O. Bristeau, *Derivation of a non-hydrostatic shallow-water model; comparison with Saint–Venant and Boussinesq systems*, 2008 (submitted for publication) .

[150] S. Schochet, *The mathematical theory of low Mach number flows*, M2AN Math. Model. Numer. Anal. **39** (3) (2005) 441–458.

[151] L. Sundbye, *Existence for the Cauchy problem for the viscous shallow water equations*, Rocky Mountain J. Math. **28** (3) (1998) 1135–1152.

[152] L. Sundbye, *Global existence for Dirichlet problem for the viscous shallow water equations*, J. Math. Anal. Appl. **202** (1996) 236–258.

[153] R. Temam, M. Ziane, *Some mathematical problems in geophysical fluid dynamics*, in: S. Friedlander, D. Serre (Eds.), Handbook of Mathematical Fluid Dynamics, vol. III, 2004, pp. 535–657.

[154] R. Temam, M. Ziane, *Navier–Stokes equations in three-dimensional thin domains with various boundary conditions*, Adv. Differential Equations (4) (1996) 499–546.

[155] Y. Teramoto, *On the Navier–Stokes flows down an inclined plane*, J. Math. Kyoto Univ. **32-3** (1992) 593–619.

[156] H. Uecker, *Self-similar decay of spatially localized perturbations of the Nusselt solutions for the inclined flim problem*, Arch. Rational Mech. Anal. **184** (2007) 401–447.

[157] J. Vanneste, *Enhanced dissipation for quasigeostrophic motion over small-scale topography*, J. Fluid. Mech. **407** (2000) 105–122.

[158] J.-P. Vila, *Two moments closure equations of shallow-water type for thin film laminar flow gravity driven*, Paper in preparation and private communication, 2007 .

[159] C. Vreugdenhill, *Two layers shallow water flow in two dimensions. A numerical study*, J. Comput. Phys. **33** (1979) 169–184.

[160] W. Wang, C.-J. Xu, *The Cauchy problem for viscous shallow water equations*, Rev. Mat. Iberoamericana **21** (1) (2005) 1–24.

[161] Z.P. Xin, *Blow of smooth solutions to the compressible Navier–Stokes equation with compact density*, Comm. Pure Appl. Math. **51** (3) (1998) 229–240.

[162] V.I. Yudovich, *The flow of a perfect, incompressible liquid through a given region*, Soviet Phys. Dokl. **7** (1962) 789–791.

[163] J.D.D. Zabsonré, C. Lucas, E.D. Fernández-Nieto, *An energetically consistent viscous sedimentation model*, Math. Models Methods Appl. Sci. (M3AS) **19** (3) (2009) 477–499.

[164] E. Zadrzyńska, *Free boundary problems for non-stationary Navier–Stokes equations*, Dissertationes Matematicae, Warszawa, 2004.

[165] V. Zeitlin, F. Bouchut, S. Medvedev, G. Reznik, A. Stegner, *Nonlinear dynamics of rotating shallow water: Methods and advances*, in: Edited Series on Advances in Nonlinear Science and Complexity, vol. 2, Elsevier, 2007.

CHAPTER 2

Fast Reaction Limit of Competition-Diffusion Systems

Danielle Hilhorst

CNRS and Laboratoire de Mathématiques, Université de Paris-Sud 11, F-91405 Orsay Cedex, France

Masayasu Mimura

Meiji Institute for Advanced Study of Mathematical Sciences, Meiji University, 1-1 Higashi Mita, Tama-ku, Kawasaki, 214-8571, Japan

Hirokazu Ninomiya

Department of Mathematics, Meiji University, 1-1 Higashi Mita, Tama-ku, Kawasaki, 214-8571, Japan

Contents

HANDBOOK OF DIFFERENTIAL EQUATIONS
Evolutionary Equations, Volume 5
Edited by C.M. Dafermos and M. Pokorný

DOI: 10.1016/S1874-5717(08)00209-0

Abstract

The purpose of this chapter is to investigate the singular limit of reaction-diffusion systems or, more precisely, the fast reaction limit of competition-diffusion systems. It often turns out that such systems converge to free boundary problems, which may have the form of Stefan problems. On the other hand, some reaction-diffusion systems can converge to cross-diffusion systems. As an application of the fast reaction limit, we also discuss the relationship between Turing's instability and cross-diffusion induced instability.

Keywords: Competition-diffusion systems, fast reaction limit, singular limit analysis, two-phase Stefan problem, cross-diffusion systems

1. Introduction

Mechanisms in nature are often quite complicated and the several spatio-temporal stages of these phenomena are related to each other. The model systems, which describe these relationships, include large and/or small parameters. In some cases, these parameters correspond to the scale in space or time. The mixture of several ranges of forces also requires large parameters. Singular limit analysis has been developed in recent decades to treat these systems. In this chapter we focus on one type of singular limit of reaction-diffusion systems, namely the fast reaction limit of competition-diffusion systems. When a system includes reaction terms which are fast when compared to the other terms, we refer to the singular limit as the *fast reaction limit* or more precisely the instantaneous reaction limit, which expresses the fact that instantaneous dynamics is also included in the system. In particular we can expect that the solution converges to the attractor of the dynamics governed by the fast reaction. In this chapter we obtain two kinds of limit problem: (i) free boundary problems, which may have the form of Stefan-like problems; (ii) cross-diffusion systems. Singular limit analysis permits us to connect smooth reaction-diffusion systems to their fast reaction limits. The interest of our study is two-fold: on the one hand, we derive the limit problem for a large number of reaction-diffusion systems with a large reaction coefficient, and it often happens that the numerical solution of the limit problem may be easily computed; on the other hand, we provide approximations of Stefan problems and cross-diffusion systems by means of reaction-diffusion systems. The biological context where these systems arise will be presented in Section 2.

We summarize below the mathematical results of this chapter. A model problem is given by the following two-component problem [9], which we will refer to as Problem (P^k),

$$(P^k) \qquad \begin{cases} u_{1t} = d_1 \Delta u_1 + f(u_1) - s_1 k u_1 u_2, & \text{in } \Omega \times \mathbb{R}^+, \\ u_{2t} = d_2 \Delta u_2 + g(u_2) - s_2 k u_1 u_2, & \text{in } \Omega \times \mathbb{R}^+, \end{cases}$$

with

$$\frac{\partial u_1}{\partial \nu} = \frac{\partial u_2}{\partial \nu} = 0, \quad \text{on } \partial\Omega \times \mathbb{R}^+,$$
$$u_1(x,0) = u_{01}^k(x), \qquad u_2(x,0) = u_{02}^k(x), \quad x \in \Omega,$$

where Ω is a smooth domain of $\mathbb{R}^N$, and where $k, s_1, s_2, d_1, d_2, \lambda$ and μ are positive constants, $f(s) = \lambda s(1-s)$, $g(s) = \mu s(1-s)$, and

$$u_{01}^k, u_{02}^k \in C(\overline{\Omega}), 0 \le u_{01}^k, u_{02}^k \le 1,$$
$$u_{01}^k \rightharpoonup u_{01}, \qquad u_{02}^k \rightharpoonup u_{02}, \quad \text{in } L^2(\Omega) \text{ as } k \to \infty.$$

The functions u_1 and u_2 represent the densities of two biological populations which are competing with each other, λ and μ are the intraspecific competition rates, and $s_1 k$ and $s_2 k$ are the interspecific competition rates. If k is very large, one expects that the two species are nearly spatially segregated.

In other contexts, Problem (P^k) is studied together with the inhomogeneous Dirichlet boundary conditions [7],

$$u_1 = m_1^k, \qquad u_2 = m_2^k, \quad \text{on } \partial\Omega \times \mathbb{R}^+.$$

Since we can treat this case in a similar manner to a case with Neumann boundary conditions, we discuss only the case with the Neumann boundary conditions. An essential idea for Dirichlet boundary conditions is to use test functions which vanish on $\partial\Omega$ to first obtain inner estimates (see [7], for more details).

Our main question is the following: what is the singular limit of the solution (u_1^k, u_2^k) of Problem (P^k) as $k \to \infty$? A first result is the following:

THEOREM 1.1 (*Dancer, Hilhorst, Mimura and Peletier [9, Lemma 3.1]*). *Let Ω be a bounded domain in $\mathbb{R}^N$ $(N \geq 1)$ with a smooth boundary $\partial\Omega$. Let (u_1^k, u_2^k) be the solution of Problem (P^k). Then*

$$u_1^k \to u_1 := s_1 z^+, \quad u_2^k \to u_2 := s_2 z^- \quad \textit{strongly in } L^2(Q_T) \textit{ as } k \to \infty,$$

for all $T > 0$, where $Q_T := \Omega \times (0, T)$,

$$r^+ = \max(r, 0), \qquad r^- = -\min(r, 0)$$

and where the function z is the unique weak solution of the problem

$$(P) \qquad \begin{cases} z_t = \Delta d(z) + h(z), & \textit{in } \Omega \times \mathbb{R}^+, \\ \dfrac{\partial z}{\partial \nu} = 0, & \textit{on } \partial\Omega \times \mathbb{R}^+, \\ z(\cdot, 0) = z_0 := \dfrac{u_{01}}{s_1} - \dfrac{u_{02}}{s_2}, & \textit{on } \Omega, \end{cases}$$

with

$$d(r) := d_1 r^+ - d_2 r^-, \qquad h(r) = \frac{f(s_1 r^+)}{s_1} - \frac{g(s_2 r^-)}{s_2}. \tag{1.1}$$

This result completely characterizes the limit (u_1, u_2) of the sequence (u_1^k, u_2^k) as $k \to \infty$. Problem (P) is a free boundary problem where the free boundary, which coincides with the set $\{(x, t) \in \Omega \times \mathbb{R}^+$ such that $z(x, t) = 0\}$, separates the regions where $u_1 > 0$ from those where $u_2 > 0$. We refer to Problem (P) as a free boundary problem in a weak form since the free boundary does not explicitly appear in its formulation. It is also interesting to write it in a strong form, with explicit relations on the interface. The following result holds.

THEOREM 1.2 (*Dancer, Hilhorst, Mimura and Peletier [9, Theorem 3.6]*). *Assume that, at each time $t \in [0, T]$, there exists a closed hypersurface $\Gamma(t)$ and two subdomains $\Omega_1(t), \Omega_2(t)$ such that*

$$\overline{\Omega} = \overline{\Omega_1(t)} \cup \overline{\Omega_2(t)}, \qquad \Gamma(t) = \overline{\Omega_1(t)} \cap \overline{\Omega_2(t)},$$
$$z(\cdot, t) > 0 \quad \textit{on } \Omega_1(t), \qquad z(\cdot, t) < 0 \quad \textit{on } \Omega_2(t),$$

where z is the unique weak solution of Problem (P). Suppose furthermore that $t \mapsto \Gamma(t)$ is smooth enough and that $(u_1, u_2) := (s_1 z^+, s_2 z^-)$ are smooth up to $\Gamma(t)$. Then the functions u_1 and u_2 satisfy

$$
(P^*) \qquad \begin{cases} u_{1t} = d_1 \Delta u_1 + f(u_1), & \text{in } Q_u := \bigcup \{\Omega_1(t), t \in [0,T]\}, \\ u_{2t} = d_2 \Delta u_2 + g(u_2), & \text{in } Q_v := \bigcup \{\Omega_2(t), t \in [0,T]\}, \\ u_1 = u_2 = 0, & \text{on } \Gamma := \bigcup \{\Gamma(t), t \in t \in [0,T]\}, \\ \dfrac{d_1}{s_1}\dfrac{\partial u_1}{\partial n} = -\dfrac{d_2}{s_2}\dfrac{\partial u_2}{\partial n}, & \text{on } \Gamma, \\ \dfrac{\partial u_1}{\partial \nu} = \dfrac{\partial u_2}{\partial \nu} = 0, & \text{on } \partial\Omega \times [0,T], \end{cases}
$$

together with the initial conditions

$$
\frac{u_1(x,0)}{s_1} = \left(\frac{u_{01}(x)}{s_1} - \frac{u_{02}(x)}{s_2}\right)^+,
$$
$$
\frac{u_2(x,0)}{s_2} = \left(\frac{u_{01}(x)}{s_1} - \frac{u_{02}(x)}{s_2}\right)^- \qquad \text{in } \Omega,
$$

where n is the unit normal vector on $\Gamma(t)$ oriented from $\Omega_1(t)$ to $\Omega_2(t)$. Moreover

$$
\Gamma(0) = \left\{ x \in \Omega \,\middle|\, \frac{u_{01}(x)}{s_1} = \frac{u_{02}(x)}{s_2} \right\}.
$$

This theorem shows that the system converges to a free boundary value problem. Since u_1 and u_2 vanish on the interface $\Gamma(t)$, they are continuous (see Figure 1). However, the derivatives $\partial u_1/\partial n$ and $\partial u_2/\partial n$ are discontinuous near the interface. This interface is called a *corner layer*. This limit problem is a Stefan problem with zero latent heat. We come to the next question: how can we approximate a Stefan problem with positive latent heat by means of a reaction-diffusion system? To that purpose we introduce an extra equation, which has the form of an ordinary differential equation. We now deal with a system of three coupled equations, where the extra unknown function w^k is an approximation of the characteristic function of the habitat of the population of density u_1^k (see Section 3 for more details). We will refer to this problem as Problem (Q^k):

$$
(Q^k) \qquad \begin{cases} u_{1t} = d_1 \Delta u_1 + f(u_1) - s_1 k u_1 u_2 & \\ \qquad - \lambda s_1 k (1-w) u_1, & \text{in } \Omega \times \mathbb{R}^+, \\ u_{2t} = d_2 \Delta u_2 + g(u_2) - s_2 k u_1 u_2 - \lambda s_2 k w u_2, & \text{in } \Omega \times \mathbb{R}^+, \\ w_t = k(1-w)u_1 - k w u_2, & \text{in } \Omega \times \mathbb{R}^+, \\ \dfrac{\partial u_1}{\partial \nu} = \dfrac{\partial u_2}{\partial \nu} = 0, & \text{in } \partial\Omega \times \mathbb{R}^+, \\ u_1(x,0) = u_{01}^k(x), \qquad u_2(x,0) = u_{02}^k(x), & \\ \qquad w(x,0) = w_0^k(x), & x \in \Omega, \end{cases}
$$

where $\lambda \geq 0$ will turn out to be the latent heat coefficient, and where

$0 \leq u_{01}^k \leq \alpha, 0 \leq u_{02}^k \leq \beta, 0 \leq w_0 \leq 1$ for some positive constants α and β.
Define $\Omega_r := \{x \in \Omega \mid B(x, 2r) \subset \Omega\}$ and choose $\hat{r} > 0$ small enough;
for each $r \in (0, \hat{r})$, there exists a positive function $\vartheta(\xi)$
satisfying $\lim_{|\xi| \to 0} \vartheta(\xi) = 0$ such that

$$\int_{\Omega_r} |u_{01}^k(x+\xi) - u_{01}^k(x)| \, \mathrm{d}x \leq \vartheta(\xi),$$

$$\int_{\Omega_r} |u_{02}^k(x+\xi) - u_{02}^k(x)| \, \mathrm{d}x \leq \vartheta(\xi),$$

$$\int_{\Omega_r} |w_0^k(x+\xi) - w_0^k(x)| \, \mathrm{d}x \leq \vartheta(\xi), \text{ and where}$$

$u_{01}^k \to u_{01}, u_{02}^k \to u_{02}, w_0^k \to w_0$ strongly in $L^2(\Omega)$ as $k \to \infty$.

THEOREM 1.3 (*Hilhorst, Iida, Mimura and Ninomiya [18, Theorem 3.13][19, Theorem 1.1]*). *Let Ω be a bounded domain in $\mathbb{R}^N$ $(N \geq 1)$ with a smooth boundary $\partial\Omega$. Let (u_1^k, u_2^k, w^k) be the solution of Problem (Q^k). Then for all $T > 0$,*

$$u_1^k \to u_1, \qquad u_2^k \to u_2, \qquad w_k \to w \quad \text{in } L^2(Q_T) \text{ as } k \to \infty.$$

Let z be the unique weak solution of the Stefan problem with positive latent heat λ,

$$(Q) \qquad \begin{cases} z_t = \Delta d(\varphi_\lambda(z)) + h(\varphi_\lambda(z)), & \text{in } \Omega \times \mathbb{R}^+, \\ \dfrac{\partial d(\varphi_\lambda(z))}{\partial \nu} = 0, & \text{on } \partial\Omega \times \mathbb{R}^+, \\ z(\cdot, 0) = \dfrac{u_{01}}{s_1} - \dfrac{u_{02}}{s_2} + \lambda w_0 & \text{on } \in \Omega, \end{cases}$$

where $d(r), h(r)$ are defined in (1.1) and

$$\varphi_\lambda(r) = (r - \lambda)^+ - r^-.$$

The limit functions are given by

$$u_1 = s_1 \varphi_\lambda(z^+), \qquad u_2 = s_2 \varphi_\lambda(z^-), \qquad w = \frac{z - \varphi_\lambda(z)}{\lambda}.$$

Just as above, we can associate a strong form to this free boundary problem, which is indeed the classical Stefan problem with positive latent heat.

THEOREM 1.4 (*Hilhorst, Iida, Mimura and Ninomiya [18, Theorem 3.13]*). *Assume that, at each time $t \in [0, T]$, there exists a closed hypersurface $\Gamma(t)$ and two subdomains $\Omega_1(t), \Omega_2(t)$ such that*

$$\overline{\Omega} = \overline{\Omega_1(t)} \cup \overline{\Omega_2(t)}, \qquad \Gamma(t) = \overline{\Omega_1(t)} \cap \overline{\Omega_2(t)},$$
$$z(\cdot, t) > 0 \quad on\ \Omega_1(t), \qquad z(\cdot, t) < 0 \quad on\ \Omega_2(t),$$

where z is the unique weak solution of Problem (Q). Suppose furthermore that $t \mapsto \Gamma(t)$ is smooth enough and that $(u_1, u_2) := (s_1 z^+, s_2 z^-)$ are smooth up to $\Gamma(t)$. Then the functions u_1 and u_2 satisfy

$$(Q^*) \qquad \begin{cases} u_{1t} = d_1 \Delta u_1 + f(u_1), & in\ \Omega_1 := \bigcup\{\Omega_1(t), t \in [0,T]\}, \\ u_{2t} = d_2 \Delta u_2 + g(u_2), & in\ \Omega_2 := \bigcup\{\Omega_2(t), t \in [0,T]\}, \\ u_1 = u_2 = 0, & on\ \Gamma := \bigcup\{\Gamma(t), t \in [0,T]\}, \\ \lambda V_n = -\dfrac{d_1}{s_1}\dfrac{\partial u_1}{\partial n} - \dfrac{d_2}{s_2}\dfrac{\partial u_2}{\partial n}, & on\ \Gamma, \\ \dfrac{\partial u_1}{\partial \nu} = \dfrac{\partial u_2}{\partial \nu} = 0, & on\ \partial\Omega \times [0,T], \end{cases}$$

together with the initial conditions

$$\frac{u_1(x,0)}{s_1} = \left(\frac{u_{01}(x)}{s_1} - \frac{u_{02}(x)}{s_2}\right)^+,$$
$$\frac{u_2(x,0)}{s_2} = \left(\frac{u_{01}(x)}{s_1} - \frac{u_{02}(x)}{s_2}\right)^- \quad in\ \Omega.$$

Moreover

$$\Gamma(0) = \left\{ x \in \Omega \,\middle|\, \frac{u_{01}(x)}{s_1} = \frac{u_{02}(x)}{s_2} \right\}.$$

Thus w exhibits a sharp transition layer at the interface, while u_1 and u_2 only have corner layers (see Figure 2). Namely, w has a jump at the interface, while u_1 and u_2 are continuous. Once we consider Theorems 1.3 and 1.4, then the proof of the results of Theorems 1.1 and 1.2 for the two phase Stefan problem with zero latent heat follows, by setting $\lambda = 0$.

The last part of the fast reaction limit for Stefan-like problems is devoted to error estimates, this is inspired by the articles of Murakawa [41,42]. The following result holds true both for Problem (Q^k) in the case that $\lambda > 0$ and for Problem (P^k) in the case that $\lambda = 0$.

THEOREM 1.5 (*Murakawa [41]*). *Let (u_1^k, u_2^k, w^k) be the solution of Problem (Q^k) with initial data $(u_{10}^k, u_{20}^k, w_0^k)$, or let (u_1^k, u_2^k) be the solution of Problem (P^k) with initial data (u_{10}^k, u_{20}^k) if $\lambda = 0$. There exists a positive constant C which does not depend on λ such that*

$$\|u_1^k - u_1\|_{L^2(Q_T)} + \|u_2^k - u_2\|_{L^2(Q_T)} + \left\| \int_0^t (\theta^k - \theta) \right\|_{L^\infty(0,T;H^1(\Omega))}$$
$$+ \|z^k - z\|_{L^\infty(0,T;(H^1(\Omega))^*)} \le C \left(k^{-1/2} + \sigma(k)\right)^{1/2},$$

where

$$\theta^k = \frac{d_1}{s_1}u_1^k - \frac{d_2}{s_2}u_2^k, \qquad \theta = d(\varphi_\lambda(z)), \qquad z^k = \frac{u_1^k}{s_1} - \frac{u_2^k}{s_2} + \lambda w^k, \tag{1.2}$$

$$z_0^k = \frac{u_{10}^k}{s_1} - \frac{u_{20}^k}{s_2} + \lambda w_0^k, \qquad \sigma(k) = \|z_0^k - z_0\|^2_{L^2(\Omega)}. \tag{1.3}$$

Another topic of interest is the approximation of cross-diffusion systems by means of reaction-diffusion systems. We consider the cross-diffusion system,

$$\begin{cases} u_{1t} = \Delta[(d_1 + \alpha u_2)u_1] + (r_1 - a_1u_1 - b_1u_2)u_1, & t > 0, x \in \Omega, \\ u_{2t} = d_2\Delta u_2 + (r_2 - b_2u_1 - a_2u_2)u_2, & t > 0, x \in \Omega, \\ \dfrac{\partial u_1}{\partial \nu} = \dfrac{\partial u_2}{\partial \nu} = 0, & t > 0, x \in \partial\Omega, \\ u_1(x,0) = u_{01}(x), u_2(x,0) = u_{02}(x), & x \in \Omega, \end{cases} \tag{1.4}$$

to which we associate the reaction-diffusion system

$$\begin{cases} v_{1t} = d_1\Delta v_1 + (r_1 - a_1(v_1 + v_2) - b_1v_3)v_1 \\ \qquad + k[h_{2\to1}(v_3)v_2 - h_{1\to2}(v_3)v_1], \quad t > 0, x \in \Omega, \\ v_{2t} = (d_1 + \alpha M_2)\Delta v_2 + (r_1 - a_1(v_1 + v_2) - b_1v_3)v_2 \\ \qquad + k[h_{1\to2}(v_3)v_1 - h_{2\to1}(v_3)v_2], \quad t > 0, x \in \Omega, \\ v_{3t} = d_2\Delta v_3 + (r_2 - b_2(v_1 + v_2) - a_2v_3)v_3, \quad t > 0, x \in \Omega. \\ \dfrac{\partial v_1}{\partial \nu} = \dfrac{\partial v_2}{\partial \nu} = \dfrac{\partial v_3}{\partial \nu} = 0, \quad t > 0, x \in \partial\Omega, \\ v_1(x,0) = v_{01}(x), v_2(x,0) = v_{02}(x), v_3(x,0) = v_{03}(x), \quad x \in \Omega, \end{cases} \tag{1.5}$$

where d_i, r_i, a_i, b_i $(i = 1, 2)$ are positive constants. We can show the following theorem.

THEOREM 1.6 (*Iida, Mimura and Ninomiya [23, Theorem 1]*). *Let* $(u_1, u_2) = (u_1(x,t), u_2(x,t))$ *be the solution of the cross-diffusion system* (1.4). *Suppose that* (u_1, u_2) *are sufficiently smooth on* $[0, T] \times \overline{\Omega}$ *and satisfy*

$$0 \le u_1(x,t) \le M_1, \qquad 0 \le u_2(x,t) \le M_2 \quad \text{on } [0, T] \times \Omega, \tag{1.6}$$

for some positive numbers T, M_1 *and* M_2 *with* $M_2 \ge r_2/a_2$. *Choose smooth functions* $h_{1\to2}$ *and* $h_{2\to1}$ *satisfying*

$$h_{1\to2}(s) \equiv (h_{1\to2}(s) + h_{2\to1}(s))\,\frac{s}{M_2} \tag{1.7}$$

for $s \in [0, M_2]$, *and*

$$h_{1\to 2}(s) \geq 0, \tag{1.8}$$

$$h_{2\to 1}(s) \geq 0, \tag{1.9}$$

$$h_{1\to 2}(s) + h_{2\to 1}(s) > 0 \tag{1.10}$$

for $s \in [0, M_2]$. *Define the initial data* (v_{01}, v_{02}, v_{03}) *by*

$$\begin{cases} v_{01}(x) \equiv \left\{1 - \dfrac{u_{02}(x)}{M_2}\right\} u_{01}(x), \\ v_{02}(x) \equiv \dfrac{u_{02}(x)}{M_2} u_{01}(x), \\ v_{03}(x) \equiv u_{02}(x) \end{cases} \tag{1.11}$$

over Ω. *Let* $(v_1, v_2, v_3) = (v_1^k(x,t), v_2^k(x,t), v_3^k(x,t))$ *be the solution of* (1.5) *and suppose that there exist positive numbers* k_0 *and* M_0 *satisfying*

$$|v_1^k(x,t)| + |v_2^k(x,t)| + |v_3^k(x,t)| \leq M_0 \tag{1.12}$$

for $(x,t) \in \overline{\Omega} \times [0,T]$ *and* $k \geq k_0$. *Then the difference between* $(v_1 + v_2, v_3)$ *and* (u_1, u_2) *is estimated to be*

$$\begin{cases} \sup\limits_{t\in[0,T]} \|v_1^k(\cdot,t) + v_2^k(\cdot,t) - u_1(\cdot,t)\|_{L^2(\Omega)} \leq \dfrac{C}{k}, \\ \sup\limits_{t\in[0,T]} \|v_3^k(\cdot,t) - u_2(\cdot,t)\|_{L^2(\Omega)} \leq \dfrac{C}{k} \end{cases} \tag{1.13}$$

for $k \geq k_0$. *Here* $C = C(u_1, u_2, k_0, M_0, T)$ *is a positive constant independent of* k.

This theorem implies that rather complicated diffusion such as cross-diffusion can be realized by the usual random movement together with a reaction mechanism. The proof of this theorem is based on constructing an energy functional E^k for the difference between the solution of the reaction-diffusion system and that of the cross-diffusion system (see Appendix A.3), while Problems (1.4) and (1.5) do not seem to have any Lyapunov functionals.

As far as Lyapunov functionals are concerned, the problems which we study in Sections 3 and 4 are different. The two component system Problem (P^k) does not possess any Lyapunov functional, while the limit problem (P) does possess a Lyapunov functional. In both cases, however, the dynamics governed by the fast reaction possesses a Lyapunov functional. This structure of the fast reaction limit system helps us to prove the convergence and the error estimates.

Mathematically the Lyapunov functional is extremely useful both for proving the stabilization of solution orbits at large time and for obtaining an extra uniform in k *a priori* estimate. This fact has been used by [6] to prove the large time stabilization of solutions of Problem (P^k) for large enough values of k in the case of equal diffusion coefficients.

The organization of this chapter is as follows. In Section 2.1, we present the biological context of Problem (P^k) and discuss the known results. In Section 2.2, we show the relationship between random movement and the fast reaction limit. This argument indicates that the cross-diffusion induced instability corresponds to the Turing instability of the approximating reaction-diffusion system.

In Section 3, we present the proofs of Theorems 1.3 and 1.4 about the singular limit of the solutions of Problem (Q^k). The proofs of Theorems 1.1 and 1.2 about the singular limit of the solutions of Problem (P^k) essentially follow in the same way by setting $\lambda = 0$. Finally, we give the proof of Theorem 1.5, where we obtain error estimates which hold for both the Problems (P^k) and (Q^k).

We prove the convergence of the competition-diffusion system (1.5) to the cross-diffusion system (1.4) in Section 4.

Our presentation involves concrete reaction-diffusion systems which we study by means of methods from nonlinear analysis, such as maximum principle arguments and energy estimates. We refer to articles by Bothe [2,3] for the study of slightly more general partial differential equations in an abstract setting, where he uses, in particular, arguments from the theory of m-accretive operators in Banach spaces.

Finally let us mention a number of open problems. One can extend Problem (Q^k) to the higher component system Problem (3.24), and, as shown in Figure 3, Problem (3.24) may give rise to triple junctions in the limit as $k \to \infty$. Up to now, there is no rigorous proof of this fact. Furthermore, when k is large enough, it would be very interesting to obtain results about the convergence to a stationary solution of the solution of Problem (P^k) in the case of different diffusion coefficients; moreover, similar questions are also unanswered for solutions of Problem (Q^k) and more generally for solutions of reaction-diffusion systems containing both partial and ordinary differential equations. The fast reaction terms, where we can show the convergence, are restricted. It is a challenging problem to extend our study to the cases, where the interspecific reaction terms are not equal in Problems (P^k) and (Q^k). It is also of great interest to study numerical analysis aspects of fast reaction limits. Let us mention the results of Ikota, Mimura and Nakaki [24] in this direction.

2. Biological context and earlier results

The problems which we consider arise in two very different application contexts, on the one hand, the spatial segregation limit of biological populations, and on the other hand, the relationship between Turing instability and the cross-diffusion induced instability.

2.1. *Spatial segregation limit of biological populations*

The understanding of the interaction of biological species arising in ecological systems has recently developed as a central problem in population ecology. In particular, problems of coexistence and exclusion of competing species have been theoretically investigated using continuous models based on partial and ordinary differential equations. Among many models proposed so far, reaction-diffusion equation models are used to study the spatial segregation of competing species which move by diffusion. More precisely, consider the case of several biological species (or several chemical substances) which coexist, and assume that m species living in a habitat $\Omega \subset \mathbb{R}^N$ $(N \geq 1)$ compete with each other. We denote by $u_i(x,t)$ $(i = 1, 2, \ldots, m)$ their population densities at position $x \in \Omega$ and time $t \geq 0$. The model which we mainly study is the following competition-diffusion system with Lotka-Volterra type nonlinearities:

$$u_{it} = d_i \Delta u_i + \left(r_i - a_i u_i - \sum_{j=1}^{m} b_{ij} u_j \right) u_i$$
$$(i = 1, 2, \ldots, m) \quad x \in \Omega, t > 0, \tag{2.1}$$

where Ω is bounded with the smooth boundary $\partial\Omega$ and all the rates d_i, r_i, a_i and b_{ij} are positive constants. The constant d_i is the diffusion rate, r_i the intrinsic growth rate, a_i the intraspecific competition rate, that is the competition between members of the *same* species u_i, and b_{ij} the interspecific competition rate, that is, the competition between members of the *different* species u_i and u_j. We impose the no-flux boundary conditions on the boundary $\partial\Omega$,

$$\frac{\partial u_i}{\partial \nu} = 0, \quad (i = 1, 2, \ldots, m) \quad x \in \partial\Omega, t > 0, \tag{2.2}$$

where ν is the outward normal unit vector to $\partial\Omega$. The initial conditions are given by

$$u_i(0, x) = u_{0i}(x) \geq 0 \quad (i = 1, 2, \ldots, m) \quad x \in \Omega. \tag{2.3}$$

In order to analyse the system (2.1)–(2.3) we discuss the simplest case of (2.1) with $m = 2$, namely

$$\begin{cases} u_{1t} = d_1 \Delta u_1 + (r_1 - a_1 u_1 - b_1 u_2) u_1 & x \in \Omega, t > 0, \\ u_{2t} = d_2 \Delta u_2 + (r_2 - a_2 u_2 - b_2 u_1) u_2 & x \in \Omega, t > 0, \end{cases} \tag{2.4}$$

with the boundary conditions

$$\frac{\partial u_1}{\partial \nu} = 0, \qquad \frac{\partial u_2}{\partial \nu} = 0, \quad x \in \partial\Omega, t > 0. \tag{2.5}$$

We first present some historical remarks for the system (2.4)–(2.5). Hirsch [20] shows that the stable attractor of the system consists of equilibrium solutions (also see Matano and Mimura [35]). This indicates that existence and stability of non-negative equilibrium solutions is important for the study of the asymptotic behavior of solutions. Kishimoto and Weinberger [31] showed that, if Ω is convex, all spatially non-constant equilibrium solutions are unstable. This implies that solutions of (2.4), (2.5) become spatially homogeneous asymptotically, in other words, stable equilibrium solutions of the diffusion-less system (2.7) of (2.4) are very important, in order to know the asymptotic behavior of the solutions of (2.4), even if the diffusion term is present. In fact, we suppose that two species are strongly competing, that is, the interspecific competition rate is stronger than the intraspecific one so that we require that

$$\frac{a_1}{b_2} < \frac{r_1}{r_2} < \frac{b_1}{a_2}. \tag{2.6}$$

Under this assumption the stable spatially constant equilibrium solutions of (2.4), (2.5) are given by $(u_1, u_2) = (r_1/a_1, 0)$ and $(0, r_2/a_2)$. This means that one of the competing species survives and the other becomes extinct. In ecological terms, this implies that the two competing species can never coexist under strong competition. This is called *Gause's competitive exclusion*.

On the other hand, if the domain Ω is not convex, the structure of equilibrium solutions becomes complicated, depending on the shape of Ω [11]. In fact, if Ω takes a suitable dumb-bell shape in two dimensions, there exist stable spatially inhomogeneous equilibrium solutions which exhibit spatial segregation in the sense that u_1 and u_2 take values close to (r_1/a_1) in one subregion and close to $(0, r_2/a_2)$ in the other. Thus the results above give us information on the asymptotic behaviour of solutions. However, from the viewpoint of ecological applications, it is more interesting to know the transient behaviour of solutions. For this purpose we consider the situation where the diffusion rates d_1 and d_2 are sufficiently small or all of the other rates r_i, a_i and b_i are sufficiently large and satisfy (2.6). We rewrite (2.4) as

$$\begin{cases} u_{1t} = \varepsilon^2 \Delta u_1 + (r_1 - a_1 u_1 - b_1 u_2) u_1, & x \in \Omega, t > 0, \\ u_{2t} = d\varepsilon^2 \Delta u_2 + (r_2 - a_2 u_2 - b_2 u_1) u_2, & x \in \Omega, t > 0, \end{cases} \tag{2.7}$$

in which ε is a small parameter. If the competing species segregate according to (2.7), it is natural to define the subregions $\Omega_1(t) = \{x \in \Omega : (u_1, u_2)(x,t) \approx (r_1/a_1, 0)\}$ and $\Omega_2(t) = \{x \in \Omega : (u_1, u_2)(x,t) \approx (0, r_2/a_2)\}$.

In order to study the dynamics of the segregation between u_1 and u_2, we take the limit $\varepsilon \downarrow 0$ in (2.7) so that the internal layers which exist for small values of $\varepsilon > 0$ become sharp interfaces, say $\Gamma(t)$, which is the boundary between the two regions $\Omega_1(t)$ and $\Omega_2(t)$. Using singular limit analysis, Ei and Yanagida [12] derived the following evolution equation to describe the motion of the interface $\Gamma(t)$,

$$V = \varepsilon L(d)(N-1)\kappa + c, \tag{2.8}$$

where V is the normal velocity of the interface, κ the mean curvature of the interface, $L(d)$ a positive constant depending on d such that $L(1) = 1$ and c the velocity of the travelling wave solution (u_1, u_2) of the one-dimensional system corresponding to (2.4) with $d_1 = 1$ and $d_2 = d$, namely

$$\begin{cases} u_{1t} = u_{1xx} + (r_1 - a_1 u_1 - b_1 u_2) u_1, & x \in \mathbb{R}, t > 0, \\ u_{2t} = d u_{2xx} + (r_2 - b_2 u_1 - a_2 u_2) u_2, & x \in \mathbb{R}, t > 0, \end{cases} \tag{2.9}$$

with the boundary conditions at infinity

$$(u_1, u_2)(-\infty, t) = \left(\frac{r_1}{a_1}, 0\right) \quad \text{and} \quad (u_1, u_2)(\infty, t) = \left(0, \frac{r_2}{a_2}\right). \tag{2.10}$$

Kan-on [26] proved that the velocity of the travelling wave solution of Problem (2.9)–(2.10) is unique for fixed values of the rates r_i, a_i and b_i $(i = 1, 2)$. In particular, if a_1 is a free parameter and the other parameters are fixed and satisfy the inequalities (2.6), then there exists a unique constant $a^* > 0$ such that $c = 0$ if $a_1 = a^*$, $c > 0$ if $a_1 > a^*$, and $c < 0$ if $a_1 < a^*$. For the special case when $c = 0$ and $L = 1$, (2.8) becomes the equation of motion by mean curvature, which has been extensively studied. The manifold $\Gamma(t)$ obtained from (2.8) provides information on the dynamics of the spatial segregation between the two competing species.

This result clearly shows the similarity between this class of problem and the Allen-Cahn equation first studied by Keller, Sternberg and Rubinstein [30], where the limiting interface moves according to its mean curvature.

In this chapter, we consider a different situation from the one obtained above, namely the case that only the interspecific competition rates b_1 and b_2 are very large. In order to study this situation, it is convenient to rewrite (2.4) as

$$\begin{cases} u_{1t} = d_1 \Delta u_1 + f(u_1) - s_1 k u_1 u_2, & x \in \Omega, t > 0, \\ u_{2t} = d_2 \Delta u_2 + g(u_2) - s_2 k u_1 u_2, & x \in \Omega, t > 0, \end{cases} \tag{2.11}$$

with

$$\frac{\partial u_1}{\partial \nu} = \frac{\partial u_2}{\partial \nu} = 0, \quad x \in \partial\Omega, t > 0, \tag{2.12}$$

$$u_1(x, 0) = u_{01}(x), \qquad u_2(x, 0) = u_{02}(x), \quad x \in \Omega, \tag{2.13}$$

where $f(u_1) = (r_1 - a_1 u_1)u_1$, $g(u_2) = (r_2 - a_2 u_2)u_2$ and s_1, s_2 and k are positive constants. If s_1/s_2 is small, for instance, the influence of competition on u_1 is weaker than that on u_2. If the competition rate k is very large, one can expect that the two species hardly coexist and are spatially segregated.

Heuristically one can understand it in the following way. If we use $\varepsilon = 1/k$, then the system can be rewritten as

$$\begin{cases} \varepsilon u_{1t} = \varepsilon d_1 \Delta u_1 + \varepsilon f(u_1) - s_1 u_1 u_2, & x \in \Omega, t > 0, \\ \varepsilon u_{2t} = \varepsilon d_2 \Delta u_2 + \varepsilon g(u_2) u_2 - s_2 u_1 u_2, & x \in \Omega, t > 0. \end{cases}$$

Formally, we have $u_1 u_2 = 0$ as the limit when ε tends to zero. The purpose of this chapter is to derive the limiting system as $\varepsilon \to 0$, which is called the *fast reaction limit*.

Let $\Gamma(t)$ be the interface which separates the two subregions

$$\Omega_1(t) = \{x \in \Omega : u_1(x, t) > 0, u_2(x, t) = 0\},$$

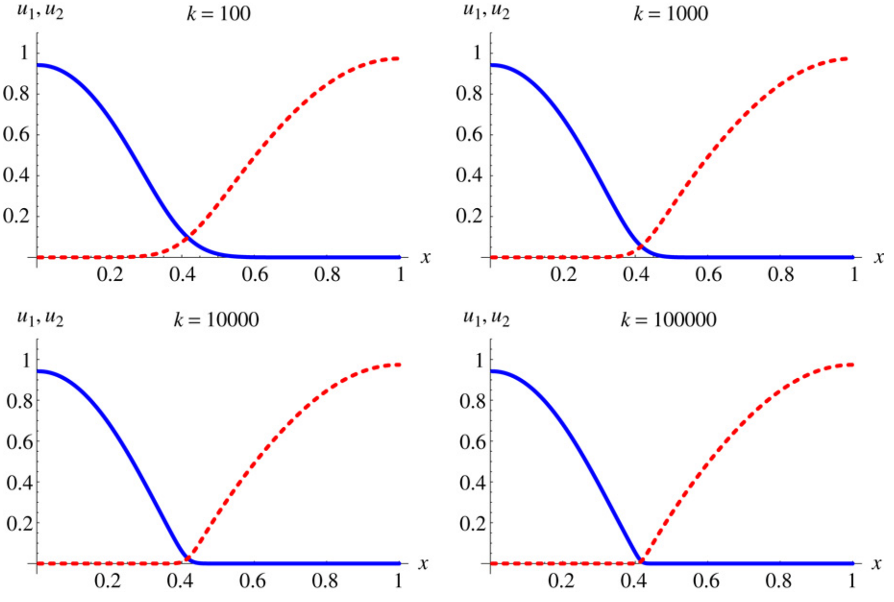

Fig. 1. Dependence on k of the spatial profiles of one-dimensional solutions of (P^k); solid curve : u_1, dotted curve : u_2.

and

$$\Omega_2(t) = \{x \in \Omega : u_1(x,t) = 0, u_2(x,t) > 0\},$$

in Ω (see Figure 1).

Then u_1 and u_2 satisfy

$$\begin{aligned} u_{1t} &= d_1 \Delta u_1 + f(u_1), \quad x \in \Omega_1(t),\ t > 0 \\ u_{2t} &= d_2 \Delta u_2 + g(u_2), \quad x \in \Omega_2(t),\ t > 0 \\ \frac{\partial u_1}{\partial \nu} &= 0, \qquad \frac{\partial u_2}{\partial \nu} = 0, \quad x \in \partial\Omega, t > 0. \end{aligned}$$

On the interface,

$$u_1 = 0, \qquad u_2 = 0, \quad x \in \Gamma(t) \text{ for } t > 0,$$

and

$$\frac{d_1}{s_1}\frac{\partial u_1}{\partial n} = -\frac{d_2}{s_2}\frac{\partial u_2}{\partial n}, \quad x \in \Gamma(t) \text{ for } t > 0.$$

In order to visualize how the segregation of two competing species depends on the value of k, we present two-dimensional numerical simulations of System (2.11), together with the boundary conditions (2.12) in a rectangular domain (see Figure 1). We take k as a free parameter and keep the other parameters fixed. For values of k which are neither large nor small, it is shown that u_1 and u_2 exhibit spatial segregation with a rather wide zone of overlap. When the value of k increases, the zone of overlap becomes narrower. Thus, taking the limit $k \to \infty$, one can expect that u_1 and u_2 have disjoint supports (habitats) with only one common curve, which separates the habitats of the two competing species.

2.2. *Cross diffusion and random movements*

When the individuals of the species move randomly, the population density satisfies a diffusion equation. Adding the competitive effect, we consider the competition-diffusion system (2.4) in the previous subsection. However, the individuals in the field may not move around randomly [28]. This situation can not be described by (2.4) any more. By considering the transition possibilities, the following three types of diffusion for the population density n are introduced in [43, Section 5.4]:

(i) when transition possibilities from a place A to another place B and from B to A are equal (neutral transition):

$$n_t = \nabla(D(x)\nabla n)$$

(ii) when the transition possibilities depend only on the departure points (repulsive transition):

$$n_t = \Delta(D(x)n)$$

(iii) when the transition possibilities depend only on the arrival points (attractive transition):

$$n_t = D(x)^2\nabla\left(\frac{\nabla n}{D(x)}\right).$$

If the movement depends on the environmental conditions, it is natural that the transition possibilities depend on the departure points. Thus the diffusion in (2.4) should be modified as in (ii). One of the ways to formulate such a situation is to replace $d_1\Delta u_1$ and $d_2\Delta u_2$ in (2.4) with $\Delta[D_1(x)u_1]$ and $\Delta[D_2(x)u_2]$ respectively, where $D_i(x)$ indicates the motility for the population of the species u_i at the position x (for example, see also [45, Section 4.2.3]). More individuals of u_i tend to leave the place x where $D_i(x)$ is larger (see [43]). Shigesada, Kawasaki and Teramoto in [44] extended this idea to the environmental pressures due to the inter-specific interferences by replacing $D_i(x)$ with a linearly increasing function of the competitor's density, in order to formulate spatial segregation phenomena between the competing species. Thus they constructed cross-diffusion models for the competitive interaction between two species:

$$\begin{cases} u_{1t} = \Delta[(d_1+\alpha u_2)u_1] + (r_1 - a_1u_1 - b_1u_2)u_1, & t>0, x\in\Omega, \\ u_{2t} = \Delta[(d_2+\beta u_1)u_2] + (r_2 - b_2u_1 - a_2u_2)u_2, & t>0, x\in\Omega, \end{cases} \tag{2.14}$$

where α, β stand for the cross-diffusion pressures and are non-negative constants. These constants α and β are called cross-diffusion coefficients. Mathematically, (2.14) becomes a quasi-linear parabolic system and it is not so easy to handle ([1] for the local existence). The global existence problem of solutions has been studied by several authors (for instance [34, 8] and the references therein). From spatial-segregation viewpoints, the stationary problem for (2.14) has been studied in three basically different analytical approaches: Bifurcation approach [36], Singular perturbation approach [37,25,33] and Elliptic approach [32]. These results indicate that the structure of equilibrium solutions of (2.14) sensitively depends on parameters in the systems and is extremely complicated, even in one dimension.

In order to understand the meaning of the cross-diffusion effect in (2.14), let us go back to the simplest equation for the repulsive transition, i.e., the movement of a single species in the inhomogeneous medium which is specified by a given function $V(x)$ $(0 \le V(x) \le 1)$. Suppose that the place x where $V(x)$ is larger is more unfavorable for the species. By replacing u_2 in (2.14) with $V(x)$ and neglecting the growth term, we obtain the following scalar linear diffusion equation like (ii) for the population density n:

$$\begin{cases} n_t = \Delta\left[(d + \alpha V(x))\, n\right], & t > 0, x \in \Omega, \\ \dfrac{\partial n}{\partial \nu} = 0, & t > 0, x \in \partial\Omega, \\ n(0, x) = n_0(x), & x \in \Omega, \end{cases} \tag{2.15}$$

where $d + \alpha V(x)$ stands for the motility of the species in a small area around x (cf. [43] or [44]). Since

$$(d + \alpha V(x))n = d \cdot \{1 - V(x)\}n + (d + \alpha) \cdot V(x)n,$$

we can split the variable n into $n_1 := \{1 - V(x)\}n$ and $n_2 := V(x)n$, and rewrite (2.15) as

$$(n_1 + n_2)_t = \Delta\left[dn_1 + (d + \alpha)n_2\right].$$

This indicates that the population n governed by (2.15) becomes the sum of partial population n_1 and n_2 with low motility d and high motility $d+\alpha$, respectively. We also notice that the ratio between the two population depends only on $V(x)$ at x. Thus we may suppose that each individual has two types of mobile state, i.e., one is less active (resp. active) and the other is active (resp. more active), and that it physiologically converts one type into the other as a very quick response to the environmental pressure $V(x)$. This quick response may lead the population n to a quasi-steady state. To construct a system for n_1 and n_2, we denote by $p_{2\to1}(x)$ (resp. $p_{1\to2}(x)$) the probability that an active individual becomes less active (resp. one that a less active individual becomes active). Under the above situation, $p_{2\to1}(x)$ (resp. $p_{1\to2}(x)$) should be lower (resp. higher) where the place x is more unfavorable for the species. We also assume that the transition between the two mobile states is much faster than the random diffusion of n_1 and n_2 and denote by $\tau_{2\to1}/k$ (resp. by $\tau_{1\to2}/k$) an average time in converting the active state into the less active one (resp. an average time in converting reversely) where k is a positive large parameter. By using

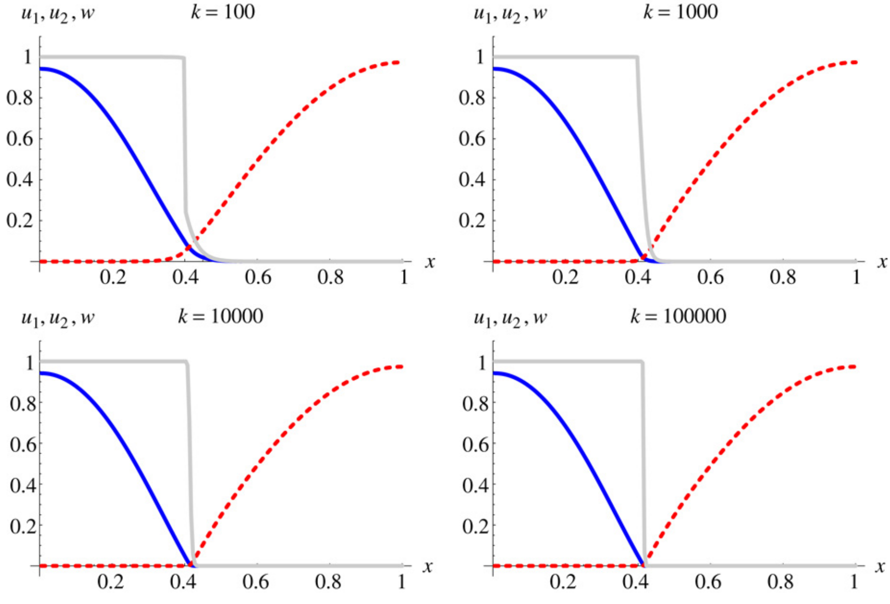

Fig. 2. Dependence on k of the spatial profiles of one-dimensional solutions of (Q^k); solid curve: u_1, dotted curve: u_2, the gray curve: w.

$h_{1\to 2}(x) := p_{1\to 2}(x)/\tau_{1\to 2}$ and $h_{2\to 1}(x) := p_{2\to 1}(x)/\tau_{2\to 1}$, the dynamics of n_1 and n_2 can be described by

$$\begin{cases} n_{1t} = d\Delta n_1 + k[h_{2\to 1}(x)n_2 - h_{1\to 2}(x)n_1], \\ n_{2t} = (d+\alpha)\Delta n_2 + k[h_{1\to 2}(x)n_1 - h_{2\to 1}(x)n_2]. \end{cases} \tag{2.16}$$

By an appropriate choice of a function $V(x)$, we can formally regard (2.16) as an approximation to (2.15) as follows. Adding both the equations in (2.16), we obtain

$$n_t = \Delta\left[dn_1 + (d+\alpha)n_2\right],$$
$$n_{2t} = (d+\alpha)\Delta n_2 + k[h_{1\to 2}(x)n_1 - h_{2\to 1}(x)n_2],$$

where $n = n_1 + n_2$. Dividing the second equation by k and letting $k \to \infty$ in the second equation, one can formally expect that

$$h_{2\to 1}(x)n_2 = h_{1\to 2}(x)n_1.$$

This implies that

$$n_1 = \frac{h_{2\to 1}(x)}{h_{1\to 2}(x) + h_{2\to 1}(x)}n, \qquad n_2 = \frac{h_{1\to 2}(x)}{h_{1\to 2}(x) + h_{2\to 1}(x)}n.$$

Therefore, n satisfies

$$n_t = \Delta \left[\left(d + \alpha \frac{h_{1\to 2}(x)}{h_{1\to 2}(x) + h_{2\to 1}(x)} \right) n \right].$$

Thus, $V(x)$ at the place x can be interpreted as the ratio $h_{1\to 2}(x)/\{h_{1\to 2}(x) + h_{2\to 1}(x)\}$ with use of the transition rates $h_{1\to 2}(x)$ and $h_{2\to 1}(x)$ between the two states. This formal calculation indicates that the movement by repulsive transition can be represented by random walking if individuals of the species possess the mechanism of switching from one type of random walking to the other, depending on environmental inhomogeneity.

Rather than applying this method to (2.14), we use a simplified system of (2.14) by $\beta = 0$, that is, (1.4). We split the population of one species U_1 into two types and each individual of U_1 converts its state into the other one. We denote by $v_1(x,t)$ and $v_2(x,t)$ the population densities of the less active type and the active one of U_1, respectively, and rewrite the population density of the species U_2 as $v_3(x,t)$. Thus we can construct a reaction-diffusion system for (v_1, v_2, v_3), which approximates bounded solutions (u_1, u_2) in $\Omega \times [0, T]$. Assume (1.6). Seeing that

$$\begin{aligned} &d_1 \le d_1 + \alpha u_2(x,t) \le d_1 + \alpha M_2, \\ &(d_1 + \alpha u_2)u_1 = d_1 \cdot \left(1 - \frac{u_2}{M_2}\right) u_1 + (d_1 + \alpha M_2) \cdot \frac{u_2}{M_2} u_1, \end{aligned}$$

we replace $n, n_1, n_2, d, d+\alpha$ and V in the discussion above with $u_1, v_1, v_2, d_1, d_1 + \alpha M_2$ and u_2/M_2, respectively. The transition rates $h_{1\to 2}$ and $h_{2\to 1}$ between the two mobile states of U_1 should be functions of $v_3(x,t)$ such that $h_{1\to 2}(v_3)$ and $h_{2\to 1}(v_3)$ increase and decrease, respectively, as v_3 increases. Assuming that the local reproduction rates of the two mobile states of U_1 are equal, we can formally reformulate the population dynamics of U_1 and U_2 by the following three component reaction-diffusion system (1.5). Setting $\rho = v_1 + v_2$, we rewrite (1.5) as

$$\begin{cases} \rho_t = d_1 \Delta \rho + \alpha M_2 \Delta v_2 + (r_1 - a_1 \rho - b_1 v_3)\rho, \\ v_{2t} = (d_1 + \alpha M_2)\Delta v_2 + (r_1 - a_1 \rho - b_1 v_3) v_2 \\ \qquad + k\left[h_{1\to 2}(v_3)\rho - (h_{2\to 1}(v_3) + h_{1\to 2}(v_3))\, v_2\right], \\ v_{3t} = d_2 \Delta v_3 + (r_2 - b_2 \rho - a_2 v_3) v_3. \end{cases} \tag{2.17}$$

We can expect that v_2 converges to $h_{1\to 2}(v_3)\rho/(h_{1\to 2}(v_3) + h_{2\to 1}(v_3))$ as $k \to \infty$ from the second equation of (2.17). Replacing v_2 in the first equation of (2.17) by this limit, we formally get a cross-diffusion system

$$\begin{cases} u_{1t} = \Delta \left[\left(d_1 + \alpha \dfrac{M_2 h_{1\to 2}(u_2)}{h_{1\to 2}(u_2) + h_{2\to 1}(u_2)} \right) u_1 \right] \\ \qquad + (r_1 - a_1 u_1 - b_1 u_2) u_1, & t > 0, x \in \Omega, \\ u_{2t} = d_2 \Delta u_2 + (r_2 - b_2 u_1 - a_2 u_2) u_2, & t > 0, x \in \Omega, \end{cases}$$

where (u_1, u_2) is a pair of limiting functions of (ρ, v_3) as $k \to \infty$. In particular, we expect (1.5) to be a "reaction-diffusion approximation" to (1.4), provided that

$$h_{1\to 2}(s) \equiv (h_{1\to 2}(s) + h_{2\to 1}(s)) \frac{s}{M_2} \tag{2.18}$$

for $s \in [0, M_2]$. As the simplest example of $h_{2\to 1}$ and $h_{1\to 2}$, we can take

$$h_{2\to 1}(s) = 1 - \frac{s}{M_2}, \quad h_{1\to 2}(s) = \frac{s}{M_2},$$

where $h_{2\to 1}(v_3)$ (resp. $h_{1\to 2}(v_3)$) is surely a decreasing (resp. an increasing) function of v_3.

In this context we have Theorem 1.6, which shows that the solutions of (1.4) can be approximated by those of (1.5) in a finite time interval if the solutions are bounded. Moreover, the conclusion of this theorem still holds true for several other expressions of the motility of u_1 and for those of the reproduction rates of v_1 and v_2, as the proof in Section 4 shows. Although there are many choices of $h_{1\to 2}(s)$ and $h_{2\to 1}(s)$ satisfying (1.7), the ratio $h_{2\to 1}(s)/h_{1\to 2}(s)$ is uniquely determined by (1.7). We recall that the ratio of individual transition rates between the two mobile states determines the motility of the population. For example,

(i) $h_{1\to 2}(v_3) := \frac{v_3}{M_2}$, $h_{2\to 1}(v_3) := 1 - \frac{v_3}{M_2}$,
(ii) $h_{1\to 2}(v_3) := \frac{v_3}{M_2 - v_3}$, $h_{2\to 1}(v_3) := 1$.

If we choose as

(iii) $h_{1\to 2}(v_3) := \frac{\varphi(v_3)}{1+\varphi(v_3)}$, $h_{2\to 1}(v_3) := \frac{1}{1+\varphi(v_3)}$,

or

(iv) $h_{1\to 2}(v_3) := \varphi(v_3)$, $h_{2\to 1}(v_3) := 1$,

then the corresponding cross-diffusion system is formally

$$\begin{cases} u_{1t} = \Delta \left[\left(d_1 + \dfrac{\alpha M_2 \varphi(u_2)}{1+\varphi(u_2)} \right) u_1 \right] \\ \qquad + (r_1 - a_1 u_1 - b_1 u_2) u_1, & t > 0, x \in \Omega, \\ u_{2t} = d_2 \Delta u_2 + (r_2 - b_2 u_1 - a_2 u_2) u_2, & t > 0, x \in \Omega. \end{cases}$$

In particular, for $\varphi(s) = s$ and $\tilde{\alpha} = \alpha$, the resulting cross-diffusion system is close to the original system (1.4), if u_2 is small. We will prove the extended version of this theorem in Section 4. We refer to [21] for related work.

Next we consider the application from the viewpoint of pattern formation. *Turing's instability* or *diffusion-induced instability* is one of the most important mechanisms of pattern formation. This phenomenon is observed under the situation where the inhibitor diffuses faster than the activator [46]. If the activator increases locally, then it generates the inhibitor at the same time. Because of the large diffusivity, the inhibitor also increases outside its neighborhood of the high concentration of the activator. This keeps the density of the activator low outside and the inhomogeneity of the distribution of the activator forms. Mathematically speaking, the stable equilibrium points of some ordinary differential equations become unstable by adding the diffusion. Consider the competition-diffusion system (2.4) under the weak competition condition:

$$\frac{b_1}{a_2} < \frac{r_1}{r_2} < \frac{a_1}{b_2}, \tag{2.19}$$

instead of (2.6). Under this assumption, there is only one stable spatially constant equilibrium solution to the diffusion-less system (2.7)

$$\begin{aligned}(u_1^*, u_2^*) = ((a_2r_1 - b_1r_2)/(a_1a_2 - b_1b_2),\\ (-b_2r_1 + a_1r_2)/(a_1a_2 - b_1b_2))\end{aligned}$$

and three unstable equilibrium solutions $(0,0)$, $(r_1/a_1, 0)$ and $(0, r_2/a_2)$. For the competition-diffusion system (2.4), it is well known that Turing's instability never occurs, that is, the spatially constant equilibrium solution (u_1^*, u_2^*) is always stable. Actually the comparison principle directly implies that all the solutions converge to the constant solution (u_1^*, u_2^*) when both components of initial data are positive.

On the other hand, Mimura and Kan-on reported in [36] that stable non-constant stationary solutions of the cross-diffusion system (1.4) bifurcate from the stable constant solution (u_1^*, u_2^*) under the weak competition condition. When α is small enough, (u_1^*, v_1^*) is still stable for (1.4). However, (u_1^*, u_2^*) loses its stability as α increases (see [27,35]).

These results indicate that the spatially segregating coexistence of two competing species occurs by the cross-diffusion effect, which is called *cross-diffusion induced instability* [38]. This result shows a remarkable contrast with the fact of (2.4) that (u_1^*, u_2^*) is always stable under (2.19). In Section 2.3 we will make clear the relationship between Turing's instability and the cross-diffusion induced instability for (1.4).

2.3. *Turing's instability and the cross-diffusion induced instability*

In this subsection we will explain the relationship between the diffusion-induced instability (or Turing's instability) and the cross-diffusion induced instability.

First let us consider the linearized stability of the constant stationary solution (u_1^*, u_2^*) for (1.4), that is,

$$\begin{cases} u_{1t} = \Delta[(d_1 + \alpha u_2)u_1] + f(u_1, u_2), & t > 0, x \in \Omega, \\ u_{2t} = d_2\Delta u_2 + g(u_1, u_2), & t > 0, x \in \Omega \end{cases} \tag{2.20}$$

for $\alpha > 0$, where

$$\begin{cases} f(u_1, u_2) = (r_1 - a_1u_1 - b_1u_2)u_1, \\ g(u_1, u_2) = (r_2 - b_2u_1 - a_2u_2)u_2. \end{cases}$$

The linearized operator for the right hand side of (2.20) in a neighborhood of (u_1^*, u_2^*) is

$$\begin{pmatrix} d_1\Delta + \alpha u_2^*\Delta + f_{u_1}(u_1^*, u_2^*) & \alpha u_1^*\Delta + f_{u_2}(u_1^*, u_2^*) \\ g_{u_1}(u_1^*, u_2^*) & d_2\Delta + g_{u_2}(u_1^*, u_2^*) \end{pmatrix}.$$

Then the eigenvalues μ of the linearized operator are characterized by $\Xi^*(\mu) = 0$, where σ is one of the eigenvalues of $-\Delta$ with the Neumann boundary condition and

$$\Xi^*(\mu) := \begin{vmatrix} -d_1\sigma - \alpha u_2^*\sigma + f_{u_1}(u_1^*, u_2^*) - \mu & -\alpha u_1^*\sigma + f_{u_2}(u_1^*, u_2^*) \\ g_{u_1}(u_1^*, u_2^*) & -d_2\sigma + g_{u_2}(u_1^*, u_2^*) - \mu \end{vmatrix}.$$

Next recall the reaction-diffusion system (1.5):

$$\begin{cases} v_{1t} = d_1\Delta v_1 + f_1(v_1, v_2, v_3) + k[h_{2\to1}(v_3)v_2 - h_{1\to2}(v_3)v_1], \\ v_{2t} = (d_1 + \alpha M_2)\Delta v_2 + f_2(v_1, v_2, v_3) + k[h_{1\to2}(v_3)v_1 \\ \qquad - h_{2\to1}(v_3)v_2], \\ v_{3t} = d_2\Delta v_3 + f_3(v_1, v_2, v_3) \end{cases} \tag{2.21}$$

for $t > 0$, $x \in \Omega$, where

$$\begin{cases} f_1(v_1, v_2, v_3) = [r_1 - a_1(v_1 + v_2) - b_1 v_3]v_1, \\ f_2(v_1, v_2, v_3) = [r_1 - a_1(v_1 + v_2) - b_1 v_3]v_2, \\ f_3(v_1, v_2, v_3) = [r_2 - b_2(v_1 + v_2) - a_2 v_3]v_3. \end{cases}$$

It is easily seen that

$$\begin{cases} f(v_1 + v_2, v_3) = f_1(v_1, v_2, v_3) + f_2(v_1, v_2, v_3), \\ g(v_1 + v_2, v_3) = f_3(v_1, v_2, v_3), \\ f_{u_1}(v_1 + v_2, v_3) = f_{1,v_i}(v_1, v_2, v_3) \\ \qquad\qquad + f_{2,v_i}(v_1, v_2, v_3) \quad (i = 1, 2), \\ f_{u_2}(v_1 + v_2, v_3) = f_{1,v_3}(v_1, v_2, v_3) + f_{2,v_3}(v_1, v_2, v_3), \\ g_{u_1}(v_1 + v_2, v_3) = f_{3,v_i}(v_1, v_2, v_3) \quad (i = 1, 2), \\ g_{u_2}(v_1 + v_2, v_3) = f_{3,v_3}(v_1, v_2, v_3). \end{cases} \tag{2.22}$$

By using $\rho = v_1 + v_2$, (2.21) can be rewritten into (2.17), that is,

$$\begin{cases} \rho_t = d_1\Delta\rho + \alpha M_2\Delta v_2 + f(\rho, v_3), \\ v_{2t} = (d_1 + \alpha M_2)\Delta v_2 + f_2(\rho - v_2, v_2, v_3) \\ \qquad + k\left[h_{1\to2}(v_3)\rho - (h_{2\to1}(v_3) + h_{1\to2}(v_3))\, v_2\right], \\ v_{3t} = d_2\Delta v_3 + g(\rho, v_3). \end{cases} \tag{2.23}$$

Recall that

$$h_{1\to 2}(s) \equiv (h_{1\to 2}(s) + h_{2\to 1}(s))\,\frac{s}{M_2}. \tag{2.24}$$

We see from (2.24) that (2.23) also possesses the constant equilibrium (ρ^*, v_2^*, v_3^*), where $\rho^* = u_1^*$, $v_3^* = u_2^*$ and

$$v_2^* = \frac{h_{1\to 2}(v_3^*)\rho^*}{h_{1\to 2}(v_3^*) + h_{2\to 1}(v_3^*)} = \frac{v_3^*}{M_2}\rho^* = \frac{u_2^*}{M_2}u_1^*,$$

and that (2.21) possesses the equilibrium solution $(v_1^*, v_2^*, v_3^*) = (\rho^* - v_2^*, v_2^*, v_3^*)$. The linearized eigenvalue problem for (2.21) around (v_1^*, v_2^*, v_3^*) is reduced to $\Xi^k(\mu) = 0$, where

$$\Xi^k := \begin{vmatrix} \xi_{11} & \xi_{12} & \xi_{13} \\ \xi_{21} & \xi_{22} & \xi_{23} \\ f_{3,v_1}(v_1^*, v_2^*, v_3^*) & f_{3,v_2}(v_1^*, v_2^*, v_3^*) & -d_2\sigma + f_{3,v_3}(v_1^*, v_2^*, v_3^*) - \mu \end{vmatrix},$$

$$\xi_{11} := -d_1\sigma + f_{1,v_1}(v_1^*, v_2^*, v_3^*) - kh_{1\to 2}(v_3^*) - \mu,$$

$$\xi_{12} := f_{1,v_2}(v_1^*, v_2^*, v_3^*) + kh_{2\to 1}(v_3^*),$$

$$\xi_{13} := f_{1,v_3}(v_1^*, v_2^*, v_3^*) + k(h'_{2\to 1}(v_3^*)v_2^* - h'_{1\to 2}(v_3^*)v_1^*),$$

$$\xi_{21} := f_{2,v_1}(v_1^*, v_2^*, v_3^*) + kh_{1\to 2}(v_3^*),$$

$$\xi_{22} := -(d_1 + \alpha M_2)\sigma + f_{2,v_2}(v_1^*, v_2^*, v_3^*) - kh_{2\to 1}(v_3^*) - \mu,$$

$$\xi_{23} := f_{2,v_3}(v_1^*, v_2^*, v_3^*) + k(h'_{1\to 2}(v_3^*)v_1^* - h'_{2\to 1}(v_3^*)v_2^*).$$

After some calculation, we can obtain

$$\begin{aligned}\Xi^k(\mu) &= -\frac{M_2 h_{1\to 2}(u_2^*)k}{u_2^*}\,\Xi^*(\mu) + \Xi^0(\mu) \\ &= -(h_{1\to 2}(u_2^*) + h_{2\to 1}(u_2^*))k\,\Xi^*(\mu) + \Xi^0(\mu). \end{aligned} \tag{2.25}$$

We will give the proof of the above equalities in Appendix A.1. It is easy to see that

$$\Xi^*(\mu) = \mu^2 + O(\mu), \qquad \Xi^0(\mu) = -\mu^3 + O(\mu^2)$$

as $|\mu| \to \infty$, which together with (1.10), implies that the eigenvalues of the linearized operator of (2.21) converge to those of (2.20) in a half plane $\{\mu \in \mathbb{C} \mid \mathrm{Re}\,\mu > -\gamma\}$ for an arbitrary positive number γ. The relation (2.25) between Ξ^k and Ξ^* implies two roots of $\Xi^k(\mu) = 0$ are close to those of $\Xi^*(\mu) = 0$ and that the other is negative. Since $(v_1^* + v_2^*, v_3^*)$ is a stable equilibrium point of (2.20) when α vanishes, the spatially constant equilibrium point (v_1^*, v_2^*, v_3^*) is stable for (2.21) if α vanishes. However, if α is large, then $\Xi^*(\mu) = 0$ possesses a positive root because of the cross-diffusion induced instability, which implies that the equilibrium point of (2.21) destabilizes for suitably large k when the diffusivity of v_2 is large enough. We can say that in the approximating reaction-diffusion system (2.21), v_3 plays the role of activator, while v_2 has the role of inhibitor. Thus it turns out that the reaction-diffusion system (2.21) includes the framework of a short range activator-long range inhibitor reaction-diffusion system for (v_3, v_2) and that the cross-diffusion induced instability of (2.20) can be regarded as Turing's instability of (2.21) if k tends to ∞.

3. Competition-diffusion systems and Stefan problems

In this section we present the proof of the convergence of the solution of Problem (Q^k) to the weak solution of Problem (Q) as $k \to \infty$ and show that the strong form of the free boundary Problem (Q) is given by the classical two-phase Stefan Problem (Q^*). In turn these results imply analogous results for the problems (P^k), (P) and the classical two-phase Stefan problem with zero latent heat (P^*) by setting $\lambda = 0$.

Latent heat is absent in Theorems 1.1 and 1.2, and in particular it does not appear in the interface equation. For a new model which converges to the free boundary problem with latent heat (cf. [39,40]), recall Problem (Q^k),

$$\begin{cases} u_{1t} = d_1 \Delta u_1 + f(u_1) - s_1 k u_1 u_2 - \lambda s_1 k(1-w) u_1, & x \in \Omega, t > 0, \\ u_{2t} = d_2 \Delta u_2 + g(u_2) - s_2 k u_1 u_2 - \lambda s_2 k w u_2, & x \in \Omega, t > 0, \\ w_t = k(1-w) u_1 - k w u_2, & x \in \Omega, t > 0, \\ \dfrac{\partial u_1}{\partial \nu} = \dfrac{\partial u_2}{\partial \nu} = 0, & x \in \partial\Omega, t > 0, \\ u(x,0) = u_{01}^k(x),\ u(x,0) = u_{02}^k(x),\ w(x,0) = w_0^k(x), & x \in \Omega, \end{cases} \tag{3.1}$$

where $\lambda \geq 0$. When $\lambda = 0$, (3.1) becomes decoupled and obviously reduces to (2.11). The third variable w can be regarded as an approximation of the characteristic function of the habitat of the species u_1. In order to be able to give an ecological interpretation of this problem, we first suppose that the initial distributions for u_1 and u_2 are completely segregated and impose

$$w(x,0) = W_1(x), \quad x \in \Omega,$$

where $W_1(x) = 1$ if $u_1(x,0) > 0$ and $W_1(x) = 0$ if $u_1(x,0) = 0$.

Letting $k \to \infty$ we can expect that w becomes the characteristic function of $\Omega_1(t)$, and then show that the first and second equations of Problem (Q) can be derived from (3.1). We emphasize that, when $f = g = 0$, the classical two-phase Stefan problem can be derived from (3.1). We also note that the system (3.1) with large k can be regarded as a variant of the penalty methods to solve the two-phase Stefan problem [29].

We are now in a position to interpret (3.1) from an ecological viewpoint. Let us introduce variables $w_1 = w$ and w_2 into (3.1) and rewrite it as

$$\begin{cases} u_{1t} = d_1 \Delta u_1 + f(u_1) - s_1 k u_1 u_2 - \lambda s_1 k w_2 u_1, & x \in \Omega, t > 0, \\ u_{2t} = d_2 \Delta u_2 + g(u_2) - s_2 k u_1 u_2 - \lambda s_2 k w_1 u_2, & x \in \Omega, t > 0, \\ w_{1t} = k w_2 u_1 - k w_1 u_2, & x \in \Omega, t > 0, \\ w_{2t} = -k w_2 u_1 + k w_1 u_2, & x \in \Omega, t > 0. \end{cases} \tag{3.2}$$

The initial condition for w_2 is

$$w_2(x,0) = W_2(x), \quad x \in \Omega,$$

where $W_2(x) = 1$ if $u_2(x,0) > 0$ and $W_2(x) = 0$ if $u_2(x,0) = 0$.

The complete segregation of initial distributions of u_1 and u_2 implies that $W_1(x) + W_2(x) = 1$ for each $x \in \Omega$. Here we assume that the initial segregating boundary $\Gamma(0)$ is a smooth hypersurface in Ω. Obviously $(w_1 + w_2)_t = 0$, so that

$$w_1(x,t) + w_2(x,t) = w_1(x,0) + w_2(x,0) = W_1(x) + W_2(x) = 1,$$

and it turns out that (3.2) coincides with (3.1). The system (3.2) can be interpreted as follows: u_1 and u_2 are the densities of two competing species with their own habitats $\Omega_1(t)$ and $\Omega_2(t)$, whose shapes are respectively described by the characteristic-like functions w_1 and w_2 (in fact, as k tends to ∞, both of them become the corresponding characteristic functions of the habitats $\Omega_1(t)$ and $\Omega_2(t)$). There are two different types of interaction between u_1 and u_2. One is the directly competitive interaction (due to the term uv), for obtaining their common resource. The other one is the struggle interaction (due to $\lambda s_1 w_2 u_1$ and $\lambda s_2 w_1 u_2$), for constructing their own habitats, where λs_1 (resp. λs_2) is the cost rate when u_1 (resp. u_2) attacks the habitat $\Omega_2(t)$ (resp. $\Omega_1(t)$). For this reason, we may say that (3.2) is not a conventional competition-diffusion model, but a new reaction-diffusion model of two competing species which move by diffusion.

In order to be able to state our main result, we impose the following assumptions on f, g and the initial datum (u_{01}, u_{02}, w_0).

A1 (Assumption on f and g) There exist C^1-functions $\tilde{f}(u)$ and $\tilde{g}(u)$ and positive constants K_1 and K_2 such that

$$f(u) = \tilde{f}(u)u, \qquad g(u) = \tilde{g}(u)u,$$
$$\tilde{f}(u) \le 0 \quad \text{for } u \ge K_1, \qquad \tilde{g}(u) \le 0 \quad \text{for } u \ge K_2.$$

A2 (Assumption on the initial datum)

$$(u_{01}^k, u_{02}^k, w_0^k) \in C(\overline{\Omega}) \times C(\overline{\Omega}) \times L^\infty(\Omega),$$
$$0 \le u_{01}^k \le \alpha, 0 \le u_{02}^k \le \beta, 0 \le w_0 \le 1$$

for some positive constants α and β.

Define $\Omega_r := \{x \in \Omega \mid B(x, 2r) \subset \Omega\}$; for each $r \in (0, \hat{r})$, there exists a positive function $\vartheta(\xi)$ satisfying

$$\lim_{|\xi|\to 0} \vartheta(\xi) = 0 \text{ such that}$$

$$\int_{\Omega_r} |u_{01}^k(x+\xi) - u_{01}^k(x)|\, \mathrm{d}x \le \vartheta(\xi),$$

$$\int_{\Omega_r} |u_{02}^k(x+\xi) - u_{02}^k(x)|\, \mathrm{d}x \le \vartheta(\xi),$$

$$\int_{\Omega_r} |w_0^k(x+\xi) - w_0^k(x)|\, \mathrm{d}x \le \vartheta(\xi), \text{ and where}$$

$u_{01}^k \to u_{01}, u_{02}^k \to u_{02}, w_0^k \to w_0$ strongly in $L^2(\Omega)$ as $k \to \infty$, for some functions $u_{01}, u_{02}, w_0 \in L^\infty(\Omega)$.

We do not assume $u_{01}u_{02} = (1-w_0)u_{01} = w_0 u_{02} = 0$. In particular, we do not impose that the supports of u_{01} and u_{02} are disjoint.

REMARK 3.1. The parameter λ in Problem (Q) corresponds to the latent heat in the Stefan problem. For sufficiently large k, one can expect that u_1 and u_2 exhibit corner layers on the interface $\Gamma(t)$, while w has a sharp transition layer, which clearly indicates a segregating boundary between u and v.

3.1. *Some basic properties*

In this section **A1** and **A2** are always assumed. By a solution of (3.1) in Q_T $(T > 0)$ we mean a triple of functions $(u_1^k, u_2^k, w^k) \in C([0,T]; C(\overline{\Omega}) \times C(\overline{\Omega}) \times L^\infty(\Omega))$ such that

$$u_1^k, u_2^k \in C^1((0,T]; C(\overline{\Omega})) \cap C((0,T]; W^{2,p}(\Omega)),$$
$$w^k \in C^1([0,T]; L^\infty(\Omega))$$

for each $p \in (1,\infty)$ and (u_1, u_2, w) satisfies Eq. (3.1). There exists a positive number $T = T(\|u_{01}^k\|_{C(\overline{\Omega})}, \|u_{02}^k\|_{C(\overline{\Omega})}, \|w_0^k\|_{L^\infty(\Omega)})$ such that (3.1) possesses a unique solution (u_1^k, u_2^k, w^k) in Q_T. This does not immediately follow from the theory of analytic semigroups because of the lack of a diffusion term for w. In particular, if w_0^k is discontinuous at some point, w^k as well as Δu_1^k and Δu_2^k may be discontinuous at that point at later times as well. We refer to [18, Appendix] for the proof. We denote by (u_1^k, u_2^k, w^k) a solution of (3.1) in Q_T.

LEMMA 3.2. *Let (u_1^k, u_2^k, w^k) satisfy Problem (P^k). Then :*

$$0 \le u_1^k(x,t) \le \max\{\alpha, K_1\}, \qquad 0 \le u_2^k(x,t) \le \max\{\beta, K_2\},$$
$$0 \le w^k(x,t) \le 1$$

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

PROOF. We deduce from the maximum principle that $u_1^k, u_2^k \ge 0$. Let $x \in \Omega$ be such that $w_0^k(x)$ is defined. Then the condition $0 \le w_0^k(x) \le 1$ implies that $0 \le w^k(x,t) \le 1$ for all

$t > 0$. Indeed, suppose that at a time $t = \bar{t}$, $w^k(x, \bar{t}) = 0$, then $w_t^k(x, \bar{t}) \geq 0$; similarly if at a time $\tilde{t}$, $w^k(x, \tilde{t}) = 1$, then $w_t^k(x, \tilde{t}) \leq 0$. Finally, we apply again the maximum principle to deduce that $u_1^k \leq \max\{\alpha, K_1\}$ and $u_2^k \leq \max\{\beta, K_2\}$. □

Without loss of generality, we can assume that

$$\alpha \geq K_1 \quad \text{and} \quad \beta \geq K_2$$

by choosing α and β so large that the above inequalities hold. The local existence of the solutions and Lemma 3.2 ensure that the solution (u_1^k, u_2^k, w^k) exists globally in time and satisfies

$$\begin{aligned} &0 \leq u_1^k(x,t) \leq \alpha, \qquad 0 \leq u_2^k(x,t) \leq \beta, \\ &\qquad 0 \leq w^k(x,t) \leq 1 \quad \text{for } (x,t) \in \overline{\Omega} \times [0,\infty). \end{aligned} \tag{3.3}$$

Set

$$M_f := \max\{f(u) \mid 0 \leq u \leq \alpha\}, \qquad M_g := \max\{g(u) \mid 0 \leq u \leq \beta\}.$$

LEMMA 3.3. *For any positive number T there exist positive constants C_i $(i = 1, \ldots, 5)$ independent of k and λ such that*

$$\begin{aligned} &\iint_{Q_T} (s_1 + s_2) u_1^k u_2^k \mathrm{d}x\mathrm{d}t \leq \frac{C_1}{k}, \\ &\iint_{Q_T} \lambda s_1 (1 - w^k) u_1^k \mathrm{d}x\mathrm{d}t \leq \frac{C_2}{k}, \\ &\iint_{Q_T} \lambda s_2 w^k u_2^k \mathrm{d}x\mathrm{d}t \leq \frac{C_3}{k}, \\ &\iint_{Q_T} d_1 |\nabla u_1^k|^2 \mathrm{d}x\mathrm{d}t \leq C_4, \\ &\iint_{Q_T} d_2 |\nabla u_2^k|^2 \mathrm{d}x\mathrm{d}t \leq C_5. \end{aligned}$$

In this Chapter, positive constants which do not depend on k are denoted by C_i for simplicity of notation.

PROOF. Integration of the equation for u in Q_T yields

$$\begin{aligned} &\iint_{Q_T} \left(s_1 k u_1^k u_2^k + \lambda s_1 k (1 - w^k) u_1^k \right) \mathrm{d}x\mathrm{d}t \\ &\qquad = \int_\Omega (u_{01}^k(x) - u_1^k(x,T)) \mathrm{d}x + \iint_{Q_T} f(u_1^k) \mathrm{d}x\mathrm{d}t \\ &\qquad \leq (\alpha + T M_f)|\Omega|, \end{aligned}$$

which implies the second estimate. The first and third ones can be shown similarly. Next we multiply the equation for u_1^k by u_1^k and integrate by parts on Ω. This yields

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega (u_1^k)^2\mathrm{d}x + d_1\int_\Omega |\nabla u_1^k|^2\mathrm{d}x + \int_\Omega \left(s_1 k(u_1^k)^2 u_2^k + \lambda s_1 k(1-w^k)(u_1^k)^2\right)\mathrm{d}x \le |\Omega|\alpha M_f,$$

which we integrate on $(0, T)$ to deduce the fourth estimate. The last estimate can be proved similarly. □

LEMMA 3.4. *Let T be any positive number and set*

$$\Omega_\xi := \{x \in \Omega | x + r\xi \in \Omega \text{ for } 0 \le r \le 1\}$$

with $\xi \in \mathbb{R}^N$. Then there exist positive constants C_6 and C_7 independent of k and λ such that

$$\int_0^T \int_{\Omega_\xi} (u_1^k(x+\xi, t) - u_1^k(x,t))^2\mathrm{d}x\mathrm{d}t \le \frac{C_4}{d_1}|\xi|^2, \tag{3.4}$$

$$\int_0^T \int_{\Omega_\xi} (u_2^k(x+\xi, t) - u_2^k(x,t))^2\mathrm{d}x\mathrm{d}t \le \frac{C_5}{d_2}|\xi|^2, \tag{3.5}$$

$$\int_0^{T-\tau} \int_{\Omega} (u_1^k(x, t+\tau) - u_1^k(x,t))^2\mathrm{d}x\mathrm{d}t \le C_6\tau, \tag{3.6}$$

$$\int_0^{T-\tau} \int_{\Omega} (u_2^k(x, t+\tau) - u_2^k(x,t))^2\mathrm{d}x\mathrm{d}t \le C_7\tau, \tag{3.7}$$

for $\xi \in \mathbb{R}^N$ and $\tau > 0$.

PROOF. The first and second inequalities (3.4), (3.5) follow immediately from Lemma 3.3. Indeed, we have

$$\int_0^T \int_{\Omega_\xi} (u_1^k(x+\xi, t) - u_1^k(x,t))^2\mathrm{d}x\mathrm{d}t = \int_0^T \int_{\Omega_\xi} \left\{\int_0^1 \nabla u_1^k(x + r\xi, t)\cdot\xi\,\mathrm{d}r\right\}^2 \mathrm{d}x\mathrm{d}t \le \frac{C_4}{d_1}|\xi|^2.$$

The second one can be shown in a similar way. Next we prove (3.6) and (3.7). We have that

$$\int_0^{T-\tau} \int_{\Omega} (u_1^k(x, t+\tau) - u_1^k(x,t))^2\mathrm{d}x\mathrm{d}t = \int_0^{T-\tau} \int_{\Omega} (u_1^k(x, t+\tau) - u_1^k(x,t)) \int_0^\tau u_t(x, t+r)\mathrm{d}r\mathrm{d}x\mathrm{d}t$$

$$
\begin{aligned}
&= \int_0^{T-\tau} \int_\Omega (u_1^k(x,t+\tau) - u_1^k(x,t)) \int_0^\tau \{d_1 \Delta u_1^k(x,t+r) \\
&\quad + f(u_1^k(x,t+r)) - s_1 k u_1^k(x,t+r) u_2^k(x,t+r) \\
&\quad - \lambda s_1 k(1 - w^k(x,t+r)) u_1^k(x,t+r)\} \mathrm{d}r \mathrm{d}x \mathrm{d}t.
\end{aligned}
$$

We estimate the three terms on the right-hand side. The first term can be estimated as follows:

$$
\begin{aligned}
&\left| \int_0^{T-\tau} \int_\Omega (u_1^k(x,t+\tau) - u_1^k(x,t)) \int_0^\tau d_1 \Delta u_1^k(x,t+r) \mathrm{d}r \mathrm{d}x \mathrm{d}t \right| \\
&= d_1 \left| \int_0^\tau \int_0^{T-\tau} \int_\Omega (\nabla u_1^k(x,t+\tau) - \nabla u_1^k(x,t)) \cdot \nabla u_1^k(x,t+r) \mathrm{d}x \mathrm{d}t \mathrm{d}r \right| \\
&\le 2 d_1 \tau \int_0^T \int_\Omega |\nabla u_1^k(x,t)|^2 \mathrm{d}x \mathrm{d}t \\
&\le 2 C_4 \tau.
\end{aligned}
$$

Secondly, we see that

$$
\begin{aligned}
&\left| \int_0^{T-\tau} \int_\Omega (u_1^k(x,t+\tau) - u_1^k(x,t)) \int_0^\tau f(u_1^k(x,t+r)) \mathrm{d}r \mathrm{d}x \mathrm{d}t \right| \\
&\le \alpha M_f T |\Omega| \tau.
\end{aligned}
$$

Finally, we have

$$
\begin{aligned}
&\left| \int_0^{T-\tau} \int_\Omega (u_1^k(x,t+\tau) - u_1^k(x,t)) \times \int_0^\tau s_1 k u_1^k(x,t+r) u_2^k(x,t+r) \right. \\
&\quad \left. + \lambda s_1 k (1 - w^k(x,t+r)) u_1^k(x,t+r) \mathrm{d}r \mathrm{d}x \mathrm{d}t \right| \\
&\le \alpha \int_0^\tau \int_0^{T-\tau} \int_\Omega s_1 k u_1^k(x,t+r) u_2^k(x,t+r) \\
&\quad + \lambda s_1 k (1 - w^k(x,t+r)) u_1^k(x,t+r) \mathrm{d}x \mathrm{d}t dr \\
&\le \alpha \tau \int_0^T \int_\Omega s_1 k u_1^k(x,t) u_2^k(x,t) + \lambda s_1 k (1 - w^k(x,t)) u_1^k(x,t) \mathrm{d}x \mathrm{d}t \\
&\le \alpha (C_1 + C_2) \tau.
\end{aligned}
$$

Thus we have shown that

$$
\begin{aligned}
&\int_0^{T-\tau} \int_\Omega (u_1^k(x,t+\tau) - u_1^k(x,t))^2 \mathrm{d}x \mathrm{d}t \\
&\le (2 C_4 + \alpha M_f T |\Omega| + \alpha C_1 + \alpha C_2) \tau.
\end{aligned}
$$

Similarly, we can prove the following estimate:

$$\int_0^{T-\tau}\int_\Omega (u_2^k(x,t+\tau)-u_2^k(x,t))^2 \mathrm{d}x\mathrm{d}t \\ \leq (2C_5+\beta M_g T|\Omega|+\beta C_1+\beta C_3)\tau.$$

This completes the proof of Lemma 3.4. □

3.2. *The limiting problem*

Choose a positive number T arbitrarily and fix it. We deduce from Lemmas 3.2 and 3.3 that the families $\{u_1^k\}$ and $\{u_2^k\}$ are bounded in $L^2(0,T;H^1(\Omega))$ and that the family $\{w^k\}$ is bounded in $L^\infty(Q_T)$. In order to show that the sequences $\{u_1^k\}$ and $\{u_2^k\}$ are relatively compact in $L^2(Q_T)$ we will apply the following Fréchet–Kolmogorov Theorem (e.g. [5, Corollary IV.26, p. 74]).

PROPOSITION 3.5 (*Fréchet–Kolmogorov*). *Let $\mathcal{F}$ be a bounded subset of $L^p(Q_T)$ with $1 \leq p < \infty$. Suppose that*

(i) *for any $\epsilon > 0$ and any subset $\omega \Subset Q_T$, there exists a positive constant δ (<dist$(\omega, \partial Q_T)$) such that*

$$\|f(x+\xi,t)-f(x,t)\|_{L^p(\omega)}+\|f(x,t+\tau)-f(x,t)\|_{L^p(\omega)}<\epsilon$$

for all ξ, τ, and $f \in \mathcal{F}$ satisfying $|\xi| + |\tau| < \delta$,

(ii) *for any $\epsilon > 0$, there exists $\omega \Subset Q_T$ such that*

$$\|f\|_{L^p(Q_T\setminus\omega)}<\epsilon$$

for all $f \in \mathcal{F}$.

Then $\mathcal{F}$ is precompact in $L^p(Q_T)$.

This proposition implies that the families $\{u_1^k\}$ and $\{u_2^k\}$ are precompact in $L^2(Q_T)$. Therefore there exist subsequences $\{u_1^{k_n}\}$ and $\{u_2^{k_n}\}$ and $\{w^{k_n}\}$ and functions $u_1, u_2 \in L^2(0,T;H^1(\Omega))$ and $w \in L^2(Q_T)$ such that

$$u_1^{k_n} \longrightarrow u_1, \qquad u_2^{k_n} \longrightarrow u_2 \quad \text{strongly in } L^2(Q_T), \\ \text{weakly in } L^2(0,T;H^1(\Omega)) \text{ and a.e. in } Q_T, \tag{3.8}$$

and

$$w^{k_n} \rightharpoonup w \quad \text{weakly in } L^2(Q_T) \tag{3.9}$$

as $k_n \to 0$. It follows from (3.3) that

$$u_1 \geq 0, \qquad u_2 \geq 0, \qquad 0 \leq w \leq 1 \quad \text{on } \overline{Q_T}. \tag{3.10}$$

Hence we deduce from Lemma 3.3 that

$$u_1 u_2 = (1-w)u_1 = w u_2 = 0. \tag{3.11}$$

In the sequel we will show that the functions u_1, u_2, and w are determined as the unique solution of the Stefan-like problem (Q).

However, before doing so, we prove that w^{k_n} converges strongly to its limit as $k_n \to \infty$ (see [19]). A key idea of the proof is that if we would consider Problem (Q^k) with $f = g = 0$, then the quantity

$$\frac{1}{s_1}\int_\Omega |u_1^k(x,t) - \widetilde{u_1^k}(x,t)|\mathrm{d}x + \frac{1}{s_2}\int_\Omega |u_2^k(x,t) - \widetilde{u_2^k}(x,t)|\mathrm{d}x$$
$$+\lambda \int_\Omega |w^k(x,t) - \widetilde{w^k}(x,t)|\mathrm{d}x$$

with (u_1^k, u_2^k, w^k) and $(\widetilde{u_1^k}, \widetilde{u_2^k}, \widetilde{w^k})$, two solution orbits, would decrease in time.

We set

$$\Omega_r := \{x \in \Omega \mid B(x, 2r) \subset \Omega\}.$$

LEMMA 3.6. *For each $r \in (0, \hat{r})$, there exists a positive function $\widetilde{\vartheta}(\xi)$ satisfying*

$$\lim_{|\xi|\to 0} \widetilde{\vartheta}(\xi) = 0$$

and

$$\int_0^T \int_{\Omega_{3r}} |w^k(x+\xi, t) - w^k(x,t)|\,\mathrm{d}x\,\mathrm{d}t \le \widetilde{\vartheta}(\xi)$$

for all $\xi \in \mathbb{R}^N$, $|\xi| \le r$.

We give the proof of this lemma in Appendix A.2.

LEMMA 3.7. *There exists a positive constant C such that*

$$\int_0^{T-\tau} \int_\Omega |w^k(x, t+\tau) - w^k(x,t)|\,\mathrm{d}x\,\mathrm{d}t \le C\tau$$

for all $\tau \in (0, T)$.

PROOF. We have

$$\int_0^{T-\tau} \int_\Omega |w^k(x, t+\tau) - w^k(x,t)|\,\mathrm{d}x\,\mathrm{d}t$$
$$= \int_0^{T-\tau} \int_\Omega \left| \int_0^\tau w_t^k(x, t+\sigma)\mathrm{d}\sigma \right| \mathrm{d}x\,\mathrm{d}t$$

$$
\begin{aligned}
&= k \int_0^{T-\tau} \int_\Omega \left| \int_0^\tau ((1-w^k)u_1^k - u_2^k w^k)(x, t+\sigma) \mathrm{d}\sigma \right| \mathrm{d}x \, \mathrm{d}t \\
&\le k \int_0^{T-\tau} \int_\Omega \int_0^\tau ((1-w^k)u_1^k + u_2^k w^k)(x, t+\sigma) \mathrm{d}\sigma \, \mathrm{d}x \, \mathrm{d}t.
\end{aligned}
$$

Lemma 3.3 immediately implies

$$
\begin{aligned}
&\int_0^{T-\tau} \int_\Omega |w^k(x, t+\tau) - w^k(x, t)| \, \mathrm{d}x \, \mathrm{d}t \\
&\quad \le \int_0^\tau k \left(\iint_{Q_T} ((1-w^k)u_1^k + u_2^k w^k)(x, t) \, \mathrm{d}x \, \mathrm{d}t \right) \mathrm{d}\sigma \\
&\quad \le C\tau,
\end{aligned}
$$

which completes the proof. □

Applying the Fréchet–Kolmogorov Theorem (Proposition 3.5) to the sequence $\{w^k\}_{k \ge k_0}$, we deduce its relative compactness in $L^1(Q_T)$.

LEMMA 3.8. *Let T be an arbitrary positive number. The functions u_1, u_2, and w given in (3.8) and (3.9) satisfy*

$$
\begin{aligned}
&\iint_{Q_T} \left\{ \left(\frac{u_1}{s_1} - \frac{u_2}{s_2} + \lambda w \right) \zeta_t - \nabla \left(\frac{d_1 u_1}{s_1} - \frac{d_2 u_2}{s_2} \right) \cdot \nabla \zeta \right. \\
&\quad \left. + \left(\frac{f(u_1)}{s_1} - \frac{g(u_2)}{s_2} \right) \zeta \right\} \mathrm{d}x \mathrm{d}t \\
&\qquad = - \int_\Omega \left(\frac{u_{01}}{s_1} - \frac{u_{02}}{s_2} + \lambda w_0 \right) \zeta(x, 0) \mathrm{d}x
\end{aligned}
\tag{3.12}
$$

for all functions $\zeta \in C^\infty(\overline{Q_T})$, such that $\zeta(x, T) = 0$.

PROOF. We deduce from the equations in Problem (Q^k) that

$$
\left(\frac{u_1^k}{s_1} - \frac{u_2^k}{s_2} + \lambda w^k \right)_t = \frac{d_1 \Delta u_1^k}{s_1} - \frac{d_2 \Delta u_2^k}{s_2} + \frac{f(u_1^k)}{s_1} - \frac{g(u_2^k)}{s_2},
$$

multiply it by a test function $\zeta \in C^\infty(\overline{Q_T})$ with $\zeta(x, T) = 0$ and integrate by parts to obtain the identity

$$
\begin{aligned}
&\iint_{Q_T} \left\{ - \left(\frac{u_1^k}{s_1} - \frac{u_2^k}{s_2} + \lambda w^k \right) \zeta_t + \nabla \left(\frac{d_1 u_1^k}{s_1} - \frac{d_2 u_2^k}{s_2} \right) \cdot \nabla \zeta \right. \\
&\quad \left. - \left(\frac{f(u_1^k)}{s_1} - \frac{g(u_2^k)}{s_2} \right) \zeta \right\} \mathrm{d}x \mathrm{d}t \\
&\qquad = \int_\Omega \left(\frac{u_{01}^k}{s_1} - \frac{u_{02}^k}{s_2} + \lambda w_0^k \right) \zeta(x, 0) \mathrm{d}x.
\end{aligned}
$$

Letting $k = k_n \to 0$, we deduce (3.12). □

We will formulate (3.12) in a weak form corresponding to Problem (Q). We use the notation :

$$H(r) := \begin{cases} 1 & (r > 0), \\ [0, 1] & (r = 0), \\ 0 & (r < 0). \end{cases}$$

LEMMA 3.9. *If $w \in H(z)$, then $\varphi_\lambda(z + \lambda w) = z$. In particular, the functions u_1, u_2 and w, which are given in (3.8) and (3.9), satisfy*

$$u_1 = s_1 \varphi_\lambda(z)^+, \qquad u_2 = s_2 \varphi_\lambda(z)^-, \quad \textit{and} \quad w = \frac{z - \varphi_\lambda(z)}{\lambda}, \tag{3.13}$$

where

$$z := \frac{u_1}{s_1} - \frac{u_2}{s_2} + \lambda w. \tag{3.14}$$

PROOF. The first claim of this lemma follows from the definitions of φ_λ and H. We can deduce from (3.10) and (3.11) that

$$w \in H\left(\frac{u_1}{s_1} - \frac{u_2}{s_2}\right).$$

Hence we have

$$\varphi_\lambda(z) = \frac{u_1}{s_1} - \frac{u_2}{s_2},$$

which, together with (3.10) and (3.11), implies (3.13). □

DEFINITION 3.10. A function $z \in L^\infty(Q_T)$ is a weak solution of Problem (Q) with an initial datum $z_0 \in L^\infty(\Omega)$ if

$$d(\varphi_\lambda(z)) \in L^2(0, T; H^1(\Omega))$$

and

$$\iint_{Q_T} z\zeta_t \mathrm{d}x\mathrm{d}t + \int_\Omega z_0(x)\zeta(x, 0)\mathrm{d}x = \iint_{Q_T} \{\nabla d(\varphi_\lambda(z)) \cdot \nabla\zeta - h(\varphi_\lambda(z))\zeta\}\mathrm{d}x\mathrm{d}t \tag{3.15}$$

for all functions $\zeta \in C^\infty(\overline{Q_T})$ such that $\zeta(x, T) = 0$.

REMARK 3.11. It follows from [10] that if z is a weak solution of Problem (Q) and if $\varphi_\lambda(z_0) \in C(\overline{\Omega})$, then $d(\varphi_\lambda(z))$ and thus $\varphi_\lambda(z)$ is uniformly continuous on $\overline{Q_T}$.

It is known that the classical two-phase Stefan problem under the Neumann condition can be formulated as the nonlinear system (Q) with $h \equiv 0$. Then z and $\varphi_\lambda(z)$ correspond to the internal energy and the temperature, respectively. We note that (Q) can also deal with the case where the interface fattens. We set

$$\begin{cases} \Omega_+(t) := \{x \in \Omega | \varphi_\lambda(z(x,t)) > 0\}, \\ \Omega_-(t) := \{x \in \Omega | \varphi_\lambda(z(x,t)) < 0\}, \\ \Gamma(t) := \Omega \setminus (\Omega_+(t) \cup \Omega_-(t)) \end{cases} \tag{3.16}$$

for $t \in [0, T]$ and also use the notation

$$\begin{cases} \Omega_+ := \bigcup_{0 \le t \le T} \Omega_+(t) \times \{t\}, \\ \Omega_- := \bigcup_{0 \le t \le T} \Omega_-(t) \times \{t\}, \\ \Gamma := \bigcup_{0 \le t \le T} \Gamma(t) \times \{t\}. \end{cases}$$

We could think that $\Omega_+(t)$ and $\Omega_-(t)$ symbolize two distinct phases, and $\Gamma(t)$ represents a *phase boundary* (or an *interface*) at time t.

LEMMA 3.12. *The function z defined by* (3.14) *coincides with the unique weak solution of Problem (Q) with initial datum $z_0 = u_{01}/s_1 - u_{02}/s_2 + \lambda w_0$.*

PROOF. It follows from Lemma 3.2 that $z \in L^\infty(Q_T)$. We observe that (3.13) implies

$$d(\varphi_\lambda(z)) = \frac{d_1 u_1}{s_1} - \frac{d_2 u_2}{s_2}.$$

In particular, $d(\varphi_\lambda(z)) \in L^2(0, T; H^1(\Omega))$ holds true by Lemma 3.3. We also notice that

$$h(\varphi_\lambda(z)) = \frac{f(u_1)}{s_1} - \frac{g(u_2)}{s_2}.$$

Therefore (3.12) can be rewritten as (3.15) in terms of z and $z_0 = u_{01}/s_1 - u_{02}/s_2 + \lambda w_0$. This completes the proof of this lemma. □

The uniqueness of the weak solution of the Stefan problem (Q) for $z_0 \in L^1(\Omega)$ follows from Hilhorst, Mimura and Schätzle [14]. Thus (u_{01}, u_{02}, w_0) uniquely determines z, which uniquely gives (u_1, u_2, w) by (3.13). Consequently we have proved the following result, which is a variant of Theorem 1.3.

THEOREM 3.13. *The function z defined by* (3.14) *is the unique weak solution of the Stefan problem (Q) with an initial datum $u_{01}/s_1 - u_{02}/s_2 + \lambda w_0$. As $k \to \infty$,*

$$u_1^k \longrightarrow u_1, \quad u_2^k \longrightarrow u_2 \quad \textit{strongly in } L^2(Q_T)$$

and weakly in $L^2(0, T; H^1(\Omega))$,

$$w^k \longrightarrow w \quad \textit{strongly in } L^2(Q_T).$$

Finally we state a result about the relation between (Q) and (Q^k).

THEOREM 3.14. *Let z be the unique weak solution of (Q) with initial datum z_0 and let $\Omega_\pm(t)$ and $\Gamma(t)$ be the sets defined by (3.16). Suppose that (each component of) $\Gamma(t)$ is a smooth, closed and orientable hypersurface satisfying $\Gamma(t) \cap \partial\Omega = \emptyset$ for all $t \in [0, T]$. Let n be the unit normal vector on $\Gamma(t)$ oriented from $\Omega_+(t)$ to $\Omega_-(t)$. Also assume that $\Gamma(t)$ smoothly moves with a velocity V_n in the direction of n and that the functions*

$$u_1 := s_1\varphi_\lambda(z)^+, \quad \textit{and} \quad u_2 := s_2\varphi_\lambda(z)^-$$

are smooth on $\overline{\Omega_+}$ and $\overline{\Omega_-}$, respectively. Then (Γ, u_1, u_2) satisfies

$$\begin{cases} u_{1t} = d_1\Delta u_1 + f(u_1), & \textit{in } \Omega_+(t), \\ u_{2t} = d_2\Delta u_2 + g(u_2), & \textit{in } \Omega_-(t), \\ \lambda V_n = -\dfrac{d_1}{s_1}\dfrac{\partial u_1}{\partial n} - \dfrac{d_2}{s_2}\dfrac{\partial u_2}{\partial n}, & \textit{on } \Gamma(t), \\ u_1 = u_2 = 0, & \textit{on } \Gamma(t), \\ \dfrac{\partial u_1}{\partial \nu} = \dfrac{\partial u_2}{\partial \nu} = 0, & \textit{on } \partial\Omega, \end{cases}$$

for $t \in (0, T]$ and

$$\begin{cases} \Gamma(0) = \{x \in \Omega | \varphi_\lambda(z_0(x)) = 0\}, \\ u_1(x, 0) = s_1(\varphi_\lambda(z_0(x)))^+, \quad u_2(x, 0) = s_2(\varphi_\lambda(z_0(x)))^- \quad x \in \Omega. \end{cases} \tag{3.17}$$

Here $\partial u_1/\partial n$ (resp. $\partial u_2/\partial n$) is regarded as a boundary value on $\partial\Omega_+(t)$ (resp. $\partial\Omega_-(t)$).

PROOF. It follows from the definition of u_1 and u_2 that

$$u_1 = u_2 = 0 \quad \text{on } \Gamma,$$
$$\varphi_\lambda(z) = z - \lambda = \frac{u_1}{s_1} \quad \text{in } \Omega_+, \qquad \varphi_\lambda(z) = z = -\frac{u_2}{s_2} \quad \text{in } \Omega_-,$$

which are used in the computations below.

Next we derive the parabolic equations for u_1 and u_2 as well as the Stefan condition on the interface Γ. First we rewrite the first term on the left hand-side of (3.15). We see that

$$\iint_{Q_T} z\zeta_t \mathrm{d}x\mathrm{d}t = \int_0^T \left\{ \int_{\Omega_+(t)} \left(\frac{u_1}{s_1} + \lambda w \right) \zeta_t \mathrm{d}x \right.$$
$$\left. - \int_{\Omega_-(t)} \frac{u_2}{s_2} \zeta_t \mathrm{d}x \right\} \mathrm{d}t. \tag{3.18}$$

Since $w = 1$ in $\Omega_+(t)$,

$$\iint_{Q_T} z\zeta_t \mathrm{d}x\mathrm{d}t = \int_0^T \left\{ \int_{\Omega_+(t)} \left(\frac{u_1}{s_1} + \lambda \right) \zeta_t \mathrm{d}x - \int_{\Omega_-(t)} \frac{u_2}{s_2} \zeta_t \mathrm{d}x \right\} \mathrm{d}t.$$

It follows from $u_1|_{\Gamma(t)} = 0$ that

$$\left[\int_{\Omega_+(t)} u_1 \zeta \mathrm{d}x \right]_{t=0}^{t=T} = \int_0^T \frac{\mathrm{d}}{\mathrm{d}t} \left(\int_{\Omega_+(t)} u_1 \zeta \mathrm{d}x \right) \mathrm{d}t$$
$$= \int_0^T \int_{\Omega_+(t)} (u_1 \zeta_t + u_{1t} \zeta) \mathrm{d}x\mathrm{d}t;$$

moreover,

$$\left[\int_{\Omega_+(t)} \zeta \mathrm{d}x \right]_{t=0}^{t=T} = \int_0^T \int_{\Omega_+(t)} \zeta_t \mathrm{d}x\mathrm{d}t + \int_0^T \int_{\Gamma(t)} V_n \zeta \mathrm{d}\sigma \mathrm{d}t,$$

and similarly since $u_2 \,|_{\Gamma(t)} = 0$, we have that

$$\left[\int_{\Omega_-(t)} u_2 \zeta \mathrm{d}x \right]_{t=0}^{t=T} = \int_0^T \int_{\Omega_-(t)} (u_2 \zeta_t + u_{2t} \zeta) \mathrm{d}x\mathrm{d}t.$$

Therefore we have that for test functions ζ which vanish at $t = 0$ and $t = T$,

$$\iint_{Q_T} z\zeta_t \mathrm{d}x\mathrm{d}t = -\frac{1}{s_1} \int_0^T \int_{\Omega_+(t)} u_{1t} \zeta \mathrm{d}x\mathrm{d}t + \frac{1}{s_2} \int_0^T \int_{\Omega_-(t)} u_{2t} \zeta \mathrm{d}x\mathrm{d}t$$
$$- \int_0^T \int_{\Gamma(t)} \lambda V_n \zeta \mathrm{d}\sigma \mathrm{d}t. \tag{3.19}$$

On the other hand we have that

$$\iint_{Q_T} \{ -\nabla d(\varphi_\lambda(z)) \cdot \nabla \zeta + h(\varphi_\lambda(z)) \zeta \} \mathrm{d}x\mathrm{d}t$$
$$= -\frac{1}{s_1} \int_0^T \int_{\Omega_+(t)} d_1 \nabla u_1 \cdot \nabla \zeta \mathrm{d}x\mathrm{d}t + \frac{1}{s_2} \int_0^T \int_{\Omega_-(t)} d_2 \nabla u_2 \cdot \nabla \zeta \mathrm{d}x\mathrm{d}t$$
$$+ \frac{1}{s_1} \int_0^T \int_{\Omega_+(t)} f(u_1) \zeta \mathrm{d}x\mathrm{d}t - \frac{1}{s_2} \int_0^T \int_{\Omega_-(t)} g(u_2) \zeta \mathrm{d}x\mathrm{d}t$$
$$= \int_0^T \int_{\Gamma(t)} \left\{ -\frac{d_1}{s_1} \frac{\partial u_1}{\partial n} - \frac{d_2}{s_2} \frac{\partial u_2}{\partial n} \right\} \zeta \mathrm{d}\sigma \mathrm{d}t$$
$$+ \int_0^T \int_{\partial\Omega} \left\{ -\frac{d_1}{s_1} \frac{\partial u_1}{\partial \nu} + \frac{d_2}{s_2} \frac{\partial u_2}{\partial \nu} \right\} \zeta \mathrm{d}\sigma \mathrm{d}t$$

$$+\frac{1}{s_1}\int_0^T\int_{\Omega_+(t)}\{d_1\Delta u_1+f(u_1)\}\zeta\,\mathrm{d}x\mathrm{d}t$$
$$-\frac{1}{s_2}\int_0^T\int_{\Omega_-(t)}\{d_2\Delta u_2+g(u_2)\}\zeta\,\mathrm{d}x\mathrm{d}t. \tag{3.20}$$

Next we substitute (3.19) and (3.20) into (3.15). This gives

$$\begin{aligned}&\int_0^T\int_{\Gamma(t)}\left\{-\lambda V_n-\frac{d_1}{s_1}\frac{\partial u_1}{\partial n}-\frac{d_2}{s_2}\frac{\partial u_2}{\partial n}\right\}\zeta\,\mathrm{d}\sigma\mathrm{d}t\\&\quad+\frac{1}{s_1}\int_0^T\int_{\Omega_+(t)}\{-u_{1t}+d_1\Delta u_1+f(u_1)\}\zeta\,\mathrm{d}x\mathrm{d}t\\&\quad+\frac{1}{s_2}\int_0^T\int_{\Omega_-(t)}\{u_{2t}-d_2\Delta u_2-g(u_2)\}\zeta\,\mathrm{d}x\mathrm{d}t\\&\quad+\int_0^T\int_{\partial\Omega}\left\{-\frac{d_1}{s_1}\frac{\partial u_1}{\partial \nu}+\frac{d_2}{s_2}\frac{\partial u_2}{\partial \nu}\right\}\zeta\,\mathrm{d}\sigma\mathrm{d}t=0\end{aligned} \tag{3.21}$$

for all $\zeta\in C^\infty(\overline{Q_T})$ such that $\zeta(x,0)=\zeta(x,T)=0$. Considering successively in (3.21) test functions with compact support in Ω_+ and test functions with compact support in Ω_-, we deduce the parabolic equations for u_1 and u_2. Then, without loss of generality, we may assume that $\Gamma(t)=\partial\Omega_+(t)$ for $t\in[0,T]$. Taking in (3.21) test functions which vanish on $\partial\Omega\times[0,T]$ and do not vanish on Γ, we deduce the Stefan condition describing the interface motion. Since $\partial\Omega\cap\overline{\Omega_+(t)}=\emptyset$, it is clear that

$$\frac{\partial u_1}{\partial\nu}=0\quad\text{on }\partial\Omega\times(0,T).$$

Thus, taking in (3.21) test functions which do not vanish on $\partial\Omega\times(0,T)$, we can deduce that

$$\frac{\partial u_2}{\partial\nu}=0\quad\text{on }\partial\Omega\times(0,T).$$

Finally consider (3.15) with test functions which do not vanish at $t=0$. Then we can replace (3.19) with

$$\begin{aligned}&\iint_{Q_T}z\zeta_t\,\mathrm{d}x\mathrm{d}t\\&\quad=-\int_\Omega z(x,0)\zeta(x,0)\mathrm{d}x-\frac{1}{s_1}\int_0^T\int_{\Omega_+(t)}u_{1t}\zeta\,\mathrm{d}x\mathrm{d}t\\&\qquad+\frac{1}{s_2}\int_0^T\int_{\Omega_-(t)}u_{2t}\zeta\,\mathrm{d}x\mathrm{d}t-\int_0^T\int_{\Gamma(t)}\lambda V_n\zeta\,\mathrm{d}\sigma\mathrm{d}t.\end{aligned} \tag{3.22}$$

Substituting (3.22) and (3.20) into (3.15) and using the fact that both the partial differential equations for u_1 and u_2 as well as the Stefan condition for the interface motion and the

Neumann boundary conditions are satisfied, we deduce that

$$-\int_{\Omega}\{z(x,0)-z_0(x)\}\zeta(x,0)\mathrm{d}x=0$$

for all $\zeta(\cdot,0)\in L^2(\Omega)$. Hence

$$z(x,0)=z_0(x)\quad \text{a.e. in } \Omega.$$

Thus we obtain (3.17). □

REMARK 3.15. If the initial data satisfy :

$$u_{01}u_{02}=(1-w_0)u_{01}=w_0u_{02}=0\quad \text{in } \Omega,$$

then

$$u_1(x,0)=u_{01}(x),\quad \text{and}\quad u_2(x,0)=u_{02}(x)\quad \text{for all } x\in\Omega,$$

in Problems (P^*) and (Q^*).

REMARK 3.16. When $u_1\equiv 0$, the system becomes

$$\begin{cases} u_{2t}=d_2\Delta u_2+g(u_2)-\lambda s_2kwu_2, & x\in\Omega, t>0,\\ w_t=-kwu_2, & x\in\Omega, t>0,\end{cases}$$

which, in a special case where $g=0$ and where the initial functions are constant, converges to the one-phase Stefan problem (Q^*) with $u_1\equiv 0$. We refer to the papers by Hilhorst, van der Hout and Peletier [15–17] and by Eymard, Hilhorst, van der Hout and Peletier [13]. Now let us assume that

$$\operatorname{supp} u_{02}\cap \operatorname{supp} w_0=\emptyset.$$

We use the notations

$$\begin{aligned}&\Omega_0(t):=\{x\in\Omega|\varphi_\lambda(z(x,t))=0\}\\ &\Gamma(t):=\partial\Omega_0(t)\cap\Omega\end{aligned}$$

instead of $\Omega_+(t)$. Then (3.18) is replaced by

$$\iint_{Q_T} z\zeta_t\mathrm{d}x\mathrm{d}t=\int_0^T\left\{\int_{\Omega_0(t)}\lambda w\zeta_t\mathrm{d}x-\int_{\Omega_-(t)}\frac{u_2}{s_2}\zeta_t\mathrm{d}x\right\}\mathrm{d}t.$$

Since

$$\left[\int_{\Omega_0(t)} w\zeta\,\mathrm{d}x\right]_{t=0}^{t=T}=\int_0^T\int_{\Omega_0(t)}(w\zeta_t+w_t\zeta)\mathrm{d}x\mathrm{d}t+\int_0^T\int_{\Gamma(t)} wV_n\zeta\,\mathrm{d}\sigma\,\mathrm{d}t,$$

we have for all test functions which vanish at $t = 0$ and $t = T$,

$$\iint_{Q_T} z\zeta_t \mathrm{d}x\mathrm{d}t = \frac{1}{s_2}\int_0^T \int_{\Omega_-(t)} u_{2t}\zeta \mathrm{d}x\mathrm{d}t - \lambda \int_0^T \int_{\Omega_0(t)} w_t \zeta \mathrm{d}x\mathrm{d}t \\ - \lambda \int_0^T \int_{\Gamma(t)} w V_n \zeta \mathrm{d}\sigma \mathrm{d}t \tag{3.23}$$

instead of (3.19). Substituting (3.23) and (3.20) with $u_1 \equiv 0$ into (3.15) implies

$$\int_0^T \int_{\Gamma(t)} \left\{ -\lambda w V_n - \frac{d_2}{s_2}\frac{\partial u_2}{\partial n} \right\} \zeta \mathrm{d}\sigma \mathrm{d}t - \lambda \int_0^T \int_{\Omega_0(t)} w_t \zeta \mathrm{d}x\mathrm{d}t \\ + \frac{1}{s_2}\int_0^T \int_{\Omega_-(t)} \{u_{2t} - d_2 \Delta u_2 - g(u_2)\}\zeta \mathrm{d}x\mathrm{d}t = 0.$$

We remark that the support of u_2 is expanding, and since

$$w_t = 0 \quad \text{in } \Omega_0(t),$$

it follows that

$$w(\cdot, t) = w_0 \quad \text{in } \Omega_0(t).$$

Thus we have

$$-\lambda w_0 V_n - \frac{d_2}{s_2}\frac{\partial u_2}{\partial n} = 0 \quad \text{on } \Gamma(t),$$

where w_0 in the first term is regarded as the boundary values of $\partial\Omega_0(t)$. Imposing the value

$$w_0 = 1 \quad \text{in } \Omega_0,$$

we obtain a one-phase Stefan problem, with the interface equation

$$-\lambda V_n - \frac{d_2}{s_2}\frac{\partial u_2}{\partial n} = 0 \quad \text{on } \Gamma(t).$$

REMARK 3.17. In order to illustrate extremely exciting patterns which may arise from problems slightly more complex than the four-component system (3.2), we show numerical tests (cf. Figure 3) for the solution of a six-component system of the form

$$\begin{cases} u_t^i = d_i \Delta u^i + f_i(u^i) - k \sum_{j\neq i} a_{ij} u^j u^i - k \sum_{j\neq i} w^j u^i \\ \quad (x \in \Omega,\ t > 0,\ i = 1, \ldots, 3), \\ w_t^i = k \sum_{j\neq i} w^j u^i - k \sum_{j\neq i} u^j w^i \quad (x \in \Omega, t > 0, i = 1, \ldots, 3). \end{cases} \tag{3.24}$$

This solution is quite remarkable, since the corresponding interfaces exhibit triple junctions.

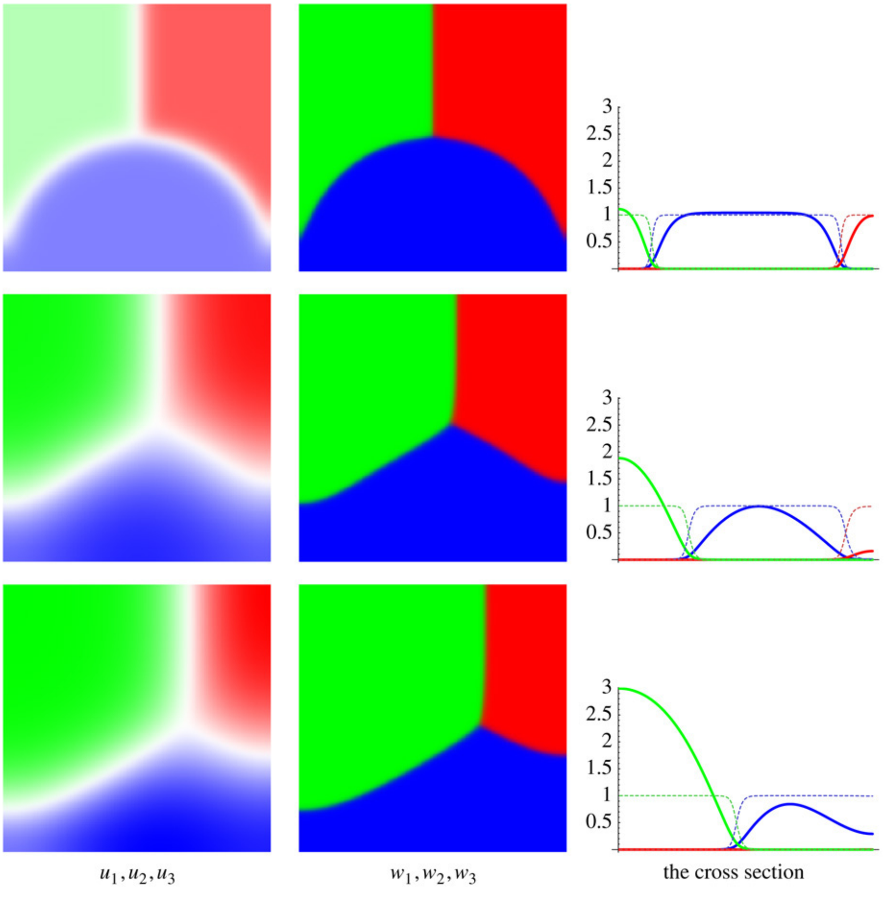

Fig. 3. The snapshots for the solutions of a 6-components system (3.24) with $k = 100, d_i = 10.0$, $a_{ij} = 1.0$ $(i \neq j)$, $f_1(u) = 2u(3-u)$, $f_2(u) = 3u(2-u)$, $f_3(u) = 5u(4-u)$ at $t = 0, 0.2, 0.4$.

3.3. *Error estimates*

The purpose of this section is to prove the error estimates of Theorem 1.5. We combine proofs from two papers of Murakawa [41,42].

To begin with, we recall the definitions of z^k, θ and θ^k given in (1.2) and (1.3):

$$z^k = \frac{u_1^k}{s_1} - \frac{u_2^k}{s_2} + \lambda w^k,$$

$$\theta = d(\varphi_\lambda(z)) = d_1(z-\lambda)^+ - d_2 z^-,$$

$$\theta^k = \frac{d_1}{s_1} u_1^k - \frac{d_2}{s_2} u_2^k,$$

where (u_1^k, u_2^k, w^k) is a solution of (Q^k) and z is a weak solution of (Q). We also define the functions:

$$\varphi_1(z) := s_1\varphi_\lambda(z)^+ = s_1(z-\lambda)^+, \qquad \varphi_2(z) := s_2\varphi_\lambda(z)^- = s_2 z^-.$$

Using these functions, we set

$$u_1 := \varphi_1(z) = s_1(z-\lambda)^+, \qquad u_2 := \varphi_2(z) = s_2 z^-,$$

so that in particular

$$\theta = \frac{d_1}{s_1}u_1 - \frac{d_2}{s_2}u_2,$$

and prove the following result.

LEMMA 3.18. *The following bounds hold:*

$$\|\varphi_1(z^k) - u_1^k\|_{L^2(Q_T)} + \|\varphi_2(z^k) - u_2^k\|_{L^2(Q_T)} \le \frac{C}{\sqrt{k}},$$
$$\|d(\varphi_\lambda(z^k)) - \theta^k\|_{L^2(Q_T)} \le \frac{C}{\sqrt{k}}.$$

PROOF. We first estimate $|\,\varphi_1(z^k) - u_1^k\,|^2$. If $z^k = u_1^k/s_1 - u_2^k/s_2 + \lambda w^k < \lambda$, then

$$\left|\varphi_1\left(\frac{u_1^k}{s_1} - \frac{u_2^k}{s_2} + \lambda w^k\right) - u_1^k\right|^2 = |u_1^k|^2 \le \frac{s_1}{s_2}u_1^k u_2^k + s_1\lambda u_1^k(1-w^k).$$

If $u_1^k/s_1 - u_2^k/s_2 + \lambda w^k \ge \lambda$, then here as well we find that

$$\left|\varphi_1\left(\frac{u_1^k}{s_1} - \frac{u_2^k}{s_2} + \lambda w^k\right) - u_1^k\right|^2 \le \frac{s_1}{s_2}u_1^k u_2^k + s_1\lambda u_1^k(1-w^k),$$

which together with Lemma 3.3 implies that

$$\left\|\varphi_1\left(\frac{u_1^k}{s_1} - \frac{u_2^k}{s_2} + \lambda w^k\right) - u_1^k\right\|_{L^2(Q_T)} \le \frac{C}{\sqrt{k}}.$$

Next we consider the term $|\varphi_2(z^k) - u_2^k|^2$. If $z^k = u_1^k/s_1 - u_2^k/s_2 + \lambda w^k > 0$, then

$$\left|\varphi_2\left(\frac{u_1^k}{s_1} - \frac{u_2^k}{s_2} + \lambda w^k\right) - u_2^k\right|^2 = |u_2^k|^2 \le s_2 u_2^k\left(\frac{u_1^k}{s_1} + \lambda w^k\right).$$

If $u_1^k/s_1 - u_2^k/s_2 + \lambda w^k \le 0$, then

$$\left|\varphi_2\left(\frac{u_1^k}{s_1} - \frac{u_2^k}{s_2} + \lambda w^k\right) - u_2^k\right|^2 \le s_2 u_2^k \left(\frac{u_1^k}{s_1} + \lambda w^k\right),$$

which together with Lemma 3.3 implies that

$$\left\|\varphi_2\left(\frac{u_1^k}{s_1} - \frac{u_2^k}{s_2} + \lambda w^k\right) - u_2^k\right\|_{L^2(Q_T)} \le \frac{C}{\sqrt{k}}.$$

This completes the proof of the first inequality in Lemma 3.18. Next we show the second one. We have that

$$d(\varphi_\lambda(z^k)) - \theta^k = d_1\left\{(z^k - \lambda)^+ - \frac{u_1^k}{s_1}\right\} - d_2\left\{(z^k)^- - \frac{u_2^k}{s_2}\right\}.$$

Therefore

$$\begin{aligned}&\|d(\varphi_\lambda(z^k)) - \theta^k\|_{L^2(Q_T)}\\ &\quad\le \frac{d_1}{s_1}\|\varphi_1(z^k) - u_1^k\|_{L^2(Q_T)} + \frac{d_2}{s_2}\|\varphi_2(z^k) - u_2^k\|_{L^2(Q_T)} \le \frac{C}{\sqrt{k}},\end{aligned}$$

which completes the proof. □

PROOF OF THEOREM 1.5. As it is done by Murakawa, we define

$$\mathcal{K} = \{\zeta \in H^1(Q_T) \mid \zeta(\cdot, T) = 0\}.$$

It then follows as in the proof of Lemma 3.8 that

$$\iint_{Q_T} z^k\zeta_t + \int_\Omega z_0^k(x)\zeta(x,0)\mathrm{d}x = \iint_{Q_T}\left\{\nabla\theta^k\nabla\zeta - \left(\frac{f(u_1^k)}{s_1} - \frac{g(u_2^k)}{s_2}\right)\zeta\right\}$$

for all $\zeta \in \mathcal{K}$. Moreover it follows from (3.15) that

$$\iint_{Q_T} z\zeta_t + \int_\Omega z_0(x)\zeta(x,0)\mathrm{d}x = \iint_{Q_T}\left\{\nabla\theta\nabla\zeta - \left(\frac{f(u_1)}{s_1} - \frac{g(u_2)}{s_2}\right)\zeta\right\}$$

for all $\zeta \in \mathcal{K}$. Thus

$$\begin{aligned}&\iint_{Q_T} e_z\zeta_t + \int_\Omega e_{z_0}(x)\zeta(x,0)\mathrm{d}x\\ &\quad= \iint_{Q_T}\left\{\nabla e_\theta\nabla\zeta - \left(\frac{f(u_1)}{s_1} - \frac{f(u_1^k)}{s_1}\right)\zeta + \left(\frac{g(u_2)}{s_2} - \frac{g(u_2^k)}{s_2}\right)\zeta\right\},\end{aligned} \tag{3.25}$$

where

$$e_z = z - z^k, \qquad e_{z0} = z_0 - z_0^k, \quad \text{and} \quad e_\theta = \theta - \theta^k.$$

We set

$$\zeta(x,t) = \begin{cases} \int_t^{t_0} e_\theta(x,s)\mathrm{d}s & \text{if } 0 \le t \le t_0, \\ 0 & \text{if } t_0 \le t \le T. \end{cases}$$

The first term on the left-hand-side of (3.25) is equal to

$$\iint_{Q_T} e_z \zeta_t = -\int_0^{t_0} \int_\Omega e_z e_\theta,$$

and

$$\begin{aligned} \int_0^{t_0} \int_\Omega e_z e_\theta &= \int_0^{t_0} \int_\Omega e_z \left\{ \theta - d(\varphi_\lambda(z^k)) \right\} + \int_0^{t_0} \int_\Omega e_z \left\{ d(\varphi_\lambda(z^k)) - \theta^k \right\} \\ &= \frac{d_1}{s_1} \int_0^{t_0} \int_\Omega e_z (u_1 - \varphi_1(z^k)) - \frac{d_2}{s_2} \int_0^{t_0} \int_\Omega e_z (u_2 - \varphi_2(z^k)) \\ &\quad + \int_0^{t_0} \int_\Omega e_z \left\{ d(\varphi_\lambda(z^k)) - \theta^k \right\}. \end{aligned}$$

We remark that

$$|\varphi_1(z) - \varphi_1(z^k)| = s_1 |(z-\lambda)^+ - (z^k - \lambda)^+| \le s_1 |z - z^k|,$$

and since φ_1 is monotone, it follows that

$$\begin{aligned} \frac{d_1}{s_1} \int_\Omega (z - z^k)(\varphi_1(z) - \varphi_1(z^k)) &= \frac{d_1}{s_1} \int_\Omega |z - z^k| |\varphi_1(z) - \varphi_1(z^k)| \\ &\ge \frac{d_1}{{s_1}^2} \int_\Omega |\varphi_1(z) - \varphi_1(z^k)|^2. \end{aligned}$$

Next we use the fact that

$$\begin{aligned} \|u_1 - u_1^k\|^2_{L^2(\Omega)} &= \|\varphi_1(z) - \varphi_1(z^k) + \varphi_1(z^k) - u_1^k\|^2_{L^2(\Omega)} \\ &\le 2\|\varphi_1(z) - \varphi_1(z^k)\|^2_{L^2(\Omega)} + 2\|\varphi_1(z^k) - u_1^k\|^2_{L^2(\Omega)}, \end{aligned}$$

to deduce that

$$\begin{aligned} &\frac{d_1}{s_1} \iint_{Q_{t_0}} (z - z^k)(\varphi_1(z) - \varphi_1(z^k)) \\ &\quad \ge \frac{d_1}{s_1^2} \left(\frac{1}{2} \|e_{u_1}\|^2_{L^2(Q_{t_0})} - \|\varphi_1(z^k) - u_1^k\|^2_{L^2(Q_{t_0})} \right), \end{aligned}$$

where $e_{u_1} := u_1 - u_1^k$. In the same way we obtain

$$-\frac{d_2}{s_2}\iint_{Q_{t_0}} (z - z^k)(\varphi_2(z) - \varphi_2(z^k)) \geq \frac{d_2}{s_2^2}\left(\frac{1}{2}\|e_{u_2}\|^2_{L^2(Q_{t_0})} - \|\varphi_2(z^k) - u_2^k\|^2_{L^2(Q_{t_0})}\right),$$

where $e_{u_2} := u_2 - u_2^k$. Moreover we deduce from the Cauchy–Schwarz inequality, the uniform boundedness of z^k and Lemma 3.18 that

$$\left|\iint_{Q_{t_0}} e_z \left\{d(\varphi_\lambda(z^k)) - \theta^k\right\}\right| \leq \|e_z\|_{L^2(Q_{t_0})}\|d(\varphi_\lambda(z^k)) - \theta^k\|_{L^2(Q_{t_0})} \leq \frac{C}{\sqrt{k}}.$$

Thus we obtain

$$\iint_{Q_T} e_z\zeta_t \leq -\frac{d_1}{s_1^2}\left(\frac{1}{2}\|e_{u_1}\|^2_{L^2(Q_{t_0})} - \|\varphi_1(z^k) - u_1^k\|^2_{L^2(Q_{t_0})}\right) - \frac{d_2}{s_2^2}\left(\frac{1}{2}\|e_{u_2}\|^2_{L^2(Q_{t_0})} - \|\varphi_2(z^k) - u_2^k\|^2_{L^2(Q_{t_0})}\right) + \frac{C}{\sqrt{k}}. \tag{3.26}$$

Recall that

$$e_\theta = \theta - \theta^k = \frac{d_1}{s_1}(u_1 - u_1^k) - \frac{d_2}{s_2}(u_2 - u_2^k) = \frac{d_1}{s_1}e_{u_1} - \frac{d_2}{s_2}e_{u_2}. \tag{3.27}$$

It follows from (3.27) that

$$\int_\Omega e_{z_0}\zeta(0) = \iint_{Q_{t_0}} e_{z_0}e_\theta = \iint_{Q_{t_0}} e_{z_0}\left(\frac{d_1}{s_1}e_{u_1} - \frac{d_2}{s_2}e_{u_2}\right) \leq (d_1 + d_2)T\|e_{z_0}\|^2_{L^2(\Omega)} + \frac{d_1}{4s_1^2}\|e_{u_1}\|^2_{L^2(Q_T)} + \frac{d_2}{4s_2^2}\|e_{u_2}\|^2_{L^2(Q_T)}. \tag{3.28}$$

Next we remark that the first term on the right-hand-side of (3.25) is equal to

$$\iint_{Q_T} \nabla e_\theta \nabla\zeta = \iint_{Q_{t_0}} \nabla e_\theta \int_t^{t_0} \nabla e_\theta = -\frac{1}{2}\iint_{Q_{t_0}} \frac{\mathrm{d}}{\mathrm{d}t}\left(\nabla\int_t^{t_0} e_\theta\right)^2 = -\frac{1}{2}\int_\Omega \left[\nabla\int_t^{t_0} e_\theta\right]_{t=0}^{t=t_0}.$$

Therefore

$$\iint_{Q_T} \nabla e_\theta \nabla \zeta = \frac{1}{2} \left\| \nabla \int_0^{t_0} e_\theta \right\|^2_{L^2(\Omega)}. \tag{3.29}$$

We consider the second and the third terms on the right-hand-side of (3.25). By (3.27),

$$\begin{aligned} &\left| - \iint_{Q_{t_0}} \left\{ \left(\frac{f(u_1)}{s_1} - \frac{f(u_1^k)}{s_1} \right) \int_t^{t_0} e_\theta \right\} \right| \\ &\quad \le C \iint_{Q_{t_0}} |e_{u_1}| \int_t^{t_0} \left(\frac{d_1}{s_1} |e_{u_1}| + \frac{d_2}{s_2} |e_{u_2}| \right). \end{aligned}$$

We have that

$$\begin{aligned} &\iint_{Q_{t_0}} \mathrm{d}t |e_{u_1}(x,t)| \int_t^{t_0} \mathrm{d}s |e_{u_1}(x,s)| \\ &\quad = \iint_{Q_{t_0}} \mathrm{d}s |e_{u_1}(x,s)| \int_0^s |e_{u_1}(x,t)| \mathrm{d}t \\ &\quad \le \int_0^{t_0} \mathrm{d}s \left(\int_\Omega |e_{u_1}(x,s)|^2 \mathrm{d}x \right)^{\frac{1}{2}} \left(\int_\Omega \left(\int_0^s |e_{u_1}(x,t) \mathrm{d}t \right)^2 \mathrm{d}x \right)^{\frac{1}{2}} \\ &\quad \le \int_0^{t_0} \mathrm{d}s \left(\varepsilon \int_\Omega |e_{u_1}(x,s)|^2 \mathrm{d}x + C_\varepsilon \int_\Omega \left(\int_0^s |e_{u_1}(x,t)| \mathrm{d}t \right)^2 \mathrm{d}x \right) \\ &\quad \le \varepsilon \|e_{u_1}\|^2_{L^2(Q_{t_0})} + C_\varepsilon T \int_0^{t_0} \|e_{u_1}\|^2_{L^2(0,s;L^2(\Omega))} \mathrm{d}s. \end{aligned}$$

Similarly

$$\begin{aligned} &\iint_{Q_{t_0}} \mathrm{d}t |e_{u_1}(x,t)| \int_t^{t_0} \mathrm{d}s |e_{u_2}(x,s)| \\ &\quad = \iint_{Q_{t_0}} \mathrm{d}s |e_{u_2}(x,s)| \int_0^s |e_{u_1}(x,t)| \mathrm{d}t \\ &\quad \le \varepsilon \|e_{u_2}\|^2_{L^2(Q_{t_0})} + C_\varepsilon T \int_0^{t_0} \|e_{u_1}\|^2_{L^2(0,s;L^2(\Omega))} \mathrm{d}s. \end{aligned}$$

Combining these inequalities, we obtain

$$\begin{aligned} &\left| - \iint_{Q_{t_0}} \left\{ \left(\frac{f(u_1)}{s_1} - \frac{f(u_1^k)}{s_1} \right) \int_t^{t_0} e_\theta \right\} \right| \\ &\quad \le \varepsilon \left(\|e_{u_1}\|^2_{L^2(Q_{t_0})} + \|e_{u_2}\|^2_{L^2(Q_{t_0})} \right) \\ &\qquad + C_\varepsilon T \int_0^{t_0} \left(\|e_{u_1}\|^2_{L^2(0,s;L^2(\Omega))} + \|e_{u_2}\|^2_{L^2(0,s;L^2(\Omega))} \right) \mathrm{d}s. \end{aligned} \tag{3.30}$$

Moreover, one can perform similar computations for the reaction function g. Substituting (3.26), (3.28), (3.29) and (3.30) into (3.25) yields

$$\begin{aligned}
&\frac{d_1}{s_1^2}\left(\frac{1}{2}\|e_{u_1}\|^2_{L^2(Q_{t_0})} - \|\varphi_1(z^k) - u_1^k\|^2_{L^2(Q_{t_0})}\right) \\
&\quad + \frac{d_2}{s_2^2}\left(\frac{1}{2}\|e_{u_2}\|^2_{L^2(Q_{t_0})} - \|\varphi_2(z^k) - u_2^k\|^2_{L^2(Q_{t_0})}\right) \\
&\quad - \frac{C}{\sqrt{k}} + \frac{1}{2}\left\|\nabla\int_0^{t_0} e_\theta\right\|^2_{L^2(\Omega)} \\
&\quad \le (d_1+d_2)T\|e_{z_0}\|^2_{L^2(\Omega)} + \frac{d_1}{4s_1^2}\|e_{u_1}\|^2_{L^2(Q_T)} + \frac{d_2}{4s_2^2}\|e_{u_2}\|^2_{L^2(Q_T)} \\
&\qquad + \varepsilon(\|e_{u_1}\|^2_{L^2(Q_{t_0})} + \|e_{u_2}\|^2_{L^2(Q_{t_0})}) + C_\varepsilon T\int_0^{t_0}(\|e_{u_1}\|^2_{L^2(0,s;L^2(\Omega))} \\
&\qquad + \|e_{u_2}\|^2_{L^2(0,s;L^2(\Omega))})\mathrm{d}s,
\end{aligned}$$

which implies

$$\begin{aligned}
&\|e_{u_1}\|^2_{L^2(Q_{t_0})} + \|e_{u_2}\|^2_{L^2(Q_{t_0})} + C_2\left\|\nabla\int_t^{t_0} e_\theta\right\|^2_{L^2(\Omega)} \\
&\quad \le C_1(\|\varphi_1(z^k) - u_1^k\|^2_{L^2(Q_{t_0})} + \|\varphi_2(z^k) - u_2^k\|^2_{L^2(Q_{t_0})}) \\
&\qquad + \frac{C_3}{\sqrt{k}} + C_4T\|e_{z_0}\|^2_{L^2(\Omega)} + C_5T\int_0^{t_0}(\|e_{u_1}\|^2_{L^2(0,s;L^2(\Omega))} \\
&\qquad + \|e_{u_2}\|^2_{L^2(0,s;L^2(\Omega))})\mathrm{d}s.
\end{aligned}$$

Lemma 3.18 implies that

$$\begin{aligned}
&\|e_{u_1}\|^2_{L^2(Q_{t_0})} + \|e_{u_2}\|^2_{L^2(Q_{t_0})} + C_2\left\|\nabla\int_0^{t_0} e_\theta\right\|^2_{L^2(\Omega)} \\
&\quad \le \frac{C_6}{\sqrt{k}} + C_4T\|e_{z_0}\|^2_{L^2(\Omega)} \\
&\qquad + C_5T\int_0^{t_0}(\|e_{u_1}\|^2_{L^2(0,s;L^2(\Omega))} + \|e_{u_2}\|^2_{L^2(0,s;L^2(\Omega))})\mathrm{d}s
\end{aligned}$$

for large k. Applying Gronwall's inequality and using the fact that e_θ is uniformly bounded, we obtain the error estimate

$$\begin{aligned}
&\|e_{u_1}\|^2_{L^2(Q_{t_0})} + \|e_{u_2}\|^2_{L^2(Q_{t_0})} + C_2\left\|\int_0^{t_0} e_\theta\right\|^2_{H^1(\Omega)} \\
&\quad \le \frac{C_6}{\sqrt{k}} + C_4T\|e_{z_0}\|^2_{L^2(\Omega)},
\end{aligned} \tag{3.31}$$

for all $t_0 \in (0, T]$. In order to complete the proof of Theorem 1.5, we still need to estimate the term e_z. Let $\phi \in H^1(\Omega)$ and let $t_0 \in (0, T)$ be arbitrary. For $\delta \in (0, \min(t_0, T - t_0))$, we define the cut-off function $\chi_\delta = \chi_\delta(t)$ as

$$\chi_\delta(t) = \begin{cases} 1 & \text{if } t \in [0, t_0 - \delta] \\ \dfrac{t_0 + \delta - t}{2\delta} & \text{if } t \in [t_0 - \delta, t_0 + \delta] \\ 0 & \text{if } t \in [t_0 + \delta, T]. \end{cases}$$

The function χ_δ converges to the characteristic function of the interval $(0, t_0)$ as $\delta \to 0$. Taking $\zeta(x,t) = \phi(x)\chi_\delta(t)$ in (3.25), we obtain

$$\begin{aligned} &-\frac{1}{2\delta}\int_{t_0-\delta}^{t_0+\delta}\int_\Omega e_z\phi + \int_\Omega e_{z_0}\phi \\ &\quad = \iint_{Q_T} \chi_\delta \nabla e_\theta \nabla\phi - \iint_{Q_T} \chi_\delta \left\{ \left(\frac{f(u_1)}{s_1} - \frac{f(u_1^k)}{s_1} \right) \right. \\ &\qquad \left. - \left(\frac{g(u_2)}{s_2} - \frac{g(u_2^k)}{s_2} \right) \right\} \phi. \end{aligned}$$

Letting $\delta \to 0$, we deduce that

$$\begin{aligned} \left| \int_\Omega e_z(t_0)\phi \right| \le \left| \int_\Omega e_{z_0}\phi \right| + \left| \iint_{Q_{t_0}} \nabla e_\theta \nabla \phi \right| \\ + \iint_{Q_{t_0}} \left\{ \left| \frac{f(u_1)}{s_1} - \frac{f(u_1^k)}{s_1} \right| + \left| \frac{g(u_2)}{s_2} - \frac{g(u_2^k)}{s_2} \right| \right\} |\phi|, \end{aligned}$$

for all $\phi \in H^1(\Omega)$. By (3.31),

$$\sup_{\|\phi\|_{H^1(\Omega)}=1} \langle e_z(t_0), \phi \rangle_{\langle H^1(\Omega), (H^1(\Omega))^* \rangle} \le C\left(\|e_{z_0}\|_{L^2(\Omega)} + k^{-1/4} \right),$$

so that

$$\|(z - z^k)(t_0)\|_{(H^1(\Omega))^*} \le C(\|e_{z_0}\|_{L^2(\Omega)} + k^{-1/4})$$

for all $t_0 \in (0, T)$ and thus

$$\|z - z^k\|_{L^\infty(0,T;(H^1(\Omega))^*)} \le C(\|e_{z_0}\|_{L^2(\Omega)} + k^{-1/4}),$$

which completes the proof of Theorem 1.5. □

4. The fast reaction limit to nonlinear diffusion systems

First we write the singular limit problem in a slightly more general setting, namely

$$\boldsymbol{u}_t = D\Delta \boldsymbol{u} + \boldsymbol{F}(\boldsymbol{u}) + k\boldsymbol{G}(\boldsymbol{u}), \tag{4.1}$$

where $\boldsymbol{u} \in \mathbb{R}^m$ and $\boldsymbol{F}, \boldsymbol{G}$ are smooth functions from $\mathbb{R}^m$ to $\mathbb{R}^m$, D is a diagonal matrix with positive components. Dividing the above equation by k, we have

$$\frac{\boldsymbol{u}_t}{k} = \frac{D}{k}\Delta \boldsymbol{u} + \frac{1}{k}\boldsymbol{F}(\boldsymbol{u}) + \boldsymbol{G}(\boldsymbol{u}).$$

Letting $k \to \infty$, we can expect that

$$\boldsymbol{0} = \boldsymbol{G}(\boldsymbol{u}).$$

The sets of the equilibria of the examples in Section 2 are

$$\{(u_1, u_2) \mid u_1 \geq 0, u_2 \geq 0\} \cup \{(u_1, u_2) \mid u_1 = 0, u_2 \geq 0\},$$
$$\{(u_1, 0, 1) \mid u_1 \geq 0\} \cup \{(0, 0, w) \mid 0 \leq w \leq 1\} \cup \{(0, u_2, 0) \mid u_2 \geq 0\},$$

respectively. Since the sets of equilibria are not smooth at $(u_1, u_2) = (0, 0)$ or $(u_1, u_2, w) = (0, 0, 0), (0, 0, 1)$, interfaces may appear. On the other hand, the set corresponding to the system (1.5) in Section 2.2 is given by

$$\{(v_1, v_2, v_3) \mid h_{1\to 2}(v_3)v_1 = h_{2\to 1}(v_3)v_2\}.$$

In the case where the null set of $\boldsymbol{G}$ is smooth, we can expect other limit problems not only free boundary value problems. For example, Bothe and Hilhorst [4] discussed the following system:

$$u_{1t} = d_1 \Delta u_1 + k(h_{2\to 1}(u_2) - h_{1\to 2}(u_1)), \tag{4.2}$$
$$u_{2t} = d_2 \Delta u_2 + k(h_{1\to 2}(u_1) - h_{2\to 1}(u_2)). \tag{4.3}$$

Here $h_{1\to 2}(s)$ and $h_{2\to 1}(s)$ are increasing in s and $h_*(s) = 0$ if and only if $s = 0$, where $* = 1 \to 2$, or $* = 2 \to 1$. Adding (4.2) and (4.3), we have

$$w_t = \Delta(d_1 u_1 + d_2 u_2), \tag{4.4}$$

where $w = u_1 + u_2$. As $k \to \infty$, we can expect that

$$h_{2\to 1}(u_2) = h_{1\to 2}(u_1), \quad \text{i.e. } u_2 = h_{2\to 1}^{-1} \circ h_{1\to 2}(u_1).$$

Thus we find at the limit,

$$w = (id + h_{2\to 1}^{-1} \circ h_{1\to 2})(u_1), \quad \text{i.e. } u_1 = (id + h_{2\to 1}^{-1} \circ h_{1\to 2})^{-1}(w),$$

which we substitute into (4.4) to obtain

$$w_t = \Delta\left((d_1 id + d_2 h_{2\to 1}^{-1} \circ h_{1\to 2})(id + h_{2\to 1}^{-1} \circ h_{1\to 2})^{-1}(w)\right).$$

We refer to [4] for a rigorous proof.

Before proving Theorem 1.6, we consider the more general problem

$$\begin{cases} \widetilde{v_1}_t = d_1\Delta\widetilde{v_1} + \tilde{\alpha}\Delta\widetilde{v_3} + f_1(\widetilde{v_1},\widetilde{v_2}) \\ \qquad + \eta_1(\widetilde{v_1},\widetilde{v_2},\widetilde{v_3}), & t>0, x\in\Omega, \\ \widetilde{v_2}_t = d_2\Delta\widetilde{v_2} + f_2(\widetilde{v_1},\widetilde{v_2}) + \eta_2(\widetilde{v_1},\widetilde{v_2},\widetilde{v_3}), & t>0, x\in\Omega, \\ \widetilde{v_3}_t = (d_1+\tilde{\alpha})\Delta\widetilde{v_3} + \eta_3(\widetilde{v_1},\widetilde{v_2},\widetilde{v_3}) \\ \qquad + k\kappa(\widetilde{v_1},\widetilde{v_2},\widetilde{v_3}), & t>0, x\in\Omega \end{cases} \tag{4.5}$$

with the boundary and initial conditions

$$\frac{\partial\widetilde{v_1}}{\partial\nu} = \frac{\partial\widetilde{v_2}}{\partial\nu} = \frac{\partial\widetilde{v_3}}{\partial\nu} = 0, \quad t>0, x\in\partial\Omega, \tag{4.6}$$

$$\widetilde{v_1}(0,x) = \widetilde{v}_{01}(x), \qquad \widetilde{v_2}(0,x) = \widetilde{v}_{02}(x),$$

$$\widetilde{v_3}(0,x) = \phi(\widetilde{v}_{02}(x))\widetilde{v}_{01}(x), \quad x\in\Omega \tag{4.7}$$

and show that it approximates the cross-diffusion system

$$\begin{cases} \widetilde{u_1}_t = \Delta[(d_1+\tilde{\alpha}\phi(\widetilde{u_2}))\widetilde{u_1}] + f(\widetilde{u_1},\widetilde{u_2}), & t>0, x\in\Omega, \\ \widetilde{u_2}_t = d_2\Delta\widetilde{u_2} + g(\widetilde{u_1},\widetilde{u_2}), & t>0, x\in\Omega \end{cases} \tag{4.8}$$

with the boundary and initial conditions

$$\frac{\partial\widetilde{u_1}}{\partial\nu} = \frac{\partial\widetilde{u_2}}{\partial\nu} = 0, \quad t>0, x\in\partial\Omega, \tag{4.9}$$

$$\widetilde{u_1}(0,x) = \widetilde{v}_{01}(x), \qquad \widetilde{u_2}(0,x) = \widetilde{v}_{02}(x), \quad x\in\Omega, \tag{4.10}$$

where we assume

$$\widetilde{v}_{01}(x)\ge 0, \quad \widetilde{v}_{02}(x)\ge 0 \quad \text{on } \Omega.$$

System (4.5) corresponds to (2.17). We note that $\widetilde{v_3}$ in (4.5) corresponds to v_2 in (2.17). For simplicity we assume that the functions ϕ, f_1, f_2, η_1, η_2, η_3 and κ allow non-negative numbers as their independent variables.

THEOREM 4.1 *([23, Theorem 2]). Let d_1, d_2, $\tilde{\alpha}$, $\tilde{M_0}$ be positive numbers and let ϕ, f_1, f_2, η_1, η_2, η_3 and κ be smooth functions satisfying*

$$\phi(s_2)\ge 0, \tag{4.11}$$

$$\eta_1(s_1,s_2,\phi(s_2)s_1) \equiv \eta_2(s_1,s_2,\phi(s_2)s_1) \equiv \kappa(s_1,s_2,\phi(s_2)s_1) \equiv 0, \tag{4.12}$$

$$\kappa_{\widetilde{v_1}}(s_1,s_2,s_3)\ge 0, \tag{4.13}$$

$$\kappa_{\widetilde{v_3}}(s_1, s_2, s_3) < 0, \tag{4.14}$$

$$\kappa_{\widetilde{v_1}\widetilde{v_3}}(s_1, s_2, s_3) \equiv 0 \tag{4.15}$$

for $(s_1, s_2, s_3) \in [0, \widetilde{M_0}]^3$. *Suppose that the solution* $(\widetilde{u_1}, \widetilde{u_2}) = (\widetilde{u_1}(x,t), \widetilde{u_2}(x,t))$ *of* (4.8), (4.9), (4.10) *is sufficiently smooth at least on* $\overline{\Omega} \times [0, T]$ *and that*

$$(\widetilde{u_1}(x,t), \widetilde{u_2}(x,t)) \in [0, \widetilde{M_0}]^2 \tag{4.16}$$

on $\overline{\Omega} \times [0, T]$, *where* T *is a positive number. Also suppose that the solution* $(\widetilde{v_1}, \widetilde{v_2}, \widetilde{v_3}) = (\widetilde{v_1}^k(x,t), \widetilde{v_2}^k(x,t), \widetilde{v_3}^k(x,t))$ *of* (4.5), (4.6), (4.7) *is sufficiently smooth at least on* $\overline{\Omega} \times [0, T]$ *and that*

$$(\widetilde{v_1}^k(x,t), \widetilde{v_2}^k(x,t), \widetilde{v_3}^k(x,t)) \in [0, \widetilde{M_0}]^3 \tag{4.17}$$

on $\overline{\Omega} \times [0, T]$ *for* $k \geq k_0$, *where* k_0 *is a positive number. Then the difference between* $(\widetilde{v_1}^k, \widetilde{v_2}^k, \widetilde{v_3}^k)$ *and* $(\widetilde{u_1}, \widetilde{u_2})$ *can be estimated by*

$$\begin{cases} \sup\limits_{t\in[0,T]} \|\widetilde{v_1}^k(\cdot,t) - \widetilde{u_1}(\cdot,t)\|_{L^2(\Omega)} \leq \dfrac{C}{k}, \\ \sup\limits_{t\in[0,T]} \|\widetilde{v_2}^k(\cdot,t) - \widetilde{u_2}(\cdot,t)\|_{L^2(\Omega)} \leq \dfrac{C}{k}, \\ \sup\limits_{t\in[0,T]} \|\widetilde{v_3}^k(\cdot,t) - \phi(\widetilde{u_2}(\cdot,t))\widetilde{u_1}(\cdot,t)\|_{L^2(\Omega)} \leq \dfrac{C}{k} \end{cases} \tag{4.18}$$

for $k \geq k_0$. *Here* $C = C(\widetilde{u_1}, \widetilde{u_2}, k_0, \widetilde{M_0}, T)$ *is a positive constant independent of* k.

We give the proof of this theorem in Appendix A.3.

PROOF OF THEOREM 1.6. It follows from (1.8) and (1.9) that

$$\begin{aligned} &(r_1 - a_1(v_1+v_2) - b_1 v_3)v_1 + k[h_{2\to1}(v_3)v_2 - h_{1\to2}(v_3)v_1] \geq 0 \\ &\qquad\qquad \text{if } v_1 = 0, v_2 \geq 0, 0 \leq v_3 \leq M_2; \\ &(r_1 - a_1(v_1+v_2) - b_1 v_3)v_2 + k[h_{1\to2}(v_3)v_1 - h_{2\to1}(v_3)v_2] \geq 0 \\ &\qquad\qquad \text{if } v_1 \geq 0, v_2 = 0, 0 \leq v_3 \leq M_2; \\ &(r_2 - b_2(v_1+v_2) - a_2 v_3)v_3 \geq 0 \quad \text{if } v_1 \geq 0, v_2 \geq 0, v_3 = 0; \\ &(r_2 - b_2(v_1+v_2) - a_2 v_3)v_3 \leq 0 \quad \text{if } v_1 \geq 0, v_2 \geq 0, v_3 = M_2. \end{aligned}$$

Hence the region $[0,\infty) \times [0,\infty) \times [0, M_2]$ for (v_1, v_2, v_3) is positively invariant in the reaction-diffusion system (1.5). In other words, (1.6) and (1.11) imply

$$v_1^k(x,t) \geq 0, \qquad v_2^k(x,t) \geq 0, \qquad 0 \leq v_3^k(x,t) \leq M_2.$$

Rewrite (u_1, u_2) and $(v_1 + v_2,\ v_3,\ v_2)$ as $(\widetilde{u_1}, \widetilde{u_2})$ and $(\widetilde{v_1}, \widetilde{v_2}, \widetilde{v_3})$, respectively. Set $\tilde{\alpha} := \alpha M_2$ and

$$
\begin{aligned}
&\phi(s_2) := \frac{s_2}{M_2},\\
&f_1(s_1, s_2) := [r_1 - a_1 s_1 - b_1 s_2]s_1,\\
&f_2(s_s, s_2) := [r_2 - b_2 s_1 - a_2 s_2]s_2,\\
&\eta_1(s_1, s_2, s_3) := 0,\\
&\eta_2(s_1, s_2, s_3) := 0,\\
&\eta_3(s_1, s_2, s_3) := [r_1 - a_1 s_1 - b_1 s_2]s_3,\\
&\kappa(s_1, s_2, s_3) := h_{1\to 2}(s_2)s_1 - \{h_{2\to 1}(s_2) + h_{1\to 2}(s_2)\}s_3
\end{aligned}
$$

for $(s_1, s_2, s_3) \in [0, \infty)^3$. Due to (1.7), (1.8), (1.9), (1.10) and (2.22), the assumptions of Theorem 4.1 are fulfilled. Therefore we obtain not only (1.13) but also

$$
\begin{aligned}
&\sup_{t\in[0,T]} \left\| v_1^k(\cdot, t) - \left\{1 - \frac{u_2(\cdot, t)}{M_2}\right\} u_1(\cdot, t)\right\|_{L^2(\Omega)} \le \frac{C}{k},\\
&\sup_{t\in[0,T]} \left\| v_2^k(\cdot, t) - \frac{u_2(\cdot, t)}{M_2} u_1(\cdot, t)\right\|_{L^2(\Omega)} \le \frac{C}{k}
\end{aligned}
$$

for $k \ge k_0$. □

REMARK 4.2. In the proof of Theorem 1.6, we do not need to introduce $\eta_1(s_1, s_2, s_3)$ and $\eta_2(s_1, s_2, s_3)$ since they vanish. However, if we take

$$
\begin{cases}
v_{1t} = d_1 \Delta v_1 + k[h_{2\to 1}(v_3)v_2 - h_{1\to 2}(v_3)v_1], & t > 0, x \in \Omega,\\
v_{2t} = (d_1 + \alpha M_2)\Delta v_2 & \\
\qquad + (r_1 - a_1(v_1 + v_2) - b_1 v_3)(v_1 + v_2) & \\
\qquad + k[h_{1\to 2}(v_3)v_1 - h_{2\to 1}(v_3)v_2], & t > 0, x \in \Omega,\\
v_{3t} = d_2 \Delta v_3 + (r_2 - b_2(v_1 + v_2) - a_2 v_3)v_3, & t > 0, x \in \Omega
\end{cases}
$$

as the approximating reaction-diffusion system to (1.4), the functions $\eta_1(s_1, s_2, s_3)$, $\eta_2(s_1, s_2, s_3)$ are necessary.

Acknowledgments

The authors would like to express their thanks to the referee who carefully checked this paper in a short time.

Appendix A.

A.1. *Proof of* (2.25)

Here we show

$$\Xi^k(\mu) = -\frac{M_2 h_{1\to 2}(u_2^*)k}{u_2^*}\Xi^*(\mu) + \Xi^0(\mu)$$
$$= -(h_{1\to 2}(u_2^*) + h_{2\to 1}(u_2^*))k\,\Xi^*(\mu) + \Xi^0(\mu).$$

PROOF. Recall that

$$\Xi^k := \begin{vmatrix} \xi_{11} & \xi_{12} & \xi_{13} \\ \xi_{21} & \xi_{22} & \xi_{23} \\ f_{3,v_1}(v_1^*, v_2^*, v_3^*) & f_{3,v_2}(v_1^*, v_2^*, v_3^*) & -d_2\sigma + f_{3,v_3}(v_1^*, v_2^*, v_3^*) - \mu \end{vmatrix},$$

where

$$\xi_{11} := -d_1\sigma + f_{1,v_1}(v_1^*, v_2^*, v_3^*) - kh_{1\to 2}(v_3^*) - \mu,$$

$$\xi_{12} := f_{1,v_2}(v_1^*, v_2^*, v_3^*) + kh_{2\to 1}(v_3^*),$$

$$\xi_{13} := f_{1,v_3}(v_1^*, v_2^*, v_3^*) + k(h'_{2\to 1}(v_3^*)v_2^* - h'_{1\to 2}(v_3^*)v_1^*),$$

$$\xi_{21} := f_{2,v_1}(v_1^*, v_2^*, v_3^*) + kh_{1\to 2}(v_3^*),$$

$$\xi_{22} := -(d_1 + \alpha M_2)\sigma + f_{2,v_2}(v_1^*, v_2^*, v_3^*) - kh_{2\to 1}(v_3^*) - \mu,$$

$$\xi_{23} := f_{2,v_3}(v_1^*, v_2^*, v_3^*) + k(h'_{1\to 2}(v_3^*)v_1^* - h'_{2\to 1}(v_3^*)v_2^*).$$

Using (2.22), we see

$$\xi_{11} + \xi_{21} = -d_1\sigma + f_{u_1}(u_1^*, u_2^*) - \mu,$$

$$\xi_{12} + \xi_{22} = -(d_1 + \alpha M_2)\sigma + f_{u_1}(u_1^*, u_2^*) - \mu,$$

$$\xi_{13} + \xi_{23} = f_{u_2}(u_1^*, u_2^*).$$

Then, adding the second row of Ξ^k to the first and subtracting the first column from the second, we have

$$\Xi^k = \begin{vmatrix} \xi_{11} + \xi_{21} & \xi_{12} + \xi_{22} & \xi_{13} + \xi_{23} \\ \xi_{21} & \xi_{22} & \xi_{23} \\ g_{u_1}(u_1^*, u_2^*) & g_{u_1}(u_1^*, u_2^*) & -d_2\sigma + g_{u_2}(u_1^*, u_2^*) - \mu \end{vmatrix}$$

$$= \begin{vmatrix} \xi_{11}+\xi_{21} & \xi_{12}+\xi_{22}-\xi_{11}-\xi_{21} & \xi_{13}+\xi_{23} \\ \xi_{21} & \xi_{22}-\xi_{21} & \xi_{23} \\ g_{u_1}(u_1^*,u_2^*) & 0 & -d_2\sigma+g_{u_2}(u_1^*,u_2^*)-\mu \end{vmatrix}$$

$$= \begin{vmatrix} -d_1\sigma+f_{u_1}(u_1^*,u_2^*)-\mu & -\alpha M_2\sigma & f_{u_2}(u_1^*,u_2^*) \\ \xi_{21} & \xi_{22}-\xi_{21} & \xi_{23} \\ g_{u_1}(u_1^*,u_2^*) & 0 & -d_2\sigma+g_{u_2}(u_1^*,u_2^*)-\mu \end{vmatrix}.$$

We note that the last determinant corresponds to the eigenvalue problem of (2.23). We derive the linear part of Ξ^k in k and rewrite it as follows:

$$\Xi^k = k \begin{vmatrix} -d_1\sigma+f_{u_1}(u_1^*,u_2^*)-\mu & -\alpha M_2\sigma & f_{u_2}(u_1^*,u_2^*) \\ h_{1\to2}(v_3^*) & -h_{2\to1}(v_3^*)-h_{1\to2}(v_3^*) & h'_{1\to2}(v_3^*)v_1^*-h'_{2\to1}(v_3^*)v_2^* \\ g_{u_1}(u_1^*,u_2^*) & 0 & -d_2\sigma+g_{u_2}(u_1^*,u_2^*)-\mu \end{vmatrix}$$
$$+\Xi^0$$

as $k\to\infty$.

By (2.24), we have

$$h'_{1\to2}(s) = (h'_{1\to2}(s)+h'_{2\to1}(s))\frac{s}{M_2}+\frac{h_{1\to2}(s)+h_{2\to1}(s)}{M_2},$$
$$\frac{h_{1\to2}(s)+h_{2\to1}(s)}{M_2} = \frac{h_{1\to2}(s)}{s}$$

and then

$$h'_{1\to2}(v_3^*)v_1^*-h'_{2\to1}(v_3^*)v_2^* = h'_{1\to2}(u_2^*)\left(u_1^*-\frac{u_2^*}{M_2}u_1^*\right)-h'_{2\to1}(u_2^*)\frac{u_2^*}{M_2}u_1^*$$
$$= \frac{h_{1\to2}(u_2^*)+h_{2\to1}(u_2^*)}{M_2}u_1^* = \frac{h_{1\to2}(u_2^*)u_1^*}{u_2^*}.$$

The linear part of Ξ^k in k can be reduced to

$$k\begin{vmatrix} -d_1\sigma+f_{u_1}(u_1^*,u_2^*)-\mu & -\alpha M_2\sigma & f_{u_2}(u_1^*,u_2^*) \\ h_{1\to2}(u_2^*) & -\dfrac{M_2 h_{1\to2}(u_2^*)}{u_2^*} & \dfrac{h_{1\to2}(u_2^*)u_1^*}{u_2^*} \\ g_{u_1}(u_1^*,u_2^*) & 0 & -d_2\sigma+g_{u_2}(u_1^*,u_2^*)-\mu \end{vmatrix}$$
$$= \frac{kh_{1\to2}(u_2^*)}{u_2^*}\begin{vmatrix} -d_1\sigma+f_{u_1}(u_1^*,u_2^*)-\mu & -\alpha M_2\sigma & f_{u_2}(u_1^*,u_2^*) \\ u_2^* & -M_2 & u_1^* \\ g_{u_1}(u_1^*,u_2^*) & 0 & -d_2\sigma+g_{u_2}(u_1^*,u_2^*)-\mu \end{vmatrix}.$$

Multiplying the second row by $-\alpha\sigma$ and adding the product to the first one, we can calculate the linear part as follows:

$$\frac{kh_{1\to 2}(u_2^*)}{u_2^*}\begin{vmatrix} -(d_1+\alpha u_2^*)\sigma + f_{u_1}(u_1^*,u_2^*) - \mu & 0 & -\alpha\sigma u_1^* + f_{u_2}(u_1^*,u_2^*) \\ u_2^* & -M_2 & u_1^* \\ g_{u_1}(u_1^*,u_2^*) & 0 & -d_2\sigma + g_{u_2}(u_1^*,u_2^*) - \mu \end{vmatrix}.$$

By the definition of Ξ^*, i.e.,

$$\Xi^*(\mu) := \begin{vmatrix} -d_1\sigma - \alpha u_2^*\sigma + f_{u_1}(u_1^*,u_2^*) - \mu & -\alpha u_1^*\sigma + f_{u_2}(u_1^*,u_2^*) \\ g_{u_1}(u_1^*,u_2^*) & -d_2\sigma + g_{u_2}(u_1^*,u_2^*) - \mu \end{vmatrix},$$

we obtain (2.25). □

A.2. *Proof of* Lemma 3.6

PROOF. Set

$$\hat{u}_i = u_i^k(x+\xi,t) - u_i^k(x,t), \qquad \hat{w} = w^k(x+\xi,t) - w^k(x,t),$$
$$\bar{u}_i = \frac{u_i^k(x+\xi,t) + u_i^k(x,t)}{2}, \qquad \bar{w} = \frac{w^k(x+\xi,t) + w^k(x,t)}{2},$$
$$u_{i,1}(x,t) = u_i^k(x+\xi,t), \qquad u_{i,2}(x,t) = u_i^k(x,t),$$
$$w_1(x,t) = w^k(x+\xi,t), \qquad w_2(x,t) = w^k(x,t).$$

Using

$$A_1B_1 - A_2B_2 = \frac{A_1+A_2}{2}(B_1-B_2) + (A_1-A_2)\frac{B_1+B_2}{2},$$

we get

$$\begin{aligned}\hat{u}_{1,t} &= d_1\Delta\hat{u}_1 + f(u_{1,1}) - f(u_{1,2}) - s_1k(\bar{u}_2\hat{u}_1 + \bar{u}_1\hat{u}_2) \\ &\quad - \lambda s_1 k(1-\bar{w})\hat{u}_1 + \lambda s_1 k\hat{w}\bar{u}_1\end{aligned}$$

$$\begin{aligned}\hat{u}_{2,t} &= d_2\Delta\hat{u}_2 + g(u_{2,1}) - g(u_{2,2}) - s_2k(\bar{u}_2\hat{u}_1 + \bar{u}_1\hat{u}_2) \\ &\quad - \lambda s_2 k\bar{w}\hat{u}_2 - \lambda s_2 k\bar{u}_2\hat{w}\end{aligned}$$

$$\hat{w}_t = k(1-\bar{w})\hat{u}_1 - k\hat{w}\bar{u}_1 - k\bar{w}\hat{u}_2 - k\bar{u}_2\hat{w}.$$

Let φ, $\Phi_\delta(r)$ be smooth positive functions satisfying

$$\varphi(x) = \begin{cases} 1 & (x\in\Omega_{3r}), \\ 0 & (x\in\Omega\setminus\Omega_{2r}), \end{cases}$$

and

$$\Phi_\delta(r) = |r| \quad (|r|\geq 2\delta), \qquad \Phi_\delta''(r)\geq 0, \qquad |\Phi_\delta'(r)|\leq 1,$$
$$\Phi_\delta(-r)\equiv\Phi_\delta(r).$$

By definition, we have

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega_r}\Phi_\delta(\hat{u}_1)\varphi\,\mathrm{d}x\\
&\quad=\int_{\Omega_r}\Phi_\delta'(\hat{u}_1)\big(d_1\Delta\hat{u}_1+f(u_{1,1})-f(u_{1,2})-s_1k(\bar{u}_2\hat{u}_1+\bar{u}_1\hat{u}_2)\\
&\qquad-\lambda s_1k(1-\bar{w})\hat{u}_1+\lambda s_1k\hat{w}\bar{u}_1\big)\varphi\,\mathrm{d}x\\
&\quad\le -d_1\int_{\Omega_r}\Phi_\delta''(\hat{u}_1)|\nabla\hat{u}_1|^2\varphi\,\mathrm{d}x-d_1\int_{\Omega_r}\Phi_\delta'(\hat{u}_1)\nabla\hat{u}_1\cdot\nabla\varphi\,\mathrm{d}x\\
&\qquad+|f|_{\mathrm{Lip}}\int_{\Omega_r}|\hat{u}_1|\varphi\,\mathrm{d}x-s_1k\int_{\Omega_r}\Phi_\delta'(\hat{u}_1)(\bar{u}_2\hat{u}_1+\bar{u}_1\hat{u}_2)\varphi\,\mathrm{d}x\\
&\qquad-\lambda s_1k\int_{\Omega_r}\Phi_\delta'(\hat{u}_1)(1-\bar{w})\hat{u}_1\varphi\,\mathrm{d}x+\lambda s_1k\int_{\Omega_r}\Phi_\delta'(\hat{u}_1)\bar{u}_1\hat{w}\varphi\,\mathrm{d}x\\
&\quad\le d_1\int_{\Omega_r}\Phi_\delta(\hat{u}_1)\Delta\varphi\,\mathrm{d}x+|f|_{\mathrm{Lip}}\int_{\Omega_r}|\hat{u}_1|\varphi\,\mathrm{d}x+s_1I_{1,\delta}\\
&\quad\le C\|\Phi_\delta(\hat{u}_1)\|_{L^2(\Omega_r)}+|f|_{\mathrm{Lip}}\int_{\Omega_r}|\hat{u}_1|\varphi\,\mathrm{d}x+s_1I_{1,\delta},
\end{aligned}
$$

where C is a positive constant independent of k and

$$
\begin{aligned}
I_{1,\delta}:=&-k\int_{\Omega_r}\Phi_\delta'(\hat{u}_1)(\bar{u}_2\hat{u}_1+\bar{u}_1\hat{u}_2)\varphi\,\mathrm{d}x\\
&-\lambda k\int_{\Omega_r}\Phi_\delta'(\hat{u}_1)(1-\bar{w})\hat{u}_1\varphi\,\mathrm{d}x+\lambda k\int_{\Omega_r}\Phi_\delta'(\hat{u}_1)\bar{u}_1\hat{w}\varphi\,\mathrm{d}x.
\end{aligned}
$$

Similarly we have

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega_r}\Phi_\delta(\hat{u}_2)\varphi\,\mathrm{d}x\le C\|\Phi_\delta(\hat{u}_2)\|_{L^2(\Omega_r)}+|g|_{\mathrm{Lip}}\int_{\Omega_r}|\hat{u}_2|\varphi\,\mathrm{d}x+s_2I_{2,\delta},\\
&\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega_r}\lambda\Phi_\delta(\hat{w})\varphi\,\mathrm{d}x=I_{3,\delta},
\end{aligned}
$$

where

$$
\begin{aligned}
I_{2,\delta}:=&-k\int_{\Omega_r}\Phi_\delta'(\hat{u}_2)(\bar{u}_2\hat{u}_1+\bar{u}_1\hat{u}_2)\varphi\,\mathrm{d}x\\
&-\lambda k\int_{\Omega_r}\Phi_\delta'(\hat{u}_2)\bar{w}\hat{u}_2\varphi\,\mathrm{d}x-\lambda k\int_{\Omega_r}\Phi_\delta'(\hat{u}_2)\bar{u}_2\hat{w}\varphi\,\mathrm{d}x,\\
I_{3,\delta}:=&\;k\int_{\Omega_r}\Phi_\delta'(\hat{w})\big((1-\bar{w})\hat{u}_1-\bar{u}_1\hat{w}-\bar{w}\hat{u}_2-\bar{u}_2\hat{w}\big)\varphi\,\mathrm{d}x.
\end{aligned}
$$

Adding these inequalities, we obtain

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega_r}\left(\frac{1}{s_1}\Phi_\delta(\hat{u}_1)+\frac{1}{s_2}\Phi_\delta(\hat{u}_2)+\lambda\Phi_\delta(\hat{w})\right)\varphi\,\mathrm{d}x\\
&\quad\le \frac{C}{s_1}\|\Phi_\delta(\hat{u}_1)\|_{L^2(\Omega_r)}+\frac{C}{s_2}\|\Phi_\delta(\hat{u}_2)\|_{L^2(\Omega_r)}+\frac{|f|_{\mathrm{Lip}}}{s_1}\int_{\Omega_r}|\hat{u}_1|\varphi\,\mathrm{d}x\\
&\qquad+\frac{|g|_{\mathrm{Lip}}}{s_2}\int_{\Omega_r}|\hat{u}_2|\varphi\,\mathrm{d}x+I_{1,\delta}+I_{2,\delta}+\lambda I_{3,\delta}.
\end{aligned}$$

Integration of the above inequality on $[0,T]$ implies

$$\begin{aligned}
&\int_{\Omega_r}\left(\frac{1}{s_1}\Phi_\delta(\hat{u}_1(\cdot,T))+\frac{1}{s_2}\Phi_\delta(\hat{u}_2(\cdot,T))+\lambda\Phi_\delta(\hat{w}(\cdot,T))\right)\varphi\,\mathrm{d}x\\
&\quad\le \int_{\Omega_r}\left(\frac{1}{s_1}\Phi_\delta(\hat{u}_1(\cdot,0))+\frac{1}{s_2}\Phi_\delta(\hat{u}_2(\cdot,0))+\lambda\Phi_\delta(\hat{w}(\cdot,0))\right)\varphi\,\mathrm{d}x\\
&\qquad+\frac{C}{s_1}\int_0^T\|\Phi_\delta(\hat{u}_1)\|_{L^2(\Omega_r)}\mathrm{d}t+\frac{C}{s_2}\int_0^T\|\Phi_\delta(\hat{u}_2)\|_{L^2(\Omega_r)}\mathrm{d}t\\
&\qquad+\frac{|f|_{\mathrm{Lip}}}{s_1}\int_0^T\int_{\Omega_r}|\hat{u}_1|\varphi\,\mathrm{d}x\mathrm{d}t+\frac{|g|_{\mathrm{Lip}}}{s_2}\int_0^T\int_{\Omega_r}|\hat{u}_2|\varphi\,\mathrm{d}x\mathrm{d}t\\
&\qquad+\int_0^T\big(I_{1,\delta}+I_{2,\delta}+\lambda I_{3,\delta}\big)\,\mathrm{d}t.
\end{aligned}$$

Letting $\delta\downarrow 0$, we obtain

$$\begin{aligned}
&\int_{\Omega_r}\left(\frac{1}{s_1}|\hat{u}_1(\cdot,T)|+\frac{1}{s_2}|\hat{u}_2(\cdot,T)|+\lambda|\hat{w}(\cdot,T)|\right)\varphi\,\mathrm{d}x\\
&\quad\le \int_{\Omega_r}\left(\frac{1}{s_1}|\hat{u}_1(\cdot,0)|+\frac{1}{s_2}|\hat{u}_2(\cdot,0)|+\lambda|\hat{w}(\cdot,0)|\right)\varphi\,\mathrm{d}x\\
&\qquad+\frac{C}{s_1}\int_0^T\|\hat{u}_1\|_{L^2(\Omega_r)}\mathrm{d}t+\frac{C}{s_2}\int_0^T\|\hat{u}_2\|_{L^2(\Omega_r)}\mathrm{d}t\\
&\qquad+\frac{|f|_{\mathrm{Lip}}}{s_1}\int_0^T\int_{\Omega_r}|\hat{u}_1|\varphi\,\mathrm{d}x\mathrm{d}t+\frac{|g|_{\mathrm{Lip}}}{s_2}\int_0^T\int_{\Omega_r}|\hat{u}_2|\varphi\,\mathrm{d}x\mathrm{d}t\\
&\qquad+\int_0^T\big(I_{1,0}+I_{2,0}+\lambda I_{3,0}\big)\,\mathrm{d}t.
\end{aligned}$$

Noting that

$$\lim_{\delta\downarrow 0}\Phi'_\delta(r)=\mathrm{sgn}(r)=\begin{cases}1 & \text{if } r>0,\\ 0 & \text{if } r=0,\\ -1 & \text{if } r<0,\end{cases}$$

we have

$$
\begin{aligned}
&\int_0^T \big(I_{1,0} + I_{2,0} + \lambda I_{3,0}\big)\,\mathrm{d}t \\
&\quad = -k\int_{\Omega_r} \operatorname{sgn}(\hat u_1)(\bar u_2 \hat u_1 + \bar u_1 \hat u_2)\varphi\,\mathrm{d}x - \lambda k \int_{\Omega_r} \operatorname{sgn}(\hat u_1)(1-\bar w)\hat u_1 \varphi\,\mathrm{d}x \\
&\qquad + \lambda k \int_{\Omega_r} \operatorname{sgn}(\hat u_1)\hat w \bar u_1 \varphi\,\mathrm{d}x - k\int_{\Omega_r} \operatorname{sgn}(\hat u_2)(\bar u_2 \hat u_1 + \bar u_1 \hat u_2)\varphi\,\mathrm{d}x \\
&\qquad - \lambda k \int_{\Omega_r} \operatorname{sgn}(\hat u_2)\bar w \hat u_2 \varphi\,\mathrm{d}x - \lambda k \int_{\Omega_r} \operatorname{sgn}(\hat u_2)\hat w \bar u_2 \varphi\,\mathrm{d}x \\
&\qquad + \lambda k \int_{\Omega_r} \operatorname{sgn}(\hat w)\big((1-\bar w)\hat u_1 - \hat w \bar u_1 - \bar w \hat u_2 - \hat w \bar u_2\big)\,\varphi\,\mathrm{d}x.
\end{aligned}
$$

For any A and $|\chi| \le 1$,

$$
|A| - A\chi \ge 0.
$$

Thus

$$
\begin{aligned}
&\int_0^T \big(I_{1,0} + I_{2,0} + \lambda I_{3,0}\big)\,\mathrm{d}t \\
&\quad = -k\int_{\Omega_r} (\bar u_2|\hat u_1| + \bar u_1 \hat u_2 \operatorname{sgn}(\hat u_1))\varphi\,\mathrm{d}x - \lambda k \int_{\Omega_r} (1-\bar w)|\hat u_1|\varphi\,\mathrm{d}x \\
&\qquad + \lambda k \int_{\Omega_r} \operatorname{sgn}(\hat u_1)\hat w \bar u_1 \varphi\,\mathrm{d}x - k\int_{\Omega_r} (\bar u_2 \hat u_1 \operatorname{sgn}(\hat u_2) + \bar u_1 |\hat u_2|)\varphi\,\mathrm{d}x \\
&\qquad - \lambda k \int_{\Omega_r} \bar w|\hat u_2|\varphi\,\mathrm{d}x - \lambda k \int_{\Omega_r} \operatorname{sgn}(\hat u_2)\hat w \bar u_2 \varphi\,\mathrm{d}x \\
&\qquad + \lambda k \int_{\Omega_r} \big((1-\bar w)\hat u_1 \operatorname{sgn}(\hat w) - |\hat w|\bar u_1 - \bar w \hat u_2 \operatorname{sgn}(\hat w) - |\hat w|\bar u_2\big)\,\varphi\,\mathrm{d}x \\
&\quad = -k\int_{\Omega_r} \bar u_2(|\hat u_1| + \hat u_1 \operatorname{sgn}(\hat u_2))\varphi\,\mathrm{d}x - k\int_{\Omega_r} \bar u_1(|\hat u_2| + \hat u_2 \operatorname{sgn}(\hat u_1))\varphi\,\mathrm{d}x \\
&\qquad - \lambda k \int_{\Omega_r} (1-\bar w)(|\hat u_1| - \hat u_1 \operatorname{sgn}(\hat w))\varphi\,\mathrm{d}x - \lambda k \int_{\Omega_r} \bar u_1(|\hat w| \\
&\qquad - \hat w \operatorname{sgn}(\hat u_1))\varphi\,\mathrm{d}x - \lambda k \int_{\Omega_r} \bar w(|\hat u_2| + \hat u_2 \operatorname{sgn}(\hat w))\varphi\,\mathrm{d}x \\
&\qquad - \lambda k \int_{\Omega_r} \bar u_2(|\hat w| + \hat w \operatorname{sgn}(\hat u_2))\varphi\,\mathrm{d}x \le 0,
\end{aligned}
$$

where the inequalities $1 - \bar w \ge 0$ and $\bar u^i > 0$ ($i = 1, 2$) have been used. We obtain

$$
\begin{aligned}
&\int_{\Omega_r} \left(\frac{1}{s_1}|\hat u_1(\cdot,T)| + \frac{1}{s_2}|\hat u_2(\cdot,T)| + \lambda|\hat w(\cdot,T)|\right)\varphi\,\mathrm{d}x \\
&\quad \le \int_{\Omega_r} \left(\frac{1}{s_1}|\hat u_1(\cdot,0)| + \frac{1}{s_2}|\hat u_2(\cdot,0)| + \lambda|\hat w(\cdot,0)|\right)\varphi\,\mathrm{d}x
\end{aligned}
$$

$$+\frac{C}{s_1}\int_0^T \|\hat{u}_1\|_{L^2(\Omega_r)}\mathrm{d}t + \frac{C}{s_2}\int_0^T \|\hat{u}_2\|_{L^2(\Omega_r)}\mathrm{d}t$$
$$+\frac{|f|_{\mathrm{Lip}}}{s_1}\int_0^T\int_{\Omega_r}|\hat{u}_1|\varphi\,\mathrm{d}x\mathrm{d}t + \frac{|g|_{\mathrm{Lip}}}{s_2}\int_0^T\int_{\Omega_r}|\hat{u}_2|\varphi\,\mathrm{d}x\mathrm{d}t.$$

The result of the lemma then follows from the assumption for u_{01}^k, u_{02}^k, and w_0^k and from Lemma 3.4. □

A.3. *Proof of* Theorem 4.1

PROOF. By (4.16) and (4.17) we can assume that f_1, f_2, η_1, η_2, η_3, κ, ϕ and their derivatives which appear in the proof are uniformly bounded independently of (x,t) and in particular of k. (For a more precise argument we need suitable truncations of f_1, f_2, etc. around $[0,\widetilde{M}_0]^3$; see, e.g., Section 2 of [22].) We denote by c_i $(i=1,2,\ldots)$ the positive constants which do not depend on k and on (x,t). It follows from (4.14) that

$$\kappa_0 := \inf_{(s_1,s_2,s_3)\in[0,\widetilde{M}_0]^3}\{-\kappa_{\widetilde{v}_3}(s_1,s_2,s_3)\} > 0. \tag{A.1}$$

Set

$$\widetilde{u}_3 := \phi(\widetilde{u}_2)\widetilde{u}_1, \qquad \widetilde{V}_1 := \widetilde{v}_1 - \widetilde{u}_1, \qquad \widetilde{V}_2 := \widetilde{v}_2 - \widetilde{u}_2,$$
$$\widetilde{V}_3 := \widetilde{v}_3 - \phi(\widetilde{u}_2)\widetilde{u}_1 = \widetilde{v}_3 - \widetilde{u}_3.$$

It follows from (4.12) that $\eta_1(\widetilde{u}_1,\widetilde{u}_2,\widetilde{u}_3) = \eta_2(\widetilde{u}_1,\widetilde{u}_2,\widetilde{u}_3) = \kappa(\widetilde{u}_1,\widetilde{u}_2,\widetilde{u}_3) = 0$. Then we get

$$\begin{cases} \widetilde{V}_{1t} = d_1\Delta\widetilde{V}_1 + \tilde{\alpha}\Delta\widetilde{V}_3 + f_1(\widetilde{u}_1+\widetilde{V}_1, \widetilde{u}_2+\widetilde{V}_2) - f(\widetilde{u}_1,\widetilde{u}_2) \\ \qquad + \eta_1(\widetilde{u}_1+\widetilde{V}_1, \widetilde{u}_2+\widetilde{V}_2, \widetilde{u}_3+\widetilde{V}_3) - \eta_1(\widetilde{u}_1,\widetilde{u}_2,\widetilde{u}_3), \\ \widetilde{V}_{2t} = d_2\Delta\widetilde{V}_2 + f_2(\widetilde{u}_1+\widetilde{V}_1, \widetilde{u}_2+\widetilde{V}_2) - g(\widetilde{u}_1,\widetilde{u}_2) \\ \qquad + \eta_2(\widetilde{u}_1+\widetilde{V}_1, \widetilde{u}_2+\widetilde{V}_2, \widetilde{u}_3+\widetilde{V}_3) - \eta_2(\widetilde{u}_1,\widetilde{u}_2,\widetilde{u}_3), \\ \widetilde{V}_{3t} = (d_1+\tilde{\alpha})\Delta\widetilde{V}_3 \\ \qquad + k\{\kappa(\widetilde{u}_1+\widetilde{V}_1, \widetilde{u}_2+\widetilde{V}_2, \widetilde{u}_3+\widetilde{V}_3) - \kappa(\widetilde{u}_1+\widetilde{V}_1, \widetilde{u}_2+\widetilde{V}_2, \widetilde{u}_3)\} \\ \qquad + k\{\kappa(\widetilde{u}_1+\widetilde{V}_1, \widetilde{u}_2+\widetilde{V}_2, \widetilde{u}_3) - \kappa(\widetilde{u}_1,\widetilde{u}_2,\widetilde{u}_3)\} \\ \qquad + \eta_3(\widetilde{u}_1+\widetilde{V}_1, \widetilde{u}_2+\widetilde{V}_2, \widetilde{u}_3+\widetilde{V}_3) - \widetilde{u}_{3t} + (d_1+\tilde{\alpha})\Delta\widetilde{u}_3 \end{cases} \tag{A.2}$$

for $t>0$ and $x\in\Omega$, with the boundary condition

$$\frac{\partial\widetilde{V}_1}{\partial\nu} = \frac{\partial\widetilde{V}_2}{\partial\nu} = \frac{\partial\widetilde{V}_3}{\partial\nu} = 0, \quad t>0, x\in\partial\Omega,$$

and the initial condition

$$\widetilde{V}_1(0,x) = \widetilde{V}_2(0,x) = \widetilde{V}_3(0,x) = 0, \quad x\in\Omega.$$

We denote a primitive function of $\kappa(\widetilde{v_1}, \widetilde{v_2}, \widetilde{v_3})$ with respect to $\widetilde{v_1}$ by

$$K(\widetilde{v_1}, \widetilde{v_2}; \widetilde{v_3}) := \int_0^{\widetilde{v_1}} \kappa(s, \widetilde{v_2}, \widetilde{v_3})\,\mathrm{d}s,$$

where we regard K as a function of $(\widetilde{v_1}, \widetilde{v_2}) \in [0,\infty)^2$ parametrized by $\widetilde{v_3} \in [0,\infty)$. Using K, we define $E^k(t)$ of $(\widetilde{V_1}(\cdot,t),\ \widetilde{V_2}(\cdot,t))$ in $L^2(\Omega)$ by

$$E^k(t) := \int_\Omega \left\{K(\widetilde{u_1}+\widetilde{V_1}, \widetilde{u_2}+\widetilde{V_2}; \widetilde{u_3}) - \widetilde{K} - \widetilde{K}_{\widetilde{v_1}}\widetilde{V_1} - \widetilde{K}_{\widetilde{v_2}}\widetilde{V_2}\right\}\mathrm{d}x,$$

where we abbreviate $K(\widetilde{u_1}, \widetilde{u_2}; \widetilde{u_3})$, $K_{\widetilde{v_1}}(\widetilde{u_1}, \widetilde{u_2}; \widetilde{u_3})$, etc. by $\widetilde{K}$, $\widetilde{K}_{\widetilde{v_1}}$, etc.; hereafter we often use similar abbreviations for simplicity of notation, such as $\kappa(\widetilde{u_1}, \widetilde{u_2}, \widetilde{u_3})$, $f(\widetilde{u_1}, \widetilde{u_2}, \widetilde{u_3})$ to $\widetilde{\kappa}$, $\widetilde{f}$, respectively. We also denote the norm and inner product in $L^2(\Omega)$ by $\|\cdot\|$ and $(\cdot,\cdot)$. Differentiating $E^k(t)$ in t yields

$$\begin{aligned}
\frac{\mathrm{d}E^k}{\mathrm{d}t} &= \big(K_{\widetilde{v_1}}(\widetilde{u_1}+\widetilde{V_1}, \widetilde{u_2}+\widetilde{V_2}; \widetilde{u_3}) - \widetilde{K}_{\widetilde{v_1}} - \widetilde{K}_{\widetilde{v_1}\widetilde{v_1}}\widetilde{V_1} - \widetilde{K}_{\widetilde{v_2}\widetilde{v_1}}\widetilde{V_2},\ \widetilde{u}_{1t}\big) \\
&\quad + \big(K_{\widetilde{v_2}}(\widetilde{u_1}+\widetilde{V_1}, \widetilde{u_2}+\widetilde{V_2}; \widetilde{u_3}) - \widetilde{K}_{\widetilde{v_2}} - \widetilde{K}_{\widetilde{v_1}\widetilde{v_2}}\widetilde{V_1} - \widetilde{K}_{\widetilde{v_2}\widetilde{v_2}}\widetilde{V_2},\ \widetilde{u}_{2t}\big) \\
&\quad + \big(K_{\widetilde{u_3}}(\widetilde{u_1}+\widetilde{V_1}, \widetilde{u_2}+\widetilde{V_2}; \widetilde{u_3}) - \widetilde{K}_{\widetilde{u_3}} - \widetilde{K}_{\widetilde{v_1}\widetilde{u_3}}\widetilde{V_1} - \widetilde{K}_{\widetilde{v_2}\widetilde{u_3}}\widetilde{V_2},\ \widetilde{u}_{3t}\big) \\
&\quad + \big(K_{\widetilde{v_1}}(\widetilde{u_1}+\widetilde{V_1}, \widetilde{u_2}+\widetilde{V_2}; \widetilde{u_3}) - \widetilde{K}_{\widetilde{v_1}},\ \widetilde{V}_{1t}\big) \\
&\quad + \big(K_{\widetilde{v_2}}(\widetilde{u_1}+\widetilde{V_1}, \widetilde{u_2}+\widetilde{V_2}; \widetilde{u_3}) - \widetilde{K}_{\widetilde{v_2}},\ \widetilde{V}_{2t}\big) \\
&\le c_1(\|\widetilde{V_1}\|^2 + \|\widetilde{V_2}\|^2 + \|\widetilde{V_3}\|^2) \\
&\quad + \big(\kappa(\widetilde{u_1}+\widetilde{V_1}, \widetilde{u_2}+\widetilde{V_2}, \widetilde{u_3}) - \widetilde{\kappa},\ d_1\Delta\widetilde{V_1} + \tilde{\alpha}\Delta\widetilde{V_3}\big) \\
&\quad + \big(K_{\widetilde{v_2}}(\widetilde{u_1}+\widetilde{V_1}, \widetilde{u_2}+\widetilde{V_2}; \widetilde{u_3}) - \widetilde{K}_{\widetilde{v_2}},\ d_2\Delta\widetilde{V_2}\big).
\end{aligned}$$

We can estimate

$$\begin{aligned}
&\big(K_{\widetilde{v_2}}(\widetilde{u_1}+\widetilde{V_1}, \widetilde{u_2}+\widetilde{V_2}; \widetilde{u_3}) - \widetilde{K}_{\widetilde{v_2}}, d_2\Delta\widetilde{V_2}\big) \\
&\quad = -d_2\Big[\big(\{K_{\widetilde{v_2}\widetilde{v_1}}(\widetilde{u_1}+\widetilde{V_1}, \widetilde{u_2}+\widetilde{V_2}; \widetilde{u_3}) - \widetilde{K}_{\widetilde{v_2}\widetilde{v_1}}\}\nabla\widetilde{u_1}, \nabla\widetilde{V_2}\big) \\
&\qquad + \big(\{K_{\widetilde{v_2}\widetilde{v_2}}(\widetilde{u_1}+\widetilde{V_1}, \widetilde{u_2}+\widetilde{V_2}; \widetilde{u_3}) - \widetilde{K}_{\widetilde{v_2}\widetilde{v_2}}\}\nabla\widetilde{u_2}, \nabla\widetilde{V_2}\big) \\
&\qquad + \big(\{K_{\widetilde{v_2}\widetilde{u_3}}(\widetilde{u_1}+\widetilde{V_1}, \widetilde{u_2}+\widetilde{V_2}; \widetilde{u_3}) - \widetilde{K}_{\widetilde{v_2}\widetilde{u_3}}\}\nabla\widetilde{u_3}, \nabla\widetilde{V_2}\big) \\
&\qquad + \big(K_{\widetilde{v_2}\widetilde{v_1}}(\widetilde{u_1}+\widetilde{V_1}, \widetilde{u_2}+\widetilde{V_2}; \widetilde{u_3})\nabla\widetilde{V_1}, \nabla\widetilde{V_2}\big) \\
&\qquad + \big(K_{\widetilde{v_2}\widetilde{v_2}}(\widetilde{u_1}+\widetilde{V_1}, \widetilde{u_2}+\widetilde{V_2}; \widetilde{u_3})\nabla\widetilde{V_2}, \nabla\widetilde{V_2}\big)\Big] \\
&\quad \le c_2(\|\widetilde{V_1}\| + \|\widetilde{V_2}\| + \|\nabla\widetilde{V_1}\| + \|\nabla\widetilde{V_2}\|)\|\nabla\widetilde{V_2}\|
\end{aligned}$$

and

$$\begin{aligned}
&\big(\kappa(\widetilde{u_1}+\widetilde{V_1}, \widetilde{u_2}+\widetilde{V_2}, \widetilde{u_3}) - \widetilde{\kappa}, d_1\Delta\widetilde{V_1}\big) \\
&\quad = -d_1\Big[\big(\{\kappa_{\widetilde{v_1}}(\widetilde{u_1}+\widetilde{V_1}, \widetilde{u_2}+\widetilde{V_2}, \widetilde{u_3}) - \widetilde{\kappa}_{\widetilde{v_1}}\}\nabla\widetilde{u_1}, \nabla\widetilde{V_1}\big)
\end{aligned}$$

$$
\begin{aligned}
&+\left(\{\kappa_{\widetilde{v_2}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3})-\widetilde{\kappa}_{\widetilde{v_2}}\}\nabla\widetilde{u_2},\nabla\widetilde{V}_1\right)\\
&+\left(\{\kappa_{\widetilde{v_3}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3})-\widetilde{\kappa}_{\widetilde{v_3}}\}\nabla\widetilde{u_3},\nabla\widetilde{V}_1\right)\\
&+\left(\kappa_{\widetilde{v_1}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3})\nabla\widetilde{V}_1,\nabla\widetilde{V}_1\right)\\
&+\left(\kappa_{\widetilde{v_2}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3})\nabla\widetilde{V}_2,\nabla\widetilde{V}_1\right)\Big]\\
\le{}&-d_1\left(\kappa_{\widetilde{v_1}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3})\nabla\widetilde{V}_1,\nabla\widetilde{V}_1\right)\\
&+c_3(\|\widetilde{V}_1\|+\|\widetilde{V}_2\|+\|\nabla\widetilde{V}_2\|)\|\nabla\widetilde{V}_1\|\\
\le{}&c_3(\|\widetilde{V}_1\|+\|\widetilde{V}_2\|+\|\nabla\widetilde{V}_2\|)\|\nabla\widetilde{V}_1\|,
\end{aligned}
$$

where we have used (4.13). Thus we obtain

$$
\begin{aligned}
\frac{dE^k}{dt}\le{}&\tilde{\alpha}\left(\kappa(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3})-\widetilde{\kappa},\Delta\widetilde{V}_3\right)\\
&+c_1(\|\widetilde{V}_1\|^2+\|\widetilde{V}_2\|^2+\|\widetilde{V}_3\|^2)\\
&+c_2(\|\widetilde{V}_1\|+\|\widetilde{V}_2\|+\|\nabla\widetilde{V}_1\|+\|\nabla\widetilde{V}_2\|)\|\nabla\widetilde{V}_2\|\\
&+c_3(\|\widetilde{V}_1\|+\|\widetilde{V}_2\|+\|\nabla\widetilde{V}_2\|)\|\nabla\widetilde{V}_1\|.
\end{aligned}
\tag{A.3}
$$

Multiply the third equation of (A.2) by $-\Delta\widetilde{V}_3$ and integrate it by parts over Ω. Recalling (4.15), we have

$$
\kappa_{\widetilde{v_1}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3}+\widetilde{V}_3)-\kappa_{\widetilde{v_1}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3})=0,
$$

which implies

$$
\begin{aligned}
&\left(\kappa(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3}+\widetilde{V}_3)-\kappa(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3}),-\Delta\widetilde{V}_3\right)\\
&=\Big((\kappa_{\widetilde{v_1}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3}+\widetilde{V}_3)\\
&\quad-\kappa_{\widetilde{v_1}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3}))\nabla(\widetilde{u_1}+\widetilde{V}_1),\nabla\widetilde{V}_3\Big)\\
&\quad+\Big((\kappa_{\widetilde{v_2}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3}+\widetilde{V}_3)\\
&\quad-\kappa_{\widetilde{v_2}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3}))\nabla(\widetilde{u_2}+\widetilde{V}_2),\nabla\widetilde{V}_3\Big)\\
&\quad+\left(\kappa_{\widetilde{v_3}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3}+\widetilde{V}_3)\nabla\widetilde{V}_3,\nabla\widetilde{V}_3\right)\\
&\quad+\Big(\{\kappa_{\widetilde{v_3}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3}+\widetilde{V}_3)\\
&\quad-\kappa_{\widetilde{v_3}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3})\}\nabla\widetilde{u_3},\nabla\widetilde{V}_3\Big)\\
&=\Big((\kappa_{\widetilde{v_2}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3}+\widetilde{V}_3)\\
&\quad-\kappa_{\widetilde{v_2}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3}))\nabla(\widetilde{u_2}+\widetilde{V}_2),\nabla\widetilde{V}_3\Big)
\end{aligned}
$$

$$
\begin{aligned}
&+\left(\kappa_{\widetilde{v_3}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3}+\widetilde{V}_3)\nabla\widetilde{V}_3,\nabla\widetilde{V}_3\right)\\
&+\left(\{\kappa_{\widetilde{v_3}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3}+\widetilde{V}_3)\right.\\
&\left.-\kappa_{\widetilde{v_3}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3})\}\nabla\widetilde{u_3},\nabla\widetilde{V}_3\right)
\end{aligned}
$$

By (A.1), we have

$$
\left(\kappa_{\widetilde{v_3}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3}+\widetilde{V}_3)\nabla\widetilde{V}_3,\nabla\widetilde{V}_3\right)\le-\kappa_0 k\|\nabla\widetilde{V}_3\|^2.
$$

Combining these inequalities, we obtain

$$
\begin{aligned}
\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\nabla\widetilde{V}_3\|^2&=(\widetilde{V}_{3t},\,-\Delta\widetilde{V}_3)=-(d_1+\tilde{\alpha})\|\Delta\widetilde{V}_3\|^2\\
&\quad+k\left(\{\kappa_{\widetilde{v_2}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3}+\widetilde{V}_3)\right.\\
&\quad\left.-\kappa_{\widetilde{v_2}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3})\}(\nabla\widetilde{u_2}+\nabla\widetilde{V}_2),\nabla\widetilde{V}_3\right)\\
&\quad+k\left(\kappa_{\widetilde{v_3}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3}+\widetilde{V}_3)\nabla\widetilde{V}_3,\nabla\widetilde{V}_3\right)\\
&\quad+k\left(\{\kappa_{\widetilde{v_3}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3}+\widetilde{V}_3)\right.\\
&\quad\left.-\kappa_{\widetilde{v_3}}(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3})\}\nabla\widetilde{u_3},\nabla\widetilde{V}_3\right)\\
&\quad-k\left(\kappa(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3})-\widetilde{\kappa},\Delta\widetilde{V}_3\right)\\
&\quad-\left(\eta_3(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3}+\widetilde{V}_3)-\widetilde{\eta_3},\Delta\widetilde{V}_3\right)\\
&\quad+\left(\nabla\{(d_1+\tilde{\alpha})\Delta\widetilde{u_3}-\widetilde{u_3}_t+\widetilde{\eta_3}\},\nabla\widetilde{V}_3\right)\\
&\le-(d_1+\tilde{\alpha})\|\Delta\widetilde{V}_3\|^2-k\left(\kappa(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3})-\widetilde{\kappa},\Delta\widetilde{V}_3\right)\\
&\quad-\kappa_0 k\|\nabla\widetilde{V}_3\|^2+c_4k(\|\widetilde{V}_3\|+\|\nabla\widetilde{V}_2\|)\|\nabla\widetilde{V}_3\|\\
&\quad+c_5(\|\widetilde{V}_1\|+\|\widetilde{V}_2\|+\|\widetilde{V}_3\|)\|\Delta\widetilde{V}_3\|+c_6\|\nabla\widetilde{V}_3\|\\
&\le-\frac{\kappa_0 k}{2}\|\nabla\widetilde{V}_3\|^2-k\left(\kappa(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3})-\widetilde{\kappa},\Delta\widetilde{V}_3\right)\\
&\quad+\frac{c_7}{k}+c_8(\|\widetilde{V}_1\|^2+\|\widetilde{V}_2\|^2+\|\widetilde{V}_3\|^2)+c_9k(\|\widetilde{V}_3\|^2+\|\nabla\widetilde{V}_2\|^2).
\end{aligned}
\tag{A.4}
$$

On the other hand, it follows from the first and second equations of (A.2) that

$$
\begin{aligned}
\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\widetilde{V}_1\|^2&=-d_1\|\nabla\widetilde{V}_1\|^2-\tilde{\alpha}(\nabla\widetilde{V}_3,\nabla\widetilde{V}_1)+\left(f_1(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2)\right.\\
&\qquad\left.-\widetilde{f}_1,\widetilde{V}_1\right)+\left(\eta_1(\widetilde{u_1}+\widetilde{V}_1,\widetilde{u_2}+\widetilde{V}_2,\widetilde{u_3}+\widetilde{V}_3)-\widetilde{\eta_1},\widetilde{V}_1\right)\\
&\le-\frac{d_1}{2}\|\nabla\widetilde{V}_1\|^2+\frac{\tilde{\alpha}^2}{2d_1}\|\nabla\widetilde{V}_3\|^2\\
&\qquad+c_{10}(\|\widetilde{V}_1\|+\|\widetilde{V}_2\|+\|\widetilde{V}_3\|)\|\widetilde{V}_1\|
\end{aligned}
\tag{A.5}
$$

and that

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\widetilde{V_2}\|^2 = -d_2\|\nabla\widetilde{V_2}\|^2 + \big(f_2(\widetilde{u_1}+\widetilde{V_1},\widetilde{u_2}+\widetilde{V_2}) - \widetilde{f_2},\widetilde{V_2}\big)$$
$$+\big(\eta_2(\widetilde{u_1}+\widetilde{V_1},\widetilde{u_2}+\widetilde{V_2},\widetilde{u_3}+\widetilde{V_3}) - \widetilde{\eta_2},\widetilde{V_2}\big)$$
$$\leq -d_2\|\nabla\widetilde{V_2}\|^2 + c_{11}(\|\widetilde{V_1}\| + \|\widetilde{V_2}\| + \|\widetilde{V_3}\|)\|\widetilde{V_2}\|. \tag{A.6}$$

Similarly, using (A.1) and the third equation of (A.2), we have

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\widetilde{V_3}\|^2 = -(d_1+\tilde{\alpha})\|\nabla\widetilde{V_3}\|^2 + k\,\big(\kappa_{\widetilde{v_3}}(\widetilde{u_1}+\widetilde{V_1},\ \widetilde{u_2}+\widetilde{V_2},\widetilde{u_3}$$
$$+\theta\widetilde{V_3})\widetilde{V_3},\widetilde{V_3}\big) + k\,\big(\kappa(\widetilde{u_1}+\widetilde{V_1},\widetilde{u_2}+\widetilde{V_2},\widetilde{u_3}) - \widetilde{\kappa},\widetilde{V_3}\big)$$
$$+\big(\eta_3(\widetilde{u_1}+\widetilde{V_1},\widetilde{u_2}+\widetilde{V_2},\widetilde{u_3}+\widetilde{V_3}) - \widetilde{u_3}_t$$
$$+(d_1+\tilde{\alpha})\Delta\widetilde{u_3},\widetilde{V_3}\big)$$
$$\leq -\frac{\kappa_0 k}{2}\|\widetilde{V_3}\|^2 + \frac{c_{12}}{k} + c_{13}k(\|\widetilde{V_1}\|^2 + \|\widetilde{V_2}\|^2), \tag{A.7}$$

where the function $\theta = \theta(x,t,k)$ is such that $\theta \in (0,1)$. Choose a positive number γ_1 satisfying

$$\gamma_1 \leq \frac{d_1\kappa_0}{\tilde{\alpha}},$$

and combine (A.3), (A.4), (A.5), (A.6) and (A.7). For sufficiently large numbers γ_2 and γ_3, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\{E^k(t) + \frac{\gamma_1}{2}\|\widetilde{V_1}\|^2 + \frac{\gamma_2}{2}\|\widetilde{V_2}\|^2 + \frac{\gamma_3}{2k}\|\widetilde{V_3}\|^2 + \frac{\tilde{\alpha}}{2k}\|\nabla\widetilde{V_3}\|^2\right\}$$
$$\leq -\frac{\tilde{\alpha}\kappa_0}{2}\|\nabla\widetilde{V_3}\|^2 - \frac{\gamma_1 d_1}{2}\|\nabla\widetilde{V_1}\|^2 - \gamma_2 d_2\|\nabla\widetilde{V_2}\|^2 - \frac{\gamma_3\kappa_0}{2}\|\widetilde{V_3}\|^2$$
$$+\frac{\gamma_1\tilde{\alpha}^2}{2d_1}\|\nabla\widetilde{V_3}\|^2 + c_2(\|\widetilde{V_1}\| + \|\widetilde{V_2}\| + \|\nabla\widetilde{V_1}\| + \|\nabla\widetilde{V_2}\|)\|\nabla\widetilde{V_2}\|$$
$$+c_3(\|\widetilde{V_1}\| + \|\widetilde{V_2}\| + \|\nabla\widetilde{V_2}\|)\|\nabla\widetilde{V_1}\| + c_9\tilde{\alpha}(\|\nabla\widetilde{V_2}\|^2 + \|\widetilde{V_3}\|^2)$$
$$+\left(c_1 + \frac{c_8\tilde{\alpha}}{k}\right)(\|\widetilde{V_1}\|^2 + \|\widetilde{V_2}\|^2 + \|\widetilde{V_3}\|^2)$$
$$+(\gamma_1 c_{10}\|\widetilde{V_1}\| + \gamma_2 c_{11}\|\widetilde{V_2}\|)(\|\widetilde{V_1}\| + \|\widetilde{V_2}\| + \|\widetilde{V_3}\|)$$
$$+\gamma_3 c_{13}(\|\widetilde{V_1}\|^2 + \|\widetilde{V_2}\|^2) + \frac{c_7\tilde{\alpha} + \gamma_3 c_{12}}{k^2}$$
$$\leq c_{14}(\|\widetilde{V_1}\|^2 + \|\widetilde{V_2}\|^2) + \frac{c_{15}}{k^2} \tag{A.8}$$

for $k \geq k_0$. It also follows from (4.13) that

$$E^k(t) \geq -c_{16}\|\widetilde{V}_1\|\|\widetilde{V}_2\| - c_{17}\|\widetilde{V}_2\|^2 \geq -\frac{\gamma_1}{4}\|\widetilde{V}_1\|^2 - \left(c_{17} + \frac{c_{16}^2}{\gamma_1}\right)\|\widetilde{V}_2\|^2.$$

Thus, taking γ_2 larger if necessary, we obtain

$$E^k(t) + \frac{\gamma_1}{2}\|\widetilde{V}_1\|^2 + \frac{\gamma_2}{2}\|\widetilde{V}_2\|^2 \geq \frac{\gamma_1}{4}(\|\widetilde{V}_1\|^2 + \|\widetilde{V}_2\|^2). \tag{A.9}$$

Therefore, we can deduce from (A.8) and Gronwall's inequality that

$$E^k(t) + \frac{\gamma_1}{2}\|\widetilde{V}_1(\cdot, t)\|^2 + \frac{\gamma_2}{2}\|\widetilde{V}_2(\cdot, t)\|^2 \leq \frac{c_{18}}{k^2}$$

for $t \in [0, T]$. This inequality and (A.9) imply that

$$\|\widetilde{V}_1(\cdot, t)\|^2 + \|\widetilde{V}_2(\cdot, t)\|^2 \leq \frac{4c_{18}}{\gamma_1 k^2}.$$

Applying this result to the right-hand side of (A.7), we find

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\widetilde{V}_3\|^2 + \kappa_0 k\|\widetilde{V}_3\|^2 \leq \frac{c_{19}}{k}.$$

Hence

$$\|\widetilde{V}_3(\cdot, t)\|^2 \leq \frac{c_{19}}{\kappa_0 k^2},$$

which completes the proof. □

References

[1] H. Amann, *Nonhomogeneous linear and quasilinear elliptic and parabolic boundary value problems*, in: H. Schmeisser, H. Triebel (Eds.), Function Spaces, Differential Operators and Nonlinear Analysis, in: Teubner-Texte Math., vol. 133, 1993, pp. 9–126.

[2] D. Bothe, *Flow invariance for perturbed nonlinear evolution equations*, Abstr. Appl. Anal. **1** (1996) 417–433.

[3] D. Bothe, *The instantaneous limit of a reaction-diffusion system*, in: Evolution Equations and Their Applications in Physical and Life Sciences (Bad Herrenalb, 1998), in: Lecture Notes in Pure and Appl. Math., vol. 215, Dekker, New York, 2001, pp. 215–224.

[4] D. Bothe, D. Hilhorst, *A reaction-diffusion system with fast reversible reaction*, J. Math. Anal. Appl. **286** (2003) 125–135.

[5] H. Brezis, *Analyse Fonctionnelle*, Masson, 1983.

[6] E.C.M. Crooks, E.N. Dancer, D. Hilhorst, *On long-time dynamics for competition-diffusion systems with inhomogeneous Dirichlet boundary conditions*, Topol. Methods Nonlinear Anal. **30** (2007) 1–36.

[7] E.C.M. Crooks, E.N. Dancer, D. Hilhorst, M. Mimura, H. Ninomiya, *Spatial segregation limit of a competition-diffusion system with Dirichlet boundary conditions*, Nonlinear Anal. Real World Appl. **5** (2004) 645–665.

[8] Y.S. Choi, R. Lui, Y. Yamada, *Existence of global solutions for the Shigesada-Kawasaki-Teramoto model with weak cross-diffusion*, Discrete Contin. Dynam. Systems **9** (2003) 1193–1200.

[9] E.N. Dancer, D. Hilhorst, M. Mimura, L.A. Peletier, *Spatial segregation limit of a competition-diffusion system*, European J. Appl. Math. **10** (1999) 97–115.

[10] E. DiBenedetto, *Continuity of weak solutions to a general porous medium equation*, Indiana Univ. Math. J. **32** (1983) 83–118.

[11] S.I. Ei, Q. Fang, M. Mimura, *Effect of domain-shape on coexistence problems in a competition-diffusion system*, J. Math. Biol. **29** (1991) 219–237.

[12] S.I. Ei, E. Yanagida, *Dynamics of interfaces in competition-diffusion systems*, SIAM J. Appl. Math. **54** (1994) 1355–1373.

[13] R. Eymard, D. Hilhorst, R. van der Hout, L.A. Peletier, *A reaction-diffusion system approximation of a one-phase Stefan problem*, in: J.L. Menaldi, E. Rofman, A. Sulem (Eds.), Optimal Control and Partial Differential Equations, IOS Press Amsterdam, Berlin, Oxford, Tokyo, Washington DC, 2001, pp. 156–170.

[14] D. Hilhorst, M. Mimura, R. Schätzle, *Vanishing latent heat limit in a Stefan-like problem arising in biology*, Nonlinear Anal. Real World Appl. **4** (2003) 261–285.

[15] D. Hilhorst, R. van der Hout, L.A. Peletier, *The fast reaction limit for a reaction-diffusion system*, J. Math. Anal. Appl. **199** (1996) 349–373.

[16] D. Hilhorst, R. van der Hout, L.A. Peletier, *Diffusion in the presence of fast reaction: The case of a general monotone reaction term*, J. Math. Sci. Univ. Tokyo **4** (1997) 469–517.

[17] D. Hilhorst, R. van der Hout, L.A. Peletier, *Nonlinear diffusion in the presence of fast reaction*, Nonlinear Anal. TMA **41** (2000) 803–823.

[18] D. Hilhorst, M. Iida, M. Mimura, H. Ninomiya, *A competition-diffusion system approximation to the classical two-phase Stefan problem*, Japan J. Ind. Appl. Math. **18** (2001) 161–180.

[19] D. Hilhorst, M. Iida, M. Mimura, H. Ninomiya, *Relative compactness in L^p of solutions of some $2m$ components competition-diffusion system*, Discrete Contin. Dynam. Systems **21** (2008) 233–244.

[20] M.W. Hirsch, *Differential equations and convergence almost everywhere of strongly monotone semiflows*, PAM Technical Report, University of California, Berkeley, 1982 .

[21] Y. Huang, *How do cross-migration models arise?* Mathematical Biosciences **195** (2005) 127–140.

[22] M. Iida, H. Ninomiya, *A reaction-diffusion approximation to a cross-diffusion system*, in: M. Chipot, H. Ninomiya (Eds.), Recent Advances on Elliptic and Parabolic Issues, World Scientific, 2006, pp. 145–164.

[23] M. Iida, M. Mimura, H. Ninomiya, *Diffusion cross-diffusion and competitive interaction*, J. Math. Biol. **53** (4) (2006) 617–641.

[24] R. Ikota, M. Mimura, T. Nakaki, *A numerical methodology for a singular limit problem*, RIMS Kokyuroku **146** (2005) 134–147.

[25] Y. Kan-on, *Stability of singularly perturbed solutions to nonlinear diffusion systems arising in population dynamics*, Hiroshima Math. J. **23** (1993) 509–536.

[26] Y. Kan-on, *Parameter dependence of propagation speed of travelling waves for competition-diffusion equations*, SIAM J. Math. Anal. **26** (1995) 340–363.

[27] Y. Kan-on, M. Mimura, *Segregation structures of competing species mediated by a diffusive predator*, in: Mathematical Topics in Population Biology Morphogenesis and Neurosciences, in: Lecture Notes in Biomath., vol. 71, 1985, pp. 123–133.

[28] P. Kareiva, G. Odell, *Swarms of predators exhibit "preytaxis" if individual predator use area-restricted search*, Amer. Natur. **130** (1987) 233–270.

[29] H. Kawarada, M. Natori, *Numerical solutions of a two-phase Stefan problem*, Theoret. Appl. Mech. **23** (1975) 511–516.

[30] J.B. Keller, P. Sternberg, J. Rubinstein, *Fast reaction, slow diffusion and curve shortening*, SIAM J. Appl. Math. **49** (1989) 116–133.

[31] K. Kishimoto, H.F. Weinberger, *The spatial homogeneity of stable equilibria of some reaction-diffusion system on convex domains*, J. Differential Equations **58** (1985) 15–21.

[32] Y. Lou, W.-M. Ni, *Diffusion vs cross-diffusion: An elliptic approach*, J. Differential Equations **154** (1999) 157–190.

[33] Y. Lou, W.-M. Ni, S. Yotsutani, *On a limiting system in the Lotka-Volterra competition with cross-diffusion*, Discrete Contin. Dynam. Systems **10** (2004) 435–458.

[34] Y. Lou, W.-M. Ni, Y. Wu, *On the global existence of a cross-diffusion system*, Discrete Contin. Dynam. Systems **4** (1998) 193–203.

[35] H. Matano, M. Mimura, *Pattern formation in competition-diffusion systems in nonconvex domains*, Publ. Res. Inst. Math. Sci. Kyoto Univ. **19** (1983) 1049–1079.

[36] M. Mimura, K. Kawasaki, *Spatial segregation in competitive interaction-diffusion equations*, J. Math. Biol. **9** (1980) 49–64.

[37] M. Mimura, Y. Nishiura, A. Tesei, T. Tsujikawa, *Coexistence problem for two competing species models with density-dependent diffusion*, Hiroshima Math. J. **14** (1984) 425–449.

[38] M. Mimura, T. Tsujikawa, *Aggregating pattern dynamics in a chemotaxis-model including growth*, Physica A **230** (1996) 499–543.

[39] M. Mimura, Y. Yamada, S. Yotsutani, *A free boundary problem in ecology*, Japan J. Appl. Math. **2** (1985) 151–186.

[40] M. Mimura, Y. Yamada, S. Yotsutani, *Stability analysis for free boundary problems in ecology*, Hiroshima Math. J. **16** (1986) 477–498.

[41] H. Murakawa, *A regularization of a reaction-diffusion system approximation to the two phase Stefan problem*, Nonlinear Anal. TMA **69** (2008) 3512–3524.

[42] H. Murakawa, *Reaction-diffusion system approximation to degenerate parabolic systems*, Nonlinearity **20** (2007) 2319–2332.

[43] A. Okubo, *Diffusion and Ecological Problems: Mathematical Models*, in: Biomathematics, vol. 10, Springer-Verlag, 1980.

[44] N. Shigesada, K. Kawasaki, E. Teramoto, *Spatial segregation of interacting species*, J. Theoret. Biol. **79** (1979) 83–99.

[45] P. Turchin, *Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals and Plants*, Sinauer, 1998.

[46] A.M. Turing, *The chemical basis of morphogenesis*, Phil. Trans. Roy. Soc. London **B237** (1952) 37–72.

CHAPTER 3

The Mathematics of Nonlinear Optics

Guy Métivier

Université Bordeaux 1, IMB UMR 5251, 351 Cours de la Libération, 33405 Talence Cedex, France
E-mail: guy.metivier@math.u-bordeaux1.fr

Contents

HANDBOOK OF DIFFERENTIAL EQUATIONS
Evolutionary Equations, Volume 5
Edited by C.M. Dafermos and M. Pokorný

DOI: 10.1016/S1874-5717(08)00210-7

Abstract

Modeling in Nonlinear Optics, and also in other fields of Physics and Mechanics, yields interesting and difficult problems due to the presence of several different scales of time, length, energy, etc. These notes are devoted to the introduction of mathematical tools that can be used in the analysis of mutliscale PDE's. We concentrate here on oscillating waves and hyperbolic equations. The main topic is to understand the propagation and the interaction of wave packets, using phase-amplitude descriptions. The main questions are first to find the reduced equations satisfied by the envelops of the fields, and second to rigorously justify them. We first motivate the mathematical analysis by giving various models from optics, including Maxwell–Bloch equations and examples of Maxwell–Euler systems. Then we present a stability analysis of solutions of nonlinear hyperbolic systems, with a particular interest in the case of singular systems where small or large parameters are present. Next, we give the main features concerning the propagations of wave trains, both in the regime of geometric optics and in the regime of diffractive optics. We present the WKB method, the propagation along rays, the diffraction effects transversal to the beam propagations, the modulation of amplitudes. We construct approximate solutions and discuss their stability, rigorously justifying, when possible, the asymptotic expansions. Finally, we discuss the most important nonlinear phenomenon of the theory, that is the wave interaction. After a small digression devoted to general considerations about the mathematical modeling of multi phase oscillations, we apply these notions to introduce important notions such as phase matching, coherence of phases and apply them in various frameworks to the construction of approximate solutions. We also present several methods that have been used for rigorously justifying the multi-phase expansion.

Keywords: Nonlinear optics, hyperbolic systems, stability of solutions, multiscale analysis, oscillations, wave packets, geometric optics, diffractive optics, dispersive optics, wave interaction, phase matching, resonance, coherence of phases, profiles, Maxwell equations

1. Introduction

Nonlinear optics is a very active field in physics, of primary importance and with an extremely wide range of applications. For an introduction to the physical approach, we can refer the reader to classical text books such as [107,10,6,8,109,49,89].

Optics is about the propagation of electromagnetic waves. Nonlinear optics is more about the study of how high intensity light interacts with and propagates through matter. As long as the amplitude of the field is moderate, the linear theory is well adapted, and complicated fields can be described as the superposition of noninteracting simpler solutions (e.g. plane waves). Nonlinear phenomena, such as double refraction or Raman effect, which are examples of wave interaction, have been known for a long time, but the major motivation for a nonlinear theory came from the discovery of the laser which made available highly coherent radiations with extremely high local intensities.

Clearly, *interaction* is a key word in nonlinear optics. Another fundamental feature is that many *different scales* are present: the wavelength, the length and the width of the beam, the duration of the pulse, the intensity of the fields, but also internal scales of the medium such as the frequencies of electronic transitions etc.

Concerning the mathematical set up, the primitive models are Maxwell's equations for the electric and magnetic intensity fields E and H and the electric and magnetic inductions D and B, coupled with constitutive relations between these fields and/or equations which model the interaction of the fields with matter. Different models are presented in Section 2. All these equations fall into the category of *nonlinear hyperbolic systems*. A very important step in modeling is to put these equations in a dimensionless form, by a suitable choice of units, and deciding which phenomena, at which scale, one wants to study (see e.g. [39]). However, even in dimensionless form, the equations may contain one or several small/large parameters. For example, one encounters *singular systems* of the form

$$A_0(u)\partial_t u + \sum_{j=1}^{d} A_j(u)\partial_{x_j} u + \lambda L_0 u = F(u) \tag{1.1}$$

with $\lambda \gg 1$. In addition to the parameters contained in the equation, there are other length scales typical to the solutions under study. In linear and nonlinear optics, one is specially interested in *waves packets*

$$u(t,x) = \mathrm{e}^{\mathrm{i}(k\cdot x-\omega t)}U(t,x) + cc, \tag{1.2}$$

where cc denotes the complex conjugate of the preceding term, k and ω are the central wave number and frequency, respectively, and where the *envelope* $U(t,x)$ has slow variations compared to the rapid oscillations of the exponential:

$$\partial_t^j U \ll \omega^j U, \qquad \partial_x^j U \ll k^j U. \tag{1.3}$$

In dimensionless form, the wave length $\varepsilon := \frac{2\pi}{k}$ is small ($\ll 1$). An important property is that k and ω must be linked by the *dispersion relation* (see Section 5). Interesting phenomena occur when $\varepsilon \approx \lambda^{-1}$. With this scaling there may be resonant interactions between the electromagnetic wave and the medium, accounting for the dependence of optical indices on the frequency and thus for diffraction of light (see e.g. [89]).

More generally, several scales can be present in the envelope U, for instance

$$U = U(t, x, \frac{t}{\sqrt{\varepsilon}}, \frac{x}{\sqrt{\varepsilon}}) \quad \text{(speckles)}, \tag{1.4}$$

$$U = U(t, x, \varepsilon t) \quad \text{(long time propagation)} \tag{1.5}$$

etc. The typical size of the envelope U (the intensity of the beam) is another very important parameter.

Nonlinear systems are the appropriate framework to describe interaction of waves: wave packets with phases $\varphi_j = k_j \cdot x - \omega_j t$ create, by multiplication of the exponentials, the new phase $\varphi = \sum \varphi_j$, which may or may not satisfy the dispersion relation. In the first case, the oscillation with phase φ is propagated (or persists) as time evolves: this is the situation of *phase matching*. In the second case, the oscillation is not propagated and only creates a lower order perturbation of the fields. These heuristic arguments are the basis of the formal derivation of envelope equations that can be found in physics books. It is part of the mathematical analysis to make them rigorous.

In particular, self interaction can create *harmonics* $n\varphi$ of a phase φ. Reducing this to a single small dimensionless parameter, this leads us, for instance, to look for solutions of the form

$$u(y) = \varepsilon^p \mathcal{U}\left(\frac{\varphi}{\varepsilon}, y\right), \qquad \varepsilon^p \mathcal{U}\left(\frac{\varphi}{\varepsilon}, \frac{y}{\sqrt{\varepsilon}}, y\right), \qquad \varepsilon^p \mathcal{U}\left(\frac{\varphi}{\varepsilon}, y, \varepsilon y\right) \tag{1.6}$$

where $y = (t, x)$, $\varphi = (\varphi_1, \dots, \varphi_m)$ and $\mathcal{U}$ is periodic in the first set of variables.

Summing up, we see that the mathematical setup concerns *high frequency* and *multiscale* solutions of a *nonlinear hyperbolic problem*. At this point we merge with many other problems arising in physics or mechanics, in particular in fluid mechanics, where this type of multiscale analysis is also of fundamental importance. For instance, we mention the problems of low Mach number flows, or fast rotating fluids, which raise very similar questions.

These notes are intended to serve as an introduction to the field. We do not even try to give a complete overview of the existing results, that would be impossible within a single book, and probably useless. We will tackle only a few basic problems, with the aim of giving methods and landmarks in the theory. Nor do we give complete proofs, instead we will focus on the key arguments and the main ideas.

Different models arising from optics are presented in Section 2. It is important to understand the variety of problems and applications which can be covered by the theory. These models will serve as examples throughout the exposition. Next, the mathematical analysis will present the following points:

• Basic results of the theory of multi-dimensional symmetric hyperbolic systems are recalled in in Section 3. In particular we state the classical theorem of local existence and local stability of smooth solutions[1]. However, this theorem is of little use when applied directly to the primitive equations. For high frequency solutions, it provides existence and stability in very short intervals of time, because fixed bounds for the derivatives require small amplitudes. Therefore this theorem, which is basic in the theory and provides a useful method, does not apply directly to high frequency and high intensity problems.

One idea to circumvent this difficulty would be to use existence theorems of solutions in energy spaces (the minimal conditions that the physical solutions are expected to satisfy). But there is no such general existence theorem in dimension ≥ 2. This is why existence theorems of energy solutions or weak solutions are important. We give examples of such theorems, noticing that, in these statements, the counterpart of global existence is a much weaker stability.

• In Section 4 we present two methods which can be used to build an existence and stability theory for high frequency/high intensity solutions. The first idea is to factor out the oscillations in the linear case or to consider directly equations for profiles $\mathcal{U}$ (see (1.6)) in the nonlinear case, *introducing the fast variables* (the placeholder for φ/ε for instance) *as independent variables*. The resulting equations are singular, meaning that they have coefficients of order ε^{-1}. The method of Section 3 does not apply in general to such systems, but it can be adapted to classes of such singular equations. which satisfy symmetry and have *good commutation properties*.

Another idea can be used if one knows an approximate solution. It is the goal of the asymptotic methods presented below to construct such approximate solutions: they satisfy the equation up to an error term of size ε^m. The exact solution is sought as the approximate solution plus a corrector, presumably of order ε^m or so. The equation for the corrector has the feature that the coefficients have two components: one is not small but highly oscillatory and known (coming from the approximate solution), the other one is not known but is small (it depends on the corrector). The theory of Section 3 can be adapted in this context to a fairly general extent. However, there is a severe restriction, which is that the order of approximation m must be large enough. In practice, this does not mean that m must necessarily be very large, but it does means that knowledge of the principal term is not sufficient.

• *Geometric optics* is a high frequency approximation of solutions. It fits with the corpuscular description of light, giving a particle-like description of the propagation along *rays*. It concerns solutions of the form $u(y) = \varepsilon^p \mathcal{U}(\frac{\varphi(y)}{\varepsilon}, y)$. The *phase* φ satisfies the *eikonal equation* and the amplitude $\mathcal{U}$ is transported along the rays. In Section 5, we give the elements of this description, using the *WKB method* (for Wentzel, Kramers and Brillouin) of asymptotic (formal) expansions. The scaling ε^p for the amplitude plays an important role in the discussion: if p is large, only linear effects are observed (at least in the leading term). There is a threshold value p_0 where the nonlinear effects are launched in the equation of propagation (typically it becomes nonlinear). Its value depends on the structure of the equation and of the nonlinearity. For a general quasi-linear system, $p_0 = 1$. This is the standard regime of *weakly nonlinear geometric optics*. However, there are special cases where for $p = 1$ the transport equation remains linear; this happens when some interaction

[1] From the point of view of applications and physics, stability is much more significant than existence.

coefficient vanishes. This phenomenon is called *transparency* and for symmetry reasons it occurs rather frequently in applications. An example is the case of waves associated with linearly degenerate modes. In this case it is natural to look for larger solutions with $p < 1$, leading to *strongly nonlinear regimes*. We give two examples where formal large amplitude expansions can be computed.

The WKB construction, when it is possible, leads to approximate solutions, $u^\varepsilon_{\rm app}$. They are functions which satisfy the equation up to an error term that is of order ε^m with m large. The main question is to study the *stability of such approximate solutions*. In the weakly nonlinear regime, the second method evoked in the presentation above in Section 4 applies and WKB solutions are stable, implying that the formal solutions are actually asymptotic expansions of exact solutions. But for strongly nonlinear expansions, the answer is not simple. Indeed, strong instabilities can occur, similar to Rayleigh instabilities in fluid mechanics. These aspects will be briefly discussed.

Another very important phenomenon is *focusing* and *caustics*. It is a linear phenomenon, which leads to concentration and amplification of amplitudes. In a nonlinear context, the large intensities can be over amplified, launching strongly nonlinear phenomena. Some results of this type are presented at the end of the section. However a general analysis of nonlinear caustics is still a wide open problem.

• *Diffractive optics.* This is the usual regime of long time or long distance propagation. Except for very intense and very localized phenomena, the length of propagation of a laser beam is much larger than its width. To analyze the problem, one introduces an additional time scale (and possibly one can also introduce an additional scale in space). This is the *slow time*, typically $T = \varepsilon t$. The propagations are governed by equations in T, so that the description allows for times that are $O(\varepsilon^{-1})$. For such propagations, the classical linear phenomenon which is observed is *diffraction of light* in the direction transversal to the beam, along long distances. The canonical model which describes this phenomenon is a Schrödinger equation, which replaces the transport equation of geometric optics. This is the well known *paraxial approximation*, which also applies to nonlinear equations. This material is presented in Section 6 with the goal of clarifying the nature and the universality of nonlinear Schrödinger equations as fundamental equations in nonlinear optics, which are often presented as the basic equations in physics books.

• *Wave interaction.* This is really where the nonlinear nature of the equations is rich in applications but also of mathematical difficulties. The physical phenomenon which is central here is the resonant interaction of waves. They can be optical waves, for instance a pump wave interacting with scattered and back scattered waves; they can also be optical waves interacting with electronic transitions yielding Raman effects, etc. An important observation is that resonance or phase matching is a *rare* phenomenon : a linear combination of characteristic phases is very likely *not* characteristic. This does not mean that the resonance phenomenon is uninteresting, on the contrary it is of fundamental importance. But this suggests that, when it occurs, it should remain limited. This is more or less correct at the level of formal or BKW expansions, if one retains only the principal terms and neglects all the alleged small residual terms. However, the mathematical analysis is much more delicate: the analysis of all the phases created by the interaction, not only those in the principal term, can be terribly complicated, and possibly beyond the scope of a description by functions of finitely many variables; the focusing effects of these phases have to be taken into account; harmonic generation can cause small divisors problems... It turns out that justification of the formal calculus requires strong assumptions, that we call *coherence*. Fortunately these assumptions are realistic in applications, and thus the theory applies to interesting examples such as the Raman interaction evoked above. Surprisingly, the coherence assumptions are very close to, if not identical with, the commutation requirements introduced in Section 4 for the equations with the fast variables. All these aspects, including examples, are presented in Section 7.

There are many other important questions which are not presented in these notes. We can mention:

– *Space propagation, transmission and boundary value problems*. In the exposition, we have adopted the point of view of describing the evolution in time, solving mainly Cauchy problems. For physicists, the propagation is more often thought of in space: the beam propagates from one point to another, or the beam enters a medium etc. In the geometric optics description, the two formulations are equivalent, as long as they are governed by the same transport equations. However, for the exact equations, this is a completely different point of view, with the main difficulty being that the equations are *not* hyperbolic in space. The correct mathematical approach is to consider transmission problems or boundary value problems. In this framework, new questions are reflection and transmission of waves at a boundary. Another important question is the generation of harmonics or scattered waves at boundaries. In addition, the boundary or transmission conditions may reveal instabilities which may or may not be excited by the incident beam, but which in any case make the mathematical analysis harder. Several references in this direction are [4,104,128,129,21, 25,97,98,39].

– *Short pulses*. We have only considered oscillating signals, that is of the form $e^{i\omega t}a$ with a slowly varying, typically, $\partial_t a \ll \omega a$. For short pulses, and even more for ultra short pulses, which do exist in physics (femtosecond lasers), the duration of the pulse (the support of a) is just a few periods. Therefore, the description with phase and amplitude is no longer adapted. Instead of periodic profiles, this leads us to consider *confined profiles* in time and space, typically in the Schwartz class. On the one hand, the problem is simpler because it is less nonlinear : the durations of interaction are much smaller, so the nonlinear effects require more intense fields. But on the other hand, the mathematical analysis is seriously complicated by passing from a discrete to a continuous Fourier analysis in the fast variables. Several references in this direction are [2,3,17–19,16].

2. Examples of equations arising in nonlinear optics

The propagation of an electro-magnetic field is governed by Maxwell's equation. The nonlinear character of the propagation has several origins: it may come from self-interaction, or from the interaction with the medium in which it propagates. The general Maxwell's equations (in Minkowski space-time) read (see e.g. [49,107,109]):

$$\begin{cases} \partial_t D - \operatorname{curl} H = -j, \\ \partial_t B + \operatorname{curl} E = 0, \\ \operatorname{div} B = 0, \\ \operatorname{div} D = q \end{cases} \tag{2.1}$$

where D is the electric displacement, E the electric field vector, H the magnetic field vector, B the magnetic induction, j the current density and q is the charge density; c is the velocity of light. They also imply the charge conservation law:

$$\partial_t q + \operatorname{div} j = 0. \tag{2.2}$$

We mainly consider the case where there are no free charges and no current flows ($j = 0$ and $q = 0$).

All the physics of interaction of light with matter is contained in the relations between the fields. The *constitutive equations* read

$$D = \varepsilon E, \qquad B = \mu H, \tag{2.3}$$

where ε is the dielectric tensor and μ the tensor of magnetic permeability. When ε and μ are known, this closes the system (2.1). Conversely, (2.3) can be seen as the definition of ε and μ and the links between the fields must be given through additional relations. In the description of interaction of light with matter, one uses the following constitutive relations (see [107,109]):

$$B = \mu H, \qquad D = \varepsilon_0 E + P = \varepsilon E \tag{2.4}$$

where P is called the *polarization* and ε the dielectric constant. In vacuum, $P = 0$. When light propagates in a dielectric medium, the light interacts with the atomic structure, creates dipole moments and induces the polarization P.

In the standard regimes of optics, the magnetic properties of the medium are not prominent and μ can be taken constant equal to $\frac{c^2}{\varepsilon_0}$ with c the speed of light in vacuum and ε_0 the dielectric constant in vacuum. Below we give several models of increasing complexity which can be derived from (2.4), varying the relation between P and E.

• *Equations in vacuum.*

In a vacuum, $\varepsilon = \varepsilon_0$ and $\mu = \frac{1}{\varepsilon_0 c^2}$ are scalar and constant. The constraint equations $\operatorname{div} E = \operatorname{div} B = 0$ are propagated in time and the evolution is governed by the classical *wave equation*

$$\partial_t^2 E - c^2 \Delta E = 0. \tag{2.5}$$

• *Linear instantaneous polarization.*

For small or moderate values of the electric field amplitude, P depends linearly in E. In the simplest case when the medium is isotropic and responds instantaneously to the electric field, P is proportional to E:

$$P = \varepsilon_0 \chi E \tag{2.6}$$

χ is the *electric susceptibility*. In this case, E satisfies the wave equation

$$n^2\partial_t^2 E - c^2\Delta E = 0. \tag{2.7}$$

where $n = \sqrt{1+\chi} \geq 1$ is the *refractive index* of the medium. In this medium, light propagates at the speed $\frac{c}{n} \leq c$.

• *Crystal optics.*

In crystals, the isotropy is broken and D is not proportional to E. The simplest model is obtained by taking the dielectric tensor ε in (2.3) to be a positive definite symmetric matrix, while $\mu = \frac{1}{\varepsilon_0 c^2}$ remains scalar. In this case the system reads:

$$\begin{cases} \partial_t(\varepsilon E) - \operatorname{curl} H = 0, \\ \partial_t(\mu H) + \operatorname{curl} E = 0, \end{cases} \tag{2.8}$$

plus the constraint equations $\operatorname{div}(\varepsilon E) = \operatorname{div} B = 0$, which are again propagated from the initial conditions. That the matrix ε is not proportional to the identity reflects the *anisotropy* of the medium. For instance, for a bi-axial crystal ε has three distinct eigenvalues.

Moreover, in the equations above, ε is constant or depends on the space variable modeling the *homogeneity or inhomogeneity* of the medium.

• *Lorentz model*

Matter does not respond instantaneously to stimulation by light. The delay is captured by writing in place of (2.6)

$$P(x,t) = \varepsilon_0 \int_{-\infty}^{t} \chi(t-t')E(x,t')\mathrm{d}t', \tag{2.9}$$

modeling a linear relation $E \mapsto P$, satisfying the causality principle. On the frequency side, that is after Fourier transform in time, this relation reads

$$\widehat{P}(x,\tau) = \varepsilon_0 \widehat{\chi}(\tau)\widehat{E}(x,\tau). \tag{2.10}$$

It is a simple model where *the electric susceptibility χ depends on the frequency τ.*

In a standard model, due to Lorentz [100], of the linear dispersive behavior of electromagnetic waves, P is given by

$$\partial_t^2 P + \partial_t P/T_1 + \omega^2 P = \gamma E \tag{2.11}$$

with positive constants ω, γ, and T_1 (see also [49], chap. I-31, and II-33). Resolving this equation yields an expression of the form (2.10). In particular

$$\widehat{\chi}(\tau) = \frac{\gamma_0}{\omega^2 - \tau^2 + \mathrm{i}\tau/T_1}, \qquad \gamma_0 = \frac{\gamma}{\varepsilon_0}. \tag{2.12}$$

The physical origin of (2.11) is a model of the electron as bound to the nucleus by a Hooke's law spring force with characteristic frequency ω; T_1 is a damping time and γ a coupling constant.

For non-isotropic crystals, the equation reads

$$\partial_t^2 P + R\partial_t P + \Omega P = \Upsilon E \tag{2.13}$$

where R, Ω and Υ are matrices which are diagonal in the the crystal frame.

• *Phenomenological modeling of nonlinear interaction*

In a first attempt, nonlinear responses of the medium can be described by writing P as a power series in E:

$$P(x,t) = \varepsilon_0 \sum_{k=1}^{\infty} \int \chi^k(t-t_1,\dots,t-t_k) \bigotimes_{j=1}^{k} E(x,t_k)\mathrm{d}t_1\dots\mathrm{d}t_k \tag{2.14}$$

where χ^k is a tensor of appropriate order. The symmetry properties of the susceptibilities χ^k reflect the symmetry properties of the medium. For instance, in a centrosymmetric and isotropic crystal, the quadratic susceptibility χ^2 vanishes.

The first term in the series represents the linear part and often splits P into its linear (and main) part P_L and its nonlinear part P_{NL}.

$$P = P_L + P_{NL}. \tag{2.15}$$

Instantaneous responses correspond to susceptibilities χ^k which are Dirac measures at $t_j = t$. One can also mix delayed and instantaneous responses.

• *Two examples*

In a centrosymmetric, homogeneous and isotropic medium (such as glass or liquid), the first nonlinear term is cubic. A model for P with a *Kerr nonlinearity* is $P = P_L + P_{NL}$ with

$$\partial_t^2 P_L + \partial_t P_L/T_1 + \omega^2 P_L = \gamma E, \tag{2.16}$$

$$P_{NL} = \gamma_{NL}|E|^2 E. \tag{2.17}$$

In a nonisotropic crystal (such as KDP), the nonlinearity is quadratic and model equations for P are

$$\partial_t^2 P_L + R\partial_t P_L + \Omega P_L = \Upsilon E, \tag{2.18}$$

$$P_{NL} = \gamma_{NL}(E_2E_3, E_1E_2, E_1E_2)^t. \tag{2.19}$$

• *The anharmonic model*

To explain nonlinear dispersive phenomena, a simple and natural model is to replace the linear restoring force (2.11) with a nonlinear law (see [6,108])

$$\partial_t^2 P + \partial_t P/T_1 + (\nabla V)(P) = bE. \tag{2.20}$$

For small amplitude solutions, the main nonlinear effect is governed by the Taylor expansion of V at the origin, in presence of symmetries, the first term is cubic, yielding the equation

$$\partial_t^2 P + \partial_t P/T_1 + \omega^2 P - \alpha |P|^2 P = \gamma E. \tag{2.21}$$

• *Maxwell–Bloch equations*

Bloch's equation are widely used in nonlinear optics textbooks as a theoretical background for the description of the interaction between light and matter and the propagation of laser beams in nonlinear media. They link P and the electronic state of the medium, which is described through a simplified quantum model, see e.g. [8,107,10,109]. The formalism of density matrices is convenient to account for statistical averaging due, for instance, to the large number of atoms. The self-adjoint density matrix ρ satisfies

$$\mathrm{i}\varepsilon\partial_t\rho = [\Omega, \rho] - [V(E, B), \rho], \tag{2.22}$$

where Ω is the electronic Hamiltonian in absence of an external field and $V(E, B)$ is the potential induced by the external electromagnetic field. For weak fields, V is expanded into its Taylor's series (see e.g. [109]). In the dipole approximation,

$$V(E, B) = E \cdot \Gamma, \qquad P = \mathrm{tr}(\Gamma\rho) \tag{2.23}$$

where $-\Gamma$ is the dipole moment operator. An important simplification is that only a finite number of eigenstates of Ω are retained. From the physical point of view, they are associated with the electronic levels which are actually in interaction with the electromagnetic field. In this case, ρ is a complex finite dimensional $N \times N$ matrix and Γ is a $N \times N$ matrix with entries in $\mathbb{C}^3$. It is Hermitian symmetric in the sense that $\Gamma_{k,j} = \overline{\Gamma}_{j,k}$ so that $\mathrm{tr}(\Gamma\rho)$ is real. In physics books, the reduction to finite dimensional systems (2.22) comes with the introduction of phenomenological damping terms, which would force the density matrix to relax towards a thermodynamical equilibrium in absence of the external field. For simplicity, we have omitted these damping terms in the equations above. The large ones only contribute to reducing the size of the effective system and the small ones contribute to perturbations which do not alter qualitatively the phenomena. Physics books also introduce "local field corrections" to improve the model and take into account the electromagnetic field created by the electrons. This mainly results in changing the values of several constants, which is of no importance in our discussion.

The parameter ε in front of ∂_t in (2.22) plays a crucial role in the model. The quantities $\omega_{j,k}/\varepsilon := (\omega_j - \omega_k)/\varepsilon$, where the ω_j are the eigenvalues of Ω, have an important physical meaning. They are the characteristic frequencies of the electronic transitions from the level k to the level j and therefore related to the energies of these transitions. The interaction between light and matter is understood as a resonance phenomenon and the possibility of excitation of electrons by the field. This means that the energies of the electronic transitions are comparable to the energy of photons. Thus, if one chooses to normalize $\Omega \approx 1$ as we now assume, ε is comparable to the pulsation of light. The Maxwell–Bloch model described above, is expected to be correct for weak fields and small perturbations of the ground state, in particular below the ionization phenomena.

• *A two levels model* A simplified version of Bloch's equations for a two levels quantum system for the electrons, links the polarization P of the medium and the difference N between the numbers of excited and nonexcited atoms:

$$\varepsilon^2 \partial_t P + \Omega^2 P = \gamma_1 N E, \tag{2.24}$$

$$\partial_t N = -\gamma_2 \partial_t P \cdot E. \tag{2.25}$$

Here, Ω/ε is the frequency associated with the electronic transition between the two levels.

• *Interaction Laser–Plasma*

We give here another example of systems that arise in nonlinear optics. It concerns the propagation of light in a plasma, that is a ionized medium. A classical model for the plasma is a bifluid description for ions and electrons. Then Maxwell equations are coupled to Euler equations for the fluids:

$$\partial_t B + c\,\mathrm{curl}\, \times E = 0, \tag{2.26}$$

$$\partial_t E - c\,\mathrm{curl}\, \times B = 4\pi e\left((n_0 + n_e)v_e - (n_0 + n_i)v_i\right), \tag{2.27}$$

$$(n_0 + n_e)\left(\partial_t v_e + v_e \cdot \nabla v_e\right) = -\frac{\gamma_e T_e}{m_e}\nabla n_e - \frac{e(n_0 + n_e)}{m_e}\left(E + \frac{1}{c} v_e \times B\right), \tag{2.28}$$

$$(n_0 + n_i)\left(\partial_t v_i + v_i \cdot \nabla v_i\right) = -\frac{\gamma_i T_i}{m_i}\nabla n_i + \frac{e(n_0 + n_i)}{m_i}\left(E + \frac{1}{c} v_i \times B\right), \tag{2.29}$$

$$\partial_t n_e + \nabla \cdot ((n_0 + n_e)v_e) = 0, \tag{2.30}$$

$$\partial_t n_i + \nabla \cdot ((n_0 + n_i)v_i) = 0, \tag{2.31}$$

where

∗ E and B are the electric and magnetic field, respectively,

∗ v_e and v_i denote the velocities of electrons and ions, respectively,

∗ n_0 is the mean density of the plasma,

∗ n_e and n_i are the variation of density with respect to the mean density n_0 of electrons and ions, respectively.

Moreover,

∗ c is the velocity of light in the vacuum; e is the elementary electric charge,

∗ m_e and m_i are the electron's and ion's masses, respectively,

∗ T_e and T_i are the electronic and ionic temperatures, respectively and γ_e and γ_i the thermodynamic coefficients.

For a precise description of this kind of model, we refer to classical textbooks such as [37]. One of the main points is that the mass of the electrons is very small compared to the mass of the ions: $m_e \ll m_i$. Since the Lorentz force is the same for the ions and the electrons, the velocity of the ions will be negligible with respect to the velocity of the electrons. The consequence is that one can neglect the contribution of the ions in Eq. (2.27).

3. The framework of hyperbolic systems

The equations above fall into the general framework of *hyperbolic systems*. In this section we point out a few landmarks in this theory, concerning the local stability and existence theory, and some results of global existence.We refer to [35,36,52,53,55,58,94,103,113] for some references to hyperbolic systems.

3.1. *Equations*

The general setting of *quasi-linear first order systems* concerns equations of the form:

$$A_0(a, u)\partial_t u + \sum_{j=1}^{d} A_j(a, u)\partial_{x_j} u = F(a, u) \tag{3.1}$$

where a denotes a set of parameters, which may depend on and include the time-space variables $(t, x) \in \mathbb{R} \times \mathbb{R}^d$; the A_j are $N \times N$ matrices and F is a function with values in $\mathbb{R}^N$; they depend on the variables (a, u) varying in a subdomain of $\mathbb{R}^M \times \mathbb{R}^N$, and we assume that $F(0, 0) = 0$. (Second order equations such as (2.11) or (2.21) are reduced to first order by introducing $Q = \partial_t P$.)

An important case is the case of *balance laws*

$$\partial_t f_0(u) + \sum_{j=1}^{d} \partial_{x_j} f_j(u) = F(u) \tag{3.2}$$

or *conservation laws* if $F = 0$. For smooth enough solutions, the chain rule can be applied and this system is equivalent to

$$A_0(u)\partial_t u + \sum_{j=1}^{d} A_j(u)\partial_{x_j} u = F(u) \tag{3.3}$$

with $A_j(u) = \nabla_u f_j(u)$. Examples of quasi-linear systems are Maxwell's equations with the Kerr nonlinearity (2.17) or Euler–Maxwell equations.

The system is *semi-linear* when the A_j do not depend on u:

$$A_0(a)\partial_t u + \sum_{j=1}^{d} A_j(a)\partial_{x_j} u = F(b, u). \tag{3.4}$$

Examples are the anharmonic model (2.21) or Maxwell–Bloch equations.

The system is *linear* when the A_j do not depend on u and F is affine in u, i.e. of the form $F(b,u) = f + E(b)u$. This is the case of systems such as (2.7), (2.8) or the Lorentz model.

Consider a solution u_0 and the equation for small variations $u = u_0 + \varepsilon v$. Expanding as a power series in ε yields at first order *the linearized equations*:

$$A_0(a,u_0)\partial_t v + \sum_{j=1}^{d} A_j(a,u_0)\partial_{x_j} v + E(t,x)v = 0 \tag{3.5}$$

where

$$E(t,x)v = (v \cdot \nabla_u A_0)\partial_t u_0 + \sum_{j=1}^{d} (v \cdot \nabla_u A_j)\partial_{x_j} u_0 - v \cdot \nabla_u F$$

and the gradients $\nabla_u A_j$ and $\nabla_u F$ are evaluated at $(a, u_0(t,x))$.

In particular, the linearized equations from (3.2) or (3.3) near a constant solution $u_0(t,x) = \underline{u}$ are the *constant coefficients equations*

$$A_0(\underline{u})\partial_t u + \sum_{j=1}^{d} A_j(\underline{u})\partial_{x_j} u = F'(\underline{u})u. \tag{3.6}$$

The example of Maxwell's equations.

Consider Maxwell's equation with no charge and no current:

$$\partial_t D - \operatorname{curl} H = -j, \qquad \partial_t B + \operatorname{curl} E = 0, \tag{3.7}$$
$$\operatorname{div} B = 0, \qquad \operatorname{div} D = 0, \tag{3.8}$$

together with constitutive relations between the fields as explained in Section 2. This system is not immediately of the form (3.1): it is *overdetermined* as it involves more equations than unknowns and as there is no ∂_t in the second set of equations. However, it satisfies *compatibility* conditions[2]: the first two equations (3.7) imply that

$$\partial_t \operatorname{div} B = 0, \qquad \partial_t \operatorname{div} D = 0, \tag{3.9}$$

so that the *constraint conditions* (3.8) are satisfied for all time by solutions of (3.7) if and only if they are satisfied at time $t = 0$. As a consequence, one studies the evolution system (3.7) alone, which is of the form (3.1), and considers the constraints (3.8) as conditions on the initial data. With this modification, the framework of hyperbolic equations is well adapted to the various models involving Maxwell's equations.

[2]This is a special case of a much more general phenomenon for fields equations, where the equations are linked through Bianchi's identities.

For instance, the Lorentz model is the linearization of both the anharmonic model and of the Kerr Model at $\underline{E} = 0$, $\underline{P} = 0$.

3.2. *The dispersion relation & polarization conditions*

Consider a *linear constant coefficient* system such as (3.6):

$$Lu := A_0 \partial_t u + \sum_{j=1}^{d} A_j \partial_{x_j} u + Eu = f. \tag{3.10}$$

Particular solutions of the homogeneous equation $Lu = 0$ are *plane waves*:

$$u(t, x) = \mathrm{e}^{\mathrm{i}t\tau + \mathrm{i}x\cdot\xi} a \tag{3.11}$$

where (τ, ξ) satisfy the *dispersion relation*:

$$\det \left(\mathrm{i}\tau A_0 + \sum_{j=1}^{d} \mathrm{i}\xi_j A_j + E \right) = 0 \tag{3.12}$$

and the constant vector a satisfies the *polarization condition*

$$a \in \ker \left(\mathrm{i}\tau A_0 + \sum_{j=1}^{d} \mathrm{i}\xi_j A_j + E \right). \tag{3.13}$$

The matrix $\mathrm{i}\tau A_0 + \sum_{j=1}^{d} \mathrm{i}\xi_j A_j + E$ is called *the symbol* of L.

In many applications, the coefficients A_j and E are real and one is interested in real functions. In this case (3.11) is to be replaced by $u = \mathrm{Re}(\mathrm{e}^{\mathrm{i}t\tau + \mathrm{i}x\cdot\xi} a)$.

When A_0 is invertible, the Eq. (3.12) means that $-\tau$ is an eigenvalue of $\sum \xi_j A_0^{-1} A_j - \mathrm{i} A_0^{-1} E$ and the polarization condition (3.13) means that a is an eigenvector.

We now illustrate these notions with four examples involving Maxwell's equations. Following the general strategy explained above, we forget the divergence equations (3.8). However, this has the effect of adding extra and non-physical eigenvalues $\tau = 0$, with eigenspace $\{B \in \mathbb{R}\xi, D \in \mathbb{R}\xi\}$ incompatible with the divergence relations, which for plane waves require that $\xi \cdot B = \xi \cdot D = 0$. Therefore, these extra eigenvalues must be discarded in the physical interpretation of the problem.

• For the *Lorentz model*, the dispersion relation reads

$$\tau^2(\delta - \gamma_0) \left(\tau^2(\delta - \gamma_0) - c^2 \delta |\xi|^2 \right)^2 = 0, \qquad \delta = \tau^2 - \mathrm{i}\tau/T_1 - \omega^2. \tag{3.14}$$

The root $\tau = 0$ is non-physical as explained above. The roots of $\delta - \gamma_0 = 0$ (that is $\tau = \pm\sqrt{\omega^2 + \gamma}$ in the case $\frac{1}{T_1} = 0$) do not correspond to optical waves, since the corresponding waves propagate at speed 0 (see Section 5). The optical plane waves are associated with roots of the third factor. They satisfy

$$c^2|\xi|^2 = \tau^2 \left(1 + \widehat{\chi}(\tau)\right), \qquad \widehat{\chi}(\tau) = \frac{\gamma_0}{\omega^2 + i\tau/T_1 - \tau^2}. \tag{3.15}$$

For $\xi \neq 0$, they have multiplicity two and the polarization conditions are

$$E \in \xi^\perp, \qquad P = \varepsilon_0 \widehat{\chi}(\tau) E, \qquad B = -\frac{\xi}{\tau} \times E. \tag{3.16}$$

• Consider the *two level Maxwell–Bloch equations*. The linearized equations around $\underline{E} = \underline{B} = \underline{P} = 0$ and $\underline{N} = N_0 > 0$ read (in suitable units)

$$\begin{aligned} &\partial_t B + \mathrm{curl} E = 0, \qquad \partial_t E - \mathrm{curl} B = -\partial_t P, \\ &\varepsilon \partial_t^2 P + \Omega^2 P = \gamma_1 N_0 E, \qquad \partial_t N = 0. \end{aligned} \tag{3.17}$$

This is the Lorentz model with coupling constant $\gamma = \gamma_1 N_0$, augmented by the equation $\partial_t N = 0$. Thus we are back to the previous example.

• For *crystal optics*, in units where $c = 1$, the plane wave equations reads

$$\begin{cases} \tau E - \varepsilon^{-1}(\xi \times B) = 0, \\ \tau B + \xi \times E = 0. \end{cases} \tag{3.18}$$

In coordinates where ε is diagonal with diagonal entries $\alpha_1 > \alpha_2 > \alpha_3$, the dispersion relation read

$$\tau^2 \left(\tau^4 - \Psi(\xi)\tau^2 + |\xi|^2 \Phi(\xi)\right) \tag{3.19}$$

with

$$\begin{cases} \Psi(\xi) = (\alpha_1 + \alpha_2)\xi_3^2 + (\alpha_2 + \alpha_3)\xi_1^2 + (\alpha_3 + \alpha_1)\xi_2^2, \\ \Phi(\xi) = \alpha_1\alpha_2\xi_3^2 + \alpha_2\alpha_3\xi_1^2 + \alpha_3\alpha_1\xi_2^2. \end{cases}$$

For $\xi \neq 0$, $\tau = 0$ is again a double eigenvalue. The non-vanishing eigenvalues are solutions of a second order equation in τ^2, of which the discriminant is

$$\Psi^2(\xi) - 4|\xi|^2 \Phi(\xi) = P^2 + Q^2$$

with

$$P = (\alpha_1 - \alpha_2)\xi_3^2 + (\alpha_3 - \alpha_2)\xi_1^2 + (\alpha_3 - \alpha_1)\xi_2^2,$$

$$Q = 2(\alpha_1 - \alpha_2)^{\frac{1}{2}}(\alpha_1 - \alpha_3)^{\frac{1}{2}}\xi_3\xi_2.$$

For a bi-axial crystal, ε has three distinct eigenvalues. For general frequency ξ, $P^2 + Q^2 \neq 0$ and there are four simple eigenvalues, $\pm\frac{1}{2}\left(\Psi \pm \left(P^2 + Q^2\right)^{\frac{1}{2}}\right)^{\frac{1}{2}}$. There are double roots exactly when $P^2 + Q^2 = 0$, that is when

$$\xi_2 = 0, \qquad \alpha_1\xi_3^2 + \alpha_3\xi_1^2 = \alpha_2(\xi_1^2 + \xi_3^2) = \tau^2. \tag{3.20}$$

This is an example where the multiplicities of the eigenvalues change with ξ.

• Consider the *linearized Maxwell–Bloch equations* around $\underline{E} = \underline{B} = 0$ and $\underline{\rho}$ equal to the fundamental state, i.e. the eigenprojector associated with the smallest eigenvalue ω_1 of Ω. In appropriate units, they read

$$\begin{cases} \partial_t B + \operatorname{curl} E = 0, \\ \partial_t E - \operatorname{curl} B - \mathrm{i}\,\mathrm{tr}\,(\Gamma[\Omega, \rho]) + \mathrm{i}\,\mathrm{tr}\,(\Gamma(E \cdot G)) = 0, \\ \partial_t \rho + \mathrm{i}[\Omega, \rho] - \mathrm{i}E \cdot G = 0, \end{cases} \tag{3.21}$$

with $G := [\Gamma, \underline{\rho}]$. The general expression of the dispersion relation is not simple, one reason for that is the lack of isotropy of the general model shown above. To simplify, we assume that the fundamental state is simple, that Ω is diagonal with entries α_j and we denote by ω_j the *distinct* eigenvalues of Ω, with $\omega_1 = \alpha_1$. Then G has the form

$$G = \begin{pmatrix} 0 & -\Gamma_{1,2} & \dots \\ \Gamma_{2,1} & 0 & \dots \\ \vdots & \vdots & 0 \end{pmatrix}.$$

Assume that the model satisfies the following isotropy condition: for all $\omega_j > \omega_1$:

$$\sum_{\{k:\alpha_k = \omega_j\}} (E \cdot \Gamma_{1,k})\Gamma_{k,1} = \gamma_j E \tag{3.22}$$

with $\gamma_j \in \mathbb{C}$. Then the optical frequencies satisfy

$$|\xi|^2 = \tau^2(1 + \widehat{\chi}(\tau)), \qquad \widehat{\chi}(\tau) = \sum 2\frac{\mathrm{Re}\gamma_j(\omega_j - \omega_1) + \mathrm{i}\tau\mathrm{Im}\gamma_j}{(\omega_j - \omega_1)^2 - \tau^2} \tag{3.23}$$

and the associated polarization conditions are

$$E \in \xi^{\perp}, \qquad B = -\frac{\xi}{\tau} \times E, \qquad \rho_{1,k} = \overline{\rho}_{k,1} = \frac{E \cdot \Gamma_{1,k}}{\alpha_k - \alpha_1 - \tau} \tag{3.24}$$

with the other entries $\rho_{j,k}$ equal to 0.

3.3. *Existence and stability*

The equations presented in Section 2 fall into the category of *symmetric hyperbolic* systems. More precisely they satisfy the following condition:

DEFINITION 3.1 (*Symmetry*). A system (3.1) is said to be symmetric hyperbolic in the sense of Friedriechs, if there exists a matrix $S(a, u)$ such that

- it is a C^∞ function of its arguments;
- for all j, a and u, the matrices $S(a, u)A_j(a, u)$ are self-adjoint and, in addition, $S(a, u)A_0(a, u)$ is positive definite.

The Cauchy problem consists of solving the equation (3.1) together with the initial condition

$$u_{|t=0} = h. \tag{3.25}$$

The first basic result of the theory is the local existence of smooth solutions:

THEOREM 3.2 (*Local Existence*). *Suppose that the system (3.1) is symmetric hyperbolic. Then for $s > \frac{d}{2} + 1$, $h \in H^s(\mathbb{R}^d)$ and $a \in C^0([0; T]; H^s(\mathbb{R}^d))$ such that $\partial_t a \in C^0([0; T]; H^{s-1}(\mathbb{R}^d))$, there is $T' > 0$, $T' \leq T$, which depends only on the norms of a, $\partial_t a$ and h, such that the Cauchy problem has a unique solution $u \in C^0([0; T']; H^s(\mathbb{R}^d))$.*

In the semi-linear case, that is when the matrices A_j do not depend on u, the limiting lower value for the local existence is $s > \frac{d}{2}$:

THEOREM 3.3. *Consider the semi-linear system (3.4) assumed to be symmetric hyperbolic. Suppose that $a \in C^0([0; T]; H^\sigma(\mathbb{R}^d))$ is such that $\partial_t a \in C^0([0; T]; H^{\sigma-1}(\mathbb{R}^d))$ where $\sigma > \frac{d}{2} + 1$. Then, for $\frac{d}{2} < s \leq \sigma$, $h \in H^s(\mathbb{R}^d)$ and $b \in C^0([0; T]; H^s(\mathbb{R}^d))$, there is $T' > 0$, $T' \leq T$ such that the Cauchy problem with initial data h has a unique solution $u \in C^0([0; T']; H^s(\mathbb{R}^d))$.*

As it is important for understanding the remaining part of these notes we will give the main steps in the proof of this important result. The analysis of linear symmetric hyperbolic problems goes back to [50,51]. For the nonlinear version we refer to [106,103,66,123].

PROOF (*Scheme of the proof*). Solutions can be constructed through an iterative scheme

$$\begin{cases} A_0(a, u_n)\partial_t u_{n+1} + \displaystyle\sum_{j=1}^{d} A_j(a, u_n)\partial_{x_j} u_{n+1} = F(a, u_n), \\ u_{n+1|t=0} = h. \end{cases} \tag{3.26}$$

There are four steps:

1 - [Definition of the scheme.] Prove that if $u_n \in C^0([0; T]; H^s(\mathbb{R}^d))$ and $\partial_t u_n \in C^0([0; T]; H^{s-1}(\mathbb{R}^d))$, the system has a solution u_{n+1} with the same smoothness;

2 - [Boundedness in high norm.] Prove that there is $T' > 0$ such that the sequence is bounded in $C^0([0; T']; H^s(\mathbb{R}^d))$ and $C^1([0; T']; H^{s-1}(\mathbb{R}^d))$.

3 - [Convergence in low norm.] Prove that the sequence converges in $C^0([0; T']; L^2(\mathbb{R}^d))$. Together with the uniform bounds, this implies that the convergence holds in $C^0([0; T']; H^{s'}(\mathbb{R}^d))$ and in $C^1([0; T']; H^{s'-1}(\mathbb{R}^d))$ for all $s' < s$. Since $s > \frac{d}{2} + 1$, the convergence holds in $C^1([0, T'] \times \mathbb{R}^d)$ and the limit u is a solution of the Cauchy problem (3.1) (3.25). The convergence also holds in $C^0([0; T']; H_w^{s'}(\mathbb{R}^d))$ where H_w^s denotes the space H^s equipped with the weak topology and $u \in L^\infty([0, T'], H^s(\mathbb{R}^d)) \cap C^0([0; T']; H_w^{s'}(\mathbb{R}^d))$.

4 - [Strong continuity.] Use the equation to prove that u is actually continuous in time with values in $H^s(\mathbb{R}^d)$ equipped with the strong topology.

This analysis relies on the study of the linear problems

$$\begin{cases} L(a, \partial)u := A_0(a)\partial_t u + \sum_{j=1}^{d} A_j(a)\partial_{x_j} u = f, \\ u_{|t=0} = h. \end{cases} \tag{3.27}$$

The main step is to prove *a priori estimates* for the solutions of such systems.

PROPOSITION 3.4. *If the system is symmetric hyperbolic, then for u smooth enough*

$$\begin{aligned} \|u(t)\|_{H^s} \le C_0 \mathrm{e}^{(K_0+K_s)t} \, \|u(0)\|_{H^s} \\ + C_0 \int_t^{t'} \mathrm{e}^{(K_0+K_s)(t-t')} \, \|L(a, \partial)u\|_{H^s} \, \mathrm{d}t' \end{aligned} \tag{3.28}$$

where C_0 [resp. K_0] [resp. K_s] depends only on the L^∞ norm [resp. $W^{1,\infty}$ norm] [resp. $L^\infty(H^s)$ norm] of a on $[0, T] \times \mathbb{R}^d$.

They are used first to prove the existence and uniqueness of solutions and next to control the solutions. In particular, they serve to prove points 1 and 2 of the scheme above. The convergence in low norm is also a consequence of the energy estimates in L^2 ($s = 0$) applied to the differences $u_{n+1} - u_n$. The additional smoothness consists in proving that if $a \in L^\infty([0, T']; H^s)$ and $\partial_t a \in L^\infty([0, T']; H^{s-1})$, then the solution u actually belongs to $C^0([0, T']; H^s)$. □

NOTES ON THE PROOF OF PROPOSITION 3.4. When $s = 0$, the estimate (with $K_s = 0$) follows easily by multiplying the equation by $S(a)u$ and integration by parts, using the symmetry properties of SA_j and the positivity of SA_0.

When s is a positive integer, $s > \frac{d}{2} + 1$, the estimates of the derivatives are deduced from the L^2 estimates writing

$$L(a, \partial)\partial_x^\alpha u = A_0 \partial_x^\alpha \left(A_0^{-1} L(a, \partial)u \right) - A_0 [\partial_x^\alpha, A_0^{-1} L(a, \partial)]u \tag{3.29}$$

and commutation estimates for $|\alpha| \le s$:

$$\|[\partial_x^\alpha, A_0^{-1} L(a, \partial)]u(t)\|_{L^2} \le K_s \, \|u(t)\|_{H^s} \tag{3.30}$$

where K_s depends only on the H^s norm of $a(t)$.

The bound (3.30) follows from two classical nonlinear estimates which are recalled in the lemma below.

LEMMA 3.5. *For $\sigma > \frac{d}{2}$,*

(i) *$H^\sigma(\mathbb{R}^d)$ is a Banach algebra embedded in $L^\infty(\mathbb{R}^d)$.*
(ii) *For $u \in H^{\sigma-l}(\mathbb{R}^d)$ and $v \in H^{\sigma-m}(\mathbb{R}^d)$ with $l \geq 0$, $m \geq 0$ and $l + m \leq \sigma$, the product uv belongs to $H^{\sigma-l-m}(\mathbb{R}^d) \subset L^2(\mathbb{R}^d)$.*
(iii) *For $s > \frac{d}{2}$ and F a smooth function such that $F(0) = 0$, the mapping $u \mapsto F(u)$ is continuous from $H^s(\mathbb{R}^d)$ into itself and maps bounded sets to bounded sets.*

Indeed, the commutator $[\partial_x^\alpha, A_0^{-1} L(a, \partial)]u$ is a linear combination of terms

$$\partial_x^\gamma \partial_{x_k} A_j(a) \partial_x^\beta \partial_{x_j} u, \qquad |\beta| + |\gamma| \leq |\alpha| - 1. \tag{3.31}$$

Noticing that $A_j(a(t)) \in H^s(\mathbb{R}^d)$, and applying (ii) of the lemma with $\sigma = s - 1$, yields the estimate (3.30). □

REMARK 3.6. The use of L^2 based Sobolev spaces is well adapted to the framework of symmetric systems, but it is also dictated by the consideration of multidimensional problems (see [112]).

Besides the existence statement, which is interesting from a mathematical point of view, the proof of Theorem 3.2 contains an important *stability* result:

THEOREM 3.7 (*Local stability*). *Under assumptions of Theorem 3.2, if $u_0 \in C^0([0, T], H^s(\mathbb{R}^d))$ is a solution of (3.1), then there exists $\delta > 0$ such that for all initial data h such that $\|h - u_0(0)\|_{H^s} \leq \delta$, the Cauchy problem with initial data h has a unique solution $u \in C^0([0, T], H^s(\mathbb{R}^d))$. Moreover, for all $t \in [0, T]$, the mapping $h \mapsto u(t)$ defined in this way, is Lipschitz continuous in the H^{s-1} norm and continuous in the $H^{s'}$ norm, for all $s' < s$, uniformly in t.*

3.4. *Continuation of solutions*

Uniqueness in Theorem 3.2 allows one to define *the maximal interval of existence of a smooth solution*: under the assumptions of this theorem, let T^* be the supremum of $T' \in [0, T]$ such that the Cauchy problem has a solution in $C^0([0, T']; H^s(\mathbb{R}^d))$. By the uniqueness, this defines a (unique) solution $u \in C^0([0, T^*[; H^s(\mathbb{R}^d))$. With this definition, Theorem 3.2 immediately implies the following.

LEMMA 3.8. *It $T^* < T$, then*

$$\sup_{t \in [0, T^*[} \|u(t)\|_{H^s} = +\infty.$$

We mention here that a more precise result exists (see e.g. [103]):

THEOREM 3.9. *It $T^* < T$, then*

$$\sup_{t\in[0,T^*[} \|u(t)\|_{L^\infty} + \|\nabla_x u(t)\|_{L^\infty} = +\infty.$$

We also deduce from Theorem 3.2 two continuation arguments based on *a priori* estimates:

LEMMA 3.10. *Suppose that there is a constant C such that if $T' \in]0, T]$ and $u \in C^0([0, T']; H^s)$ is solution of the Cauchy problem for (3.1) with initial data $h \in H^s$, then it satisfies for $t \in [0, T']$:*

$$\|u(t)\|_{H^s} \leq C. \tag{3.32}$$

Then, the maximal solution corresponding to the initial data h is defined on $[0, T]$ and satisfies (3.32) on $[0, T]$.

LEMMA 3.11. *Suppose that there is are constants C and $C' > C \geq \|h\|_{H^s}$ such that if $u \in C^0([0, T']; H^s)$ is solution of the Cauchy problem for (3.1) with initial data $h \in H^s$ the following implication holds:*

$$\sup_{t\in[0,T']} \|u(t)\|_{H^s} \leq C' \Rightarrow \sup_{t\in[0,T']} \|u(t)\|_{H^s} \leq C. \tag{3.33}$$

Then, the maximal solution is defined on $[0, T]$ and satisfies (3.32) on $[0, T]$.

3.5. *Global existence*

As mentioned in the introduction, the general theorem of local existence is of little use for high frequency initial data, since the time of existence depends on high regularity norms and thus may be very small. In the theory of hyperbolic equations, there is a huge literature on global existence and stability results. We do not mention here the results which concern *small data* (see e.g. [66] and the references therein), since the smallness is again measured in high order Sobolev spaces and thus is difficult to apply to high frequency solutions.

There is another class of classical global existence theorems of *weak or energy* solutions for hyperbolic *maximal dissipative* equations, which use only the conservation or dissipation of energy and a weak compactness argument (see e.g. [99] and Section 7.6.4 below.)

In this section, we illustrate with an example another approach which is better adapted to our context. Indeed, for most of the physical examples, there are conserved (or dissipated) quantities, such as energies. These provide *a priori* estimates that are valid for all time, independently of the size of the data. The problem is to use this particular additional information to improve the general analysis and eventually arrive at global existence.

The case of two levels Maxwell–Bloch equations has been studied in [42]. Their results have been extended to general Maxwell–Bloch equations (2.22) in [46] and to the anharmonic model (2.20) in [77] (see also [83] for Maxwell's equations in a ferromagnetic medium). In the remaining part of this section, we present the example of two level

Maxwell–Bloch equations. Recall the equations from Section 2:

$$\begin{cases} \partial_t B + \operatorname{curl} E = 0, \\ \partial_t (E + P) - \operatorname{curl} B = 0, \\ \partial_t P + \Omega^2 P = \gamma_1 N E, \\ \partial_t N = -\gamma_2 \partial_t P \cdot E, \end{cases} \tag{3.34}$$

together with the constraints

$$\operatorname{div}(E + P) = \operatorname{div} B = 0. \tag{3.35}$$

Recall that N is the difference between the number of electrons in the excited state and in the ground state per unit of volume. N_0 is the equilibrium value of N. This system can be written as a first order semi-linear symmetric hyperbolic system for

$$U = (B, E, P, \partial_t P, N - N_0). \tag{3.36}$$

Since the system is semi-linear, with matrices A_j that are constant, the local existence theorem proves that the Cauchy problem is locally well-posed in $H^s(\mathbb{R}^3)$ for $s > \frac{3}{2}$, see Theorem 3.3. The special form of the system implies that the maximal time of existence is $T^* = +\infty$:

THEOREM 3.12. *If $s \geq 2$ and the initial data $U(0) \in H^s(\mathbb{R}^3)$ satisfies (3.35), then the Cauchy problem for (3.34) has a unique solution $U \in C^0([0, +\infty[; H^s(\mathbb{R}^3))$, which satisfies (3.35) for all time.*

NOTES ON THE PROOF (*see [42]*). The total energy

$$\mathcal{E} = N_0 \|B\|_{L^2}^2 + N_0 \|E\|_{L^2}^2 + \frac{\Omega^2}{\gamma_1} \|P\|_{L^2}^2 + \frac{1}{\gamma_1} \|\partial_t P\|_{L^2}^2 + \frac{1}{\gamma_2} \|N - N_0\|_{L^2}^2$$

is conserved, proving that U remains bounded in L^2 for all time.

There is also a pointwise conservation of

$$\frac{1}{\gamma_1}|\partial_t P|^2 + \frac{\Omega^2}{\gamma_1}|P|^2 + \frac{1}{\gamma_2}|N|^2$$

proving that P, $\partial_t P$ and N remain bounded in L^∞ for all time.

The H^1 estimates are obtained by differentiating the equations with respect to x:

$$\begin{cases} \partial_t \partial B + \operatorname{curl} \partial E = 0, \\ \partial_t \partial E - \operatorname{curl} \partial B = -\partial_t \partial P, \\ \partial_t \partial P + \Omega^2 \partial P = \gamma_1 \partial(N E), \\ \partial_t \partial N = -\gamma_2 \partial(\partial_t P \cdot E). \end{cases}$$

Then

$$\mathcal{E}_1 = \|\partial B\|_{L^2}^2 + \|\partial E\|_{L^2}^2 + \frac{\Omega^2}{\gamma_1} \|\partial P\|_{L^2}^2 + \frac{1}{\gamma_1} \|\partial_t \partial P\|_{L^2}^2 + \frac{1}{\gamma_2} \|\partial N\|_{L^2}^2$$

satisfies

$$\partial_t \mathcal{E}_1 = 2 \int \Phi \mathrm{d}x$$

with

$$\begin{aligned} \Phi &= -(\partial Q)\partial E + \partial(NE)\partial Q - \partial(QE)\partial N \\ &= -(\partial Q)\partial E + N\partial E \partial Q - \partial N Q \partial E, \end{aligned}$$

with $Q = \partial_t P$. Thus, using the known L^∞ bounds for N and Q, implies that

$$\partial_e \mathcal{E}_1 \leq C\mathcal{E}_1$$

implying that $\mathcal{E}_1(t) \leq \mathrm{e}^{Ct}\mathcal{E}_1(0)$ for all time.

The estimate of the second derivatives is more subtle, but follows the same ideas: use the known L^2, L^∞ and H^1 bounds to obtain H^2 estimates valid for all time. For the details we refer to [42]. □

Using the *a priori* H^1 bounds, it is not difficult to prove the global existence of global H^1 solutions (see [42]):

THEOREM 3.13. *For arbitrary* $U(0) \in H^1(\mathbb{R}^3)$ *satisfying* (3.35) *and such that* $(P(0), \partial_t P(0), N(0)) \in L^\infty(\mathbb{R}^3)$*, there is a unique global solution* U *such that for all* $T > 0$, $U \in L^\infty([0, T]; H^1(\mathbb{R}^3))$ *and* $(P, \partial_t P, N) \in L^\infty([0, +\infty[\times\mathbb{R}^3)$.

3.6. *Local results*

A fundamental property of hyperbolic systems is that they reproduce the physical idea that *waves propagate at a finite speed*. Consider solutions of a symmetric hyperbolic linear equation

$$L(a, \partial)u := A_0(a)\partial_t u + \sum_{j=1}^{d} A_j(a)\partial_{x_j} u + E(a)u = f \tag{3.37}$$

on domains of the form:

$$\Omega = \{(t, x) : t \geq 0, |x| + t\lambda_* \leq R\}. \tag{3.38}$$

Let $\omega = \{x : |x| \leq R\}$. One has the following result.

PROPOSITION 3.14 (*Local uniqueness*). *There is a real valued function* $\lambda_*(M)$*, which depends only the matrices* A_j *in the principal part, such that if* $\lambda_* \geq \lambda_*(M)$*, a is Lipschitz* *continuous on* $\overline{\Omega}$ *with* $|a(t,x)| \le M$ *on* Ω, $u \in H^1(\Omega)$ *satisfies* (3.37) *on* Ω *with* $f = 0$ *and* $u_{|t=0} = 0$ *on* ω, *then* $u = 0$ *on* Ω.

PROOF. By Green's formula

$$\begin{aligned} 0 &= 2\mathrm{Re} \int_{\Omega} \mathrm{e}^{-\gamma t} Lu \cdot \overline{u}\,\mathrm{d}t\mathrm{d}x \\ &= \int_{\Omega} \mathrm{e}^{-\gamma t} (\gamma A_0 u - Ku) \cdot \overline{u}\,\mathrm{d}t\mathrm{d}x \\ &\quad + \int_{\Sigma} \mathrm{e}^{-\gamma t} L(a,\nu) u \cdot u\,\mathrm{d}\Sigma \end{aligned}$$

where $K = \partial_t A_0(a) + \sum \partial_{x_j} A_j(a)$, $\Sigma = \{\lambda_* t + |x| = R\}$ and $L(a,\nu) = \sum \nu_j A_j$ is the value of L in the direction $\nu = (\nu_0, \dots, \nu_d)$, which is the exterior normal to Ω. Because ν is proportional to $(\lambda_*, x_1|x|, \dots, x_d/|x|)$, if λ_* is large enough, the matrix $L(a,\nu)$ is nonnegative. More precisely, this condition is satisfied if

$$\lambda_* \ge \lambda_*(M) = \sup_{|a|\le M} \sup_{|\xi|=1} \sup_{p} |\lambda_p(a,\xi)| \tag{3.39}$$

where the $\lambda_p(a,\xi)$ denote the eigenvalues of $A_0(a)^{-1} \sum_{j=1}^d \xi_j A_j(a)$.

If γ is large enough, the matrix $\gamma A_0 - K$ is positive definite, and the energy identity above implies that $u = 0$ on Ω. □

This result implies that the solution u of (3.37) is uniquely determined in Ω by the values of the source term f on Ω and the values of the initial data on ω. One says that Ω *is contained in the domain of determinacy of* ω.

The proposition can be improved, giving the optimal domain of determinacy Ω associated to an initial domain ω, see [93,96,84] and the references therein.

On domains Ω, one uses the following spaces:

DEFINITION 3.15. We say that u defined on Ω is continuous in time with values in L^2 if its extension by 0 outside Ω belongs to $C^0([0,T_0]; L^2(\mathbb{R}^d))$; for $s \in \mathbb{N}$, we say that u is continuous with values in H^s if the derivatives $\partial_x^\alpha u$ for $|\alpha| \le s$ are continuous in time with values in L^2. We denote these spaces by $C^0H^s(\Omega)$.

Proposition 3.14 extends to semi-linear equations, as the domain Ω does not depend on the source term $f(u)$. The energy estimates can be localized on Ω, using integration by parts on Ω, and Theorem 3.3 can be extended as follows:

THEOREM 3.16. *Consider the semi-linear system* (3.4) *assumed to be symmetric hyperbolic. Suppose that* $a \in C^0H^\sigma(\Omega)$ *is such that* $\partial_t a \in C^0H^{\sigma-1}(\Omega)$ *where* $\sigma > \frac{d}{2} + 1$ *and* $\|a\|_{L^\infty(\Omega)} \le M$. *Then, for* $\frac{d}{2} < s \le \sigma$, $h \in H^s(\omega)$ *and* $b \in C^0H^s(\Omega)$, *there exists* $T > 0$, *such that the Cauchy problem with initial data* h *has a unique solution* $u \in C^0H^s(\Omega \cap \{t \le T\})$.

For quasi-linear systems (3.1), the situation is more intricate since then the eigenvalues depend on the solution itself, so that $\lambda_*(M)$ in (3.39) must be replaced by a function

$\lambda_*(M, M')$ which dominates the eigenvalues of $A_0(a,u)^{-1} \sum \xi_j A_j(a,u)$ when $|a| \le M$ and $|u| \le M'$. Note that $\lambda_*(M, M')$ is a continuous increasing function of M', so that if $\lambda_* > \lambda_*(M, M')$ then $\lambda_* \ge \lambda_*(M, M'')$ for a $M'' > M'$.

THEOREM 3.17. *Suppose that the system (3.1) is symmetric hyperbolic. Fix M, M' and $s > \frac{d}{2} + 1$. Let Ω denote the set (3.38) with $\lambda_* > \lambda(M, M')$. Let $h \in H^s(\mathbb{R}^d)$ and $a \in C^0 H^s(\Omega)$ be such that $\|h\|_{L^\infty(\omega)} \le M'$, $\partial_t a \in C^0 H^{s-1}(\Omega)$ and $\|a\|_{L^\infty(\Omega)} \le M$. Then there exists $T > 0$, such that the Cauchy problem has a unique solution $u \in C^0 H^s(\Omega \cap \{t \le T\})$.*

4. Equations with parameters

A general feature of problems in optics is that very different scales are present: for instance, the wavelength of the light beam is much smaller than the length of propagation, the length of the beam is much larger than its width. Many models (Lorentz, anharmonic, Maxwell–Bloch, Euler–Maxwell) contain many parameters, which may be large or small. In applications to optics, we are facing two opposite requirements:

- optics concerned with high frequency regimes, that is functions with Fourier transforms localized around large values of the wave number ξ;
- we want to consider waves with large enough amplitude so that nonlinear effects are present in the propagation of the main amplitude.

Obviously, large frequencies and not too small amplitudes are incompatible with uniform H^s bounds for large s. Therefore, a direct application of Theorem 3.2 for highly oscillatory but not small data, yields existence and stability for $t \in [0, T]$ with T very small, and often much smaller than any relevant physical time in the problem. This is why, one has to keep track of the parameters in the equations and to look for existence or stability results which are independent of these parameters. In this section, we give two examples of such results, which will be used for solving envelope equations and proving the stability of approximate solutions, respectively.

Note that all the results given in this section have local analogues on domains of determinacy, in the spirit of Section 3.6.

4.1. *Singular equations*

We start with two examples which will serve as a motivation:

- The Lorentz model (2.11) (see also (2.21) and (2.24)) contains a large parameter ω in front of the zeroth order term. Similarly, Bloch's equations (2.22) contain a small parameter ε in front of the derivative ∂_t.
- To take into account the multiscale character of the phenomena, one can introduce explicitly the *fast scales* and look for solutions of (3.1) of the form

$$u(t,x) = U(t,x,\varphi(t,x)/\varepsilon) \tag{4.1}$$

where φ is valued in $\mathbb{R}^m$, and U is a function of (t, x) and additional *independent* variables $y = (y_1, \dots, y_m)$.

Both cases lead to equations of the form

$$A_0(a, u)\partial_t u + \sum_{j=1}^{d} A_j(a, u)\partial_{x_j} u + \frac{1}{\varepsilon}\mathcal{L}(a, u, \partial_x)u = F(a, u) \tag{4.2}$$

with possibly an augmented number of variables x_j and an augmented number of parameters a. In (4.2)

$$\mathcal{L}(a, u, \partial_x) = \sum_{j=1}^{d} \mathcal{L}_j(a, u)\partial_{x_j} + \mathcal{L}_0(a, u). \tag{4.3}$$

We again assume that $F(0, 0) = 0$. This setting occurs in many other fields, in particular in fluid mechanics, in the study of low Mach number flows (see e.g. [86,87,115]) or in the analysis of rotating fluids.

Multiplying by a symmetrizer $S(a, u)$, if necessary, we assume that the following condition is satisfied:

ASSUMPTION 4.1 (*Symmetry*). For $j \in \{0, \dots, d\}$, the matrices $A_j(a, u)$ are self-adjoint and in addition $A_0(a, u)$ is positive definite.

For all $j \in \{1, \dots, m\}$ the matrices $\mathcal{L}_j(a, u)$ are self-adjoint and $\mathcal{L}_0(a, u)$ is skew adjoint.

Theorem 3.2 implies that the Cauchy problem is locally well-posed for systems (4.2), but the time of existence given by this theorem in general shrinks to 0 as ε tends to 0. To have a uniform interval of existence, additional conditions are required. We first give two examples, before giving hints for a more general discussion.

4.1.1. The weakly nonlinear case Consider a system (4.2) where all the coefficients A_j and $\mathcal{L}_k$ are functions of $(\varepsilon a, \varepsilon u)$. Expanding $\mathcal{L}_k(\varepsilon a, \varepsilon u) = \underline{\mathcal{L}}_k + \varepsilon \widetilde{A}_k(\varepsilon, a, u)$ yields systems with the following structure:

$$A_0(\varepsilon a, \varepsilon u)\partial_t u + \sum_{j=1}^{d} A_j(a, u)\partial_{x_j} u + \frac{1}{\varepsilon}\underline{\mathcal{L}}(\partial_x)u = F(a, u) \tag{4.4}$$

where $\underline{\mathcal{L}}$ has the form (4.3) with *constant coefficients* $\underline{\mathcal{L}}_j$. We still assume that the symmetry Assumption 4.1 is satisfied and $F(0, 0) = 0$. The matrices A_j and F could also depend smoothly on ε, but for simplicity we forget this easy extension.

THEOREM 4.2 (*Uniform local Existence*). *Suppose that* $h \in H^s(\mathbb{R}^d)$ *and* $a \in C^0([0; T]; H^s(\mathbb{R}^d)) \cap C^1([0; T]; H^{s-1}(\mathbb{R}^d))$, *where* $s > \frac{d}{2} + 1$. *Then, there exists* $T' > 0$ *such that, for all* $\varepsilon \in]0, 1]$, *the Cauchy problem for* (4.4) *with initial data* h *has a unique solution* $u \in C^0([0; T']; H^s(\mathbb{R}^d))$.

SKETCH OF PROOF. Consider the linear version of the equation

$$L_\varepsilon(a, \partial)u := A_0(\varepsilon a)\partial_t u + \sum_{j=1}^{d} A_j(a)\partial_{x_j} u + \frac{1}{\varepsilon}\underline{\mathcal{L}}(\partial_x)u = f. \tag{4.5}$$

Thanks to the symmetry, the L^2 estimate is found immediately. The following expression holds

$$\|u(t)\|_{L^2} \leq C_0 \mathrm{e}^{K_0 t} \|u(0)\|_{L^2} + C_0 \int_t^{t'} \mathrm{e}^{K_0(t-t')} \|L_\varepsilon(a, \partial)u\|_{H^s} \,\mathrm{d}t' \tag{4.6}$$

with C_0 [resp. K_0] depending only on the L^∞ norm [resp. $W^{1,\infty}$ norm] of εa.

Next, one commutes $A_0^{-1} L_\varepsilon$ with the derivatives ∂_x^α as in (3.29). The key new observation is that the derivatives $\partial_x^\alpha(\varepsilon^{-1} A_0^{-1}(\varepsilon a)\underline{\mathcal{L}}_j)$ are bounded with respect to ε, as well as the derivatives $\partial_x^\alpha(A_0^{-1}(\varepsilon a)A_j(a))$. The precise estimate is that for $s > \frac{d}{2} + 1$, there is K_s which depends only on the H^s norm of $a(t)$, such that for $|\alpha| \leq s$, there holds

$$\|[\partial_x^\alpha, A_0^{-1} L_\varepsilon(a, \partial)]u(t)\|_{L^2} \leq K_s \|u(t)\|_{H^s}. \tag{4.7}$$

From here, the proof is as in the nonsingular case. □

4.1.2. The case of prepared data We relax here the weakly nonlinear dependence of the coefficient A_0 and consider a system:

$$A_0(a, u)\partial_t u + \sum_{j=1}^{d} A_j(a, u)\partial_{x_j} u + \frac{1}{\varepsilon}\underline{\mathcal{L}}(\partial_x)u = F(a, u) \tag{4.8}$$

and its linear version

$$L_\varepsilon(a, \partial)u := A_0(a)\partial_t u + \sum_{j=1}^{d} A_j(a)\partial_{x_j} u + \frac{1}{\varepsilon}\underline{\mathcal{L}}(\partial_x)u = f \tag{4.9}$$

where $\underline{\mathcal{L}}$ has *constant coefficients* $\underline{\mathcal{L}}_j$. We assume that the symmetry Assumption 4.1 is satisfied and $F(0, 0) = 0$.

The commutators $[\partial_x^\alpha, \varepsilon^{-1} A_0^{-1}(a) L_\varepsilon]$ are of order ε^{-1}, so that the method of proof of Theorem 4.2 cannot be used anymore. Instead, one can use the following path: the commutators $[\partial_x, \varepsilon^{-1}\underline{\mathcal{L}}]$ are excellent, since they vanish identically. Thus one can try to commute ∂_x^α and $L_\varepsilon(a)$ directly. However, this commutator contains terms $\partial_x^{\alpha-\beta} A_0(a)\partial_t \partial_x^\beta u$ and hence the mixed time-space derivative $\partial_t \partial_x^\beta u$. One cannot use the equation to replace $\partial_t u$ by the spatial derivative, since this would reintroduce singular terms. Therefore, to close the estimate one is led to estimate all the derivatives $\partial_t^{\alpha_0} \partial_x^\alpha u$. Then, the commutator argument closes, yielding an existence result on a uniform interval of time, *provided that* the initial values of *all* the derivatives $\partial_t^{\alpha_0}\partial_x^{\alpha}u$ are uniformly bounded. Let us proceed to the details.

For $s \in \mathbb{N}$, introduce the space $CH^s([0,T]\times\mathbb{R}^d)$ of functions $u \in C^0([0,T];H^s(\mathbb{R}^d))$ such that for all $k \le s$, $\partial_t^k u \in C^0([0,T];H^{s-k}(\mathbb{R}^d))$. It is equipped with the norm

$$\|u\|_{CH^s} = \sup_{t\in[0,T]} |||u(t)|||_s, \qquad |||u(t)|||_s = \sum_{k=0}^{s} \|\partial_t^k u(t)\|_{H^{s-k}(\mathbb{R}^d)}. \tag{4.10}$$

For $s > \frac{d}{2}$, the space CH^s is embedded in L^∞. Therefore, the chain rule and Lemma 3.5 imply that for smooth F, such that $F(0)=0$, the mapping $u \mapsto F(u)$ is continuous from CH^s into itself and maps bounded sets to bounded sets.

Moreover, the commutator $[\partial_{t,x}^{\alpha}, L_\varepsilon(a,\partial)]u$ is a linear combination of terms

$$\partial_{t,x}^{\gamma}A_j(a)\partial_{t,x}^{\beta}u, \tag{4.11}$$

with

$$0 \le |\gamma|-1 \le s-1, \quad 0 \le |\beta|-1, \quad |\gamma|-1+|\beta|-1 \le s-1.$$

Since $A_j(a(t)) \in CH^s$, (ii) of Lemma 3.5 applied with $\sigma = s-1$ implies for $s > \frac{d}{2}+1$ the following commutator estimates

$$\left\| [\partial_{t,x}^{\alpha}, L_\varepsilon(a,\partial_x)]u(t) \right\|_{L^2} \le C|||u(t)|||_s \tag{4.12}$$

where C depends on the norm $|||a(t)|||_s$.

For systems (4.9), the L^2 energy estimates are again straightforward from the symmetry assumption. Together with the commutator estimates above, they provide bounds for the norms $|||u(t)|||_s$, uniform and t and ε, *provided that their initial values* $|||u(0)|||_s$ *are bounded.*

The initial values $\partial_t^k u(0)$ are computed by induction using the equation: for instance

$$\partial_t u_{|t=0} = -A_0^{-1}(a_0,u_0) \left(\sum_{j=1}^{d} A_j(a_0,u_0)\partial_{x_j}u_0 + \frac{1}{\varepsilon}\underline{\mathcal{L}}(\partial_x)u_0 - F(a_0,u_0) \right) \tag{4.13}$$

where $a_0 = a_{|t=0}$ and $u_0 = u_{|t=0}$. In particular, for a fixed initial data $u_{|t=0} = h$, the term $\partial_t u_{|t=0}$ is bounded independently of ε if and only if

$$\underline{\mathcal{L}}(\partial_x)h = 0. \tag{4.14}$$

The analysis of higher order derivatives is similar. To simplify the notation, let $\partial^{(k)}u$ denote a product of derivatives of u of total order k:

$$\partial^{\alpha_1}u\ldots\partial^{\alpha_p}u \quad \text{with } \alpha_j > 0 \quad \text{and} \quad |\alpha_1|+\cdots+|\alpha_p| = k. \tag{4.15}$$

LEMMA 4.3. *For $s > (d+1)/2$ and $k \in \{0, \dots, s\}$, there are nonlinear functionals $\mathcal{F}_k^\varepsilon(a,u)$, which are finite sums of terms*

$$\varepsilon^{-l}\Phi(a,h)(\partial_{t,x}^{(p)}a)(\partial_x^{(p)}u), \quad l \le k,\ p+q \le k \tag{4.16}$$

with Φ smooth, such that for $a \in CH^s$ and $h \in H^s$ and all $\varepsilon > 0$, the local solution of the Cauchy problem for (4.8) with initial data h belongs to CH^s and

$$\partial_t^k u = \mathcal{F}_k^\varepsilon(a,h). \tag{4.17}$$

PROOF. For C^∞ functions, (4.17) is immediate by induction on k. Lemma 3.5 implies that the identities extend to coefficients $a \in CH^s$ and $u \in C^0([0,T'], H^s)$, proving by induction that $\partial_t^k u \in C^0([0,T'], H^{s-k})$. □

In particular, if u is a solution of (4.8) with initial data h, there holds

$$(\partial_t^k u)_{|t=0} = \mathcal{H}_k^\varepsilon(a,h) := \mathcal{F}_k^\varepsilon(a,h)_{|t=0} \tag{4.18}$$

Note that $\mathcal{H}_k^\varepsilon$ is singular as $\varepsilon \to 0$, since in general it is of order ε^{-k}.

THEOREM 4.4. *Suppose that $s > \frac{d}{2}+1$ and $a \in CH^s([0,T]\times\mathbb{R}^d)$ and $h \in H^s(\mathbb{R}^d)$ are such that the families $\mathcal{H}_k^\varepsilon(a,h)$ are bounded for $\varepsilon \in]0,1]$.*

Then, there exists $T' > 0$ such that for all $\varepsilon \in]0,1]$ the Cauchy problem for (4.8) with initial data h has a unique solution $u \in CH^s([0,T']\times\mathbb{R}^d)$.

SKETCH OF THE PROOF (*See e.g. [11]*). The assumption means that the initial norms $|||u(0)|||_s$ are bounded, providing uniform estimates of $|||u(t)|||_s$ for $t \le T'$, for some $T' > 0$ independent of ε. This implies that the local solution can be continued up to time T'. □

The data which satisfy the condition for $\mathcal{H}_k^\varepsilon(a,h)$ are often called *prepared data*. The first condition (4.14) is quite explicit, but the higher order conditions are less explicit, and the construction of prepared data is a nontrivial independent problem. However, there is an interesting application of Theorem 4.4 when the wave is created not by an initial data but by a forcing term which vanishes in the past: consider the problem

$$A_0(a,u)\partial_t u + \sum_{j=1}^d A_j(a,u)\partial_{x_j}u + \frac{1}{\varepsilon}\underline{\mathcal{L}}(\partial_x)u = F(a,u) + f \tag{4.19}$$

with $F(a,0) = 0$. (In the notations of (4.8), this means that $f = F(a,0)$). We consider f as one of the parameters entering the equation. We assume that f is given in $CH^s([0,T]\times\mathbb{R}^d)$ and vanishes at order s on $t = 0$:

$$\partial_t^k f_{|t=0} = 0, \qquad k \in \{0,\dots,s-1\}. \tag{4.20}$$

Then, one can check by induction, that for a vanishing initial data $h = 0$ the traces of the solution vanish:

$$\partial_t^k u_{|t=0} = 0, \qquad k \in \{0, \dots, s\}. \tag{4.21}$$

Therefore:

THEOREM 4.5. *Suppose that $s > \frac{d}{2} + 1$, $a \in CH^s([0, T] \times \mathbb{R}^d)$ and $f \in CH^s([0, T] \times \mathbb{R}^d)$ satisfies (4.20).*

Then, there exists $T' > 0$ such that for all $\varepsilon \in]0, 1]$ the Cauchy problem for (4.19) with vanishing initial data has a unique solution $u \in CH^s([0, T'] \times \mathbb{R}^d)$ and u satisfies (4.21).

4.1.3. Remarks on the commutation method The proofs of Theorems 3.2, 4.2 and 4.4 are based on commutation properties which strongly depend on the structure of the equation. In particular, it was crucial in (4.4) and (4.8) that the singular part $\underline{\mathcal{L}}(\partial_x)$ had *constant* coefficients and that A_0 depended on $(\varepsilon a, \varepsilon u)$ in the former case. Without assuming maximal generality, we investigate here to what extent the commutation method can be extended to more general equations (4.2) where $\mathcal{L}$ and A_0 could also depend on a.

1. *The principle of the method*. Consider a linear equation $L(t, x, \partial)u = f$ and assume that there is a good L^2 energy estimate. We suppose that the coefficients are smooth functions of a and that a is smooth enough. Consider a set of vector fields $\mathcal{Z} = \{Z_1, \dots, Z_m\}$, which we want to commute with the equation. The commutation properties we use are of the form:

$$L(t, x, \partial)Z_j = Z_j L(t, x, D) + \sum R_{j,k} Z_k$$

with bounded matrices $R_{j,k}$. The energy estimate is then applied to $Z_j u$. However, one must keep in mind that one could perform a change of variables $u = Vv$ or pre-multiply the equation by a matrix before commuting with the Z_j. This is equivalent to adding zero order terms to the Z_j and looking for relations of the form

$$L(t, x, \partial)(Z_j + B_j) = (Z_j + C_j)L(t, x, D) + \sum R_{j,k} Z_k + R_{j,0} \tag{4.22}$$

with matrices B_j and C_j to be found, depending on a and its derivatives. Together with the L^2 estimate, this clearly implies estimates of derivatives $Z_{j_1} \dots Z_{j_m} u$, and knowing similar estimates for the right hand side f and the initial data.

2. *Analysis of the commutation conditions.* The operator L is split into a good part, the linear combination of the Z_j, and a bad or singular part:

$$L(t, x, \partial) = \sum A_j(t, x) Z_j + \sum G_k(t, x) T_k \tag{4.23}$$

where the T_k are some other vector fields or singular terms which commute with the Z_j. Only the commutation with terms $G_k(t, x)T_k$ may cause problems, and the commutation conditions read

$$Z_j(G_k) = G_k B_j - C_j G_k, \tag{4.24}$$

assuming that

$$\text{the } [T_k, B_j] \text{ are bounded.} \tag{4.25}$$

We now analyze the geometric implications of (4.24).

LEMMA 4.6. *Suppose that the G_k are smooth matrices and $\mathcal{Z}$ is an integrable system of vector fields. Then, locally, there are smooth matrices B_j and C_j satisfying (4.24), if and only if there are smooth invertible matrices W and V, such that*

$$\forall j, \forall k, \quad [Z_j, WG_kV] = 0. \tag{4.26}$$

In this case, for all $\eta \in \mathbb{R}^m$, the rank of $\sum \eta_k G_k$ is constant along the integral leaves of $\mathcal{Z}$.

PROOF. If $\widetilde{G}_k = WG_kV$, then

$$Z_j(\widetilde{G}_k) = W\left(W^{-1}(Z_jW)G_k + Z(G_k) + G_kZ_j(V)V^{-1}\right)V.$$

Thus, if $Z(\widetilde{G}_k) = 0$, then (4.24) holds with $C_j = W^{-1}Z_j(W)$ and $B_j = -Z_j(V)V^{-1}$.

Conversely, one can assume locally that $\mathcal{Z} = \{\partial_{x_1}, \ldots, \partial_{x_m}\}$ and prove the result by induction on m. One determines locally W_1 and V_1 such that

$$\partial_{x_1}W_1 = W_1C_1, \quad \partial_{x_1}V_1 = -B_1V_1$$

and (4.24) implies $\partial_{x_1}(W_1G_1V_1) = 0$. This finishes the proof when $m = 1$. When $m > 1$, let $\widetilde{G}_k = W_1G_kV_1$. The commutation properties (4.24) are stable under such a transform: the $\widetilde{G}_k$ satisfy

$$\partial_{x_j}\widetilde{G}_k = \widetilde{G}_k\widetilde{B}_j - \widetilde{C}_j\widetilde{G}_k, \tag{4.27}$$

with new matrices $\widetilde{B}_j$ and $\widetilde{C}_j$, which vanish when $j = 1$. In particular, the $\widetilde{G}_k$ do not depend on the first variable x_1. Freezing the variable x_1 at $\underline{x}_1$, we see that the commutation relation (4.27) also holds with matrices $\widetilde{B}_j$ and $\widetilde{C}_j$ that are independent of x_1. From here, one can apply the induction hypothesis for the matrices $\widetilde{G}_k$ and the vector fields $\{\partial_{x_j}, j \geq 2\}$ and find matrices W' and V', independent of x_1, such that $\partial_{x_j}(W'\widetilde{G}_kV) = 0$ for $j \geq 2$. The property (4.26) follows with $W = W'W_1$ and $V = V_1V'$.

The condition (4.26) implies that $W(\sum \eta_kG_k)V$ is independent of the variables $(x_1, \cdots, x_m)$. In particular, its rank is constant when the other variables are fixed. □

REMARK 4.7. Suppose that $l \geq 2$ and that G_1 is invertible. Saying that the rank of $\sum \eta_kG_k$ is independent of x, is equivalent to saying that the eigenvalues of $\sum_{k\geq 2}\eta_kG_1^{-1}G_k$ are independent of x as well as their multiplicity.

The next lemma shows that the commutation properties imply that the equation can be transformed to another equivalent equation with constant coefficients in the singular part.

LEMMA 4.8. (i) *Suppose that $G(a)$ is a smooth matrix of rank independent of a. Then, locally near any point $\underline{a}$, there are smooth matrices $W(a)$ and $V(a)$ such that $W(a)G(a)V(a)$ is constant.*

(ii) *Suppose that $G_0(a)$ and $G(a)$ are smooth self-adjoint matrices with G_0 positive definite, such that the eigenvalues of $G_0^{-1}G(a)$ are independent of a. Then locally, there is a smooth matrix $V(a)$ such that, with $W = V^{-1}G_0^{-1}$, $W(a)G_0(a)V(a)$ and $W(a)G(a)V(a)$ are constant.*

In both cases, if $a(t,x)$ is such that the $Z_j a$ are bounded, the commutation relations (4.24) *are satisfied with bounded matrices B_j and C_j.*

PROOF. (i) If the rank of G is constant equal to μ, in a neighborhood of $\underline{a}$, there are smooth invertible matrices W and V such that

$$W(a)G(a)V(a) = \begin{pmatrix} \mathrm{Id}_\mu & 0 \\ 0 & 0 \end{pmatrix}.$$

(ii) Since G_0 is positive definite and G is self-adjoint, $G_0^{-1}G$ is diagonalizable. Since the eigenvalues are constant, say equal to $\underline{\lambda}_j$, there is a smooth invertible matrix V such that

$$V^{-1}(a)G_0^{-1}(a)G(a)V(a) = \mathrm{diag}(\underline{\lambda}_j). \qquad \square$$

CONCLUSION 4.9. *Lemmas* 4.6 and 4.8 *give key indications for the validity of the commutation method, applied to equations* (4.23).

4.1.4. Application 1 Consider a system with coefficients depending smoothly on a

$$A_0(a)\partial_t u + \sum_{j=1}^{d} A_j(a)\partial_{x_j} u + \frac{1}{\varepsilon}B(a) = f. \tag{4.28}$$

We assume that the matrices A_j are self-adjoint with A_0 positive definite and $B(a)$ skew-adjoint.

• For $\mathcal{Z} = \{\partial_t, \partial_{x_1}, \ldots, \partial_{x_d}\}$, the commutation condition reads

$$\text{the rank of } B(a(t,x)) \text{ is independent of } (t,x). \tag{4.29}$$

Locally in a, this is equivalent to

$$\begin{aligned}&\text{there are smooth matrices } W(t,x) \text{ and } V(t,x) \text{ such that}\\ &W(t,x)B(a(t,x))V(t,x) \text{ is constant.}\end{aligned} \tag{4.30}$$

THEOREM 4.10. *Suppose that a is a smooth function of (t, x). Under Assumption (4.30), there are energy estimates in the spaces $CH^s([0, T] \times \mathbb{R}^d)$ for the solutions of (4.28), of the form*

$$|||u(t)|||_s \leq C\mathrm{e}^{Kt}|||u(0)|||_s + C\int_0^t \mathrm{e}^{K(t-t')}|||f(t')|||_s \mathrm{d}t'. \tag{4.31}$$

• For $\mathcal{Z} = \{\partial_{x_1}, \dots, \partial_{x_d}\}$, then ∂_t becomes a "bad" vector field to be included in the second sum in (4.23). Following Remark 4.7, the condition reads

$$\text{the eigenvalues of } A_0^{-1}(a(t,x))B(a(t,x)) \text{ have constant multiplicity.} \tag{4.32}$$

Locally in a, this is equivalent to

$$\text{there is a smooth matrix } V(t,x) \text{ such that } V^{-1}(t,x)A_0^{-1}(a(t,x))B(a(t,x))V(t,x) \text{ is constant.} \tag{4.33}$$

THEOREM 4.11. *Suppose that a is a smooth function of (t, x). Under Assumption (4.33), there are energy estimates in the spaces $C^0([0, T]; H^s(\mathbb{R}^d))$ for the solutions of (4.28), of the form*

$$\|u(t)\|_s \leq C\mathrm{e}^{Kt}\|u(0)\|_s + C\int_0^t \mathrm{e}^{K(t-t')}\|f(t')\|_s \mathrm{d}t'. \tag{4.34}$$

4.1.5. Application 2 The following systems arise when one introduces fast scales (see Section 7.4):

$$A_0(a)\partial_t u + \sum_{j=1}^{d} A_j(a)\partial_{x_j}u + \frac{1}{\varepsilon}\left(\sum_{k=1}^{m} B_j(a)\partial_{\theta_j}u + E(a)u\right) = f. \tag{4.35}$$

We assume that the matrices A_j and B_k are self-adjoint with A_0 is positive definite and E is skew symmetric. The additional variables $\theta = (\theta_1, \dots, \theta_m)$ correspond to the fast variables φ/ε in (4.1). In this framework, functions are periodic in θ, and up to a normalization of periods, this means that $\theta \in \mathbb{T}^m$, $\mathbb{T} = \mathbb{R}/(2\pi\mathbb{Z})$. There are other situations where systems of the form (4.35) occur with no periodicity assumption in the variables θ_j, for instance in the low mach limit analysis Euler equation. The analysis can be adapted to some extent for this case, but we do not investigate this question in detail here.

An important assumption is that

$$\partial_{\theta_j} a = 0, \qquad j \in \{1, \dots, m\}. \tag{4.36}$$

In this case one can expand u and f in Fourier series in θ, and the Fourier coefficients $\hat{u}_\alpha$, $\alpha \in \mathbb{Z}^m$ are solutions of

$$A_0(a)\partial_t \hat{u}_\alpha + \sum_{j=1}^{d} A_j(a)\partial_{x_j} \hat{u}_\alpha + \frac{1}{\varepsilon}\mathcal{B}(a,\alpha)\hat{u}_\alpha = \hat{f}_\alpha \tag{4.37}$$

where

$$\mathcal{B}(a,\eta) = \sum_{j=1}^{m} i\eta_j B_j(a) + E(a). \tag{4.38}$$

They are systems of the form (4.28). The conditions (4.29) (4.32) read

$$\text{for all } \alpha \in \mathbb{Z}^m\text{, the rank of } \mathcal{B}(a(t,x),\alpha) \text{ is independent of } (t,x), \tag{4.39}$$

$$\begin{aligned}&\text{for all } \alpha \in \mathbb{Z}^m\text{, the eigenvalues of } A_0^{-1}(a(t,x))B(a(t,x),\alpha)\\ &\text{have constant multiplicity,}\end{aligned} \tag{4.40}$$

respectively. To get estimates uniform in α, we reinforce their equivalent formulation (4.29), (4.32) as follows:

ASSUMPTION 4.12. There are matrices $\mathcal{W}(t,x,\alpha)$ and $\mathcal{V}(t,x,\alpha)$ for $\alpha \in \mathbb{Z}^m$, which are uniformly bounded, as are their derivatives, with respect to (t,x), and a matrix $\mathcal{B}^\sharp(\alpha)$ independent of (t,x) such that

$$\mathcal{W}(t,x,\alpha)\mathcal{B}(a(t,x),\alpha)\mathcal{V}(t,x,\alpha) = \mathcal{B}^\sharp(\alpha). \tag{4.41}$$

ASSUMPTION 4.13. There are matrices $\mathcal{V}(a(t,x),\alpha)$ for $\alpha \in \mathbb{Z}^m$, which are uniformly bounded, as are their derivatives, with respect to (t,x), and a matrix $\mathcal{B}^\sharp(\alpha)$ independent of (t,x) such that

$$\mathcal{V}^{-1}(t,x,\alpha)A_0^{-1}(a(t,x))\mathcal{B}(a(t,x),\alpha)\mathcal{V}(t,x,\alpha) = \mathcal{B}^\sharp(\alpha). \tag{4.42}$$

Note that no smoothness in α is required in these assumptions.

THEOREM 4.14. *Suppose that a is a smooth function of (t,x).*

(i) *Under Assumption 4.12 there are energy estimates in the spaces $CH^s([0,T]\times\mathbb{R}^d\times\mathbb{T}^m)$ for the solutions of (4.35)*

$$|||u(t)|||_s \le C\mathrm{e}^{Kt}|||u(0)|||_s + C\int_0^t \mathrm{e}^{K(t-t')}|||f(t')|||_s \mathrm{d}t'. \tag{4.43}$$

(ii) *Under Assumption 4.13, there are energy estimates in the spaces* $C^0([0,T]; H^s(\mathbb{R}^d \times \mathbb{T}^m))$

$$\|u(t)\|_{H^s} \le C e^{Kt} \|u(0)\|_{H^s} + C \int_0^t e^{K(t-t')} \|f(t')\|_{H^s} \, dt'. \tag{4.44}$$

PROOF. The symmetry conditions immediately imply an L^2 energy estimate for (4.37). The commutation properties are satisfied by $\varepsilon^{-1}\mathcal{B}(a,\alpha)$ in the first case, and by $\{A_0(a)\partial_t, \varepsilon^{-1}\mathcal{B}(a,\alpha)\}$ in the second case, providing estimates which are uniform in α. Moreover, the equation commutes with $|\alpha|$, allowing for estimates of the ∂_θ derivatives. □

REMARK 4.15. These estimates extend to equations

$$\begin{aligned} &A_0(a,\varepsilon u)\partial_t u + \sum_{j=1}^d A_j(a,\varepsilon u)\partial_{x_j} u \\ &\quad + \frac{1}{\varepsilon}\left(\sum_{k=1}^m B_j(a,\varepsilon u)\partial_{\theta_j} u + E(a,\varepsilon u)u\right) = f \end{aligned} \tag{4.45}$$

since all the additional commutators which involve εu are nonsingular. We stress that the conditions (4.41) or (4.42) are unchanged and bear on the coefficients $A_j(a,v)$ and $B_j(a,v)$ *at* $v=0$. This yields existence theorems, which extend Theorem 4.4 for data prepared under Assumption 4.12, and Theorem 4.2 for general data under Assumption 4.13. We refer to [76] for details.

Next we give explicit conditions which ensure that the Assumptions are satisfied. The first remark is trivial, but useful.

REMARK 4.16. If $a(t,x) = \underline{a}$ is constant, the Assumptions 4.12 and 4.13 are satisfied.

Next we consider the case where $E = 0$. Though the next proposition does not apply to geometric optics, it is natural and useful for other applications.

PROPOSITION 4.17. *Suppose that* $E = 0$ *and there is* $\delta > 0$ *such that for all* a *and* η *in the sphere* S^{m-1}*, 0 is the unique eigenvalue of* $\mathcal{B}(a,\eta)$ *in the disk centered at 0 of radius* δ*, and that its multiplicity is constant.*

Then, the Assumption 4.12 is satisfied.

In Section 7 we will consider systems of the above form, such that the symmetric system $\mathcal{B}(a(t,x),\eta)$ is hyperbolic in a direction $\underline{\eta}$, which we can choose to be $\underline{\eta} = (1,0,\ldots,0)$. We use the notations

$$\eta = (\eta_1, \eta'), \qquad \mathcal{B}(a,\eta) = \eta_1 B_1(a) + \mathcal{B}'(a,\eta') \tag{4.46}$$

and B_1 is positive definite. Then the kernel of $\mathcal{B}(a(t,x),\eta)$ is nontrivial if and only if $-\eta_1$ is an eigenvalue of $B_1^{-1}\mathcal{B}'(a(t,x),\eta')$. In this case, a natural choice for a projector $P(t,x,\eta)$ on $\ker \mathcal{B}$ is to consider the spectral projector, which is orthogonal for the scalar product defined by $B_1(a(t,x))$. If $-\tau_1$ is not an eigenvalue, define $P(t,x,\eta) = 0$.

PROPOSITION 4.18. *With $E = 0$, assume that the symmetric symbol $\mathcal{B}$ is hyperbolic with time-like co-direction $\underline{\eta} \in \mathbb{R}^m \setminus \{0\}$ and use the notations above. Then, locally in (t, x), the Assumption 4.12 is satisfied if*

(i) *for all $\eta \in \mathbb{R}^m$ the rank of $\mathcal{B}(a(t, x), \eta)$ is independent of (t, x),*
(ii) *for all $\eta \in \mathbb{R}^m$ the projectors $P(t, x, \eta)$ are smooth in (t, x), and the set $\{P(\cdot, , \eta)\}_{\{\eta\in\mathbb{R}^m\}}$ is bounded in C^∞.*

In particular, the second condition is satisfied when the eigenvalues of $B_1^{-1}\mathcal{B}'(a(t, x), \eta')$ have their multiplicity independent of (t, x, η'), for η' in the unit sphere.

PROOF. Condition (i) implies that for all η', the eigenvalues of $B_1^{-1}\mathcal{B}'(a(t, x), \eta')$, which are all real by the hyperbolicity assumption, are independent of (t, x). By (ii), locally, there are smooth families of eigenvectors, which are orthonormal for the scalar product $A_0(a(t, x))$, and uniformly bounded in C^∞ independently of η'. This yields matrices $\mathcal{V}(t, x, \eta')$ such that $\mathcal{V}(\cdot, \eta')$ and $\mathcal{V}^{-1}(\cdot, \eta')$ are uniformly bounded in C^∞ and

$$\mathcal{V}^{-1}(t, x, \eta')B_1^{-1}a(t, x)\mathcal{B}'(a(t, x), \eta')\mathcal{V}(t, x, \eta') = \operatorname{diag}(\lambda_j(\eta')) := \mathcal{B}^\flat(\eta').$$

Thus, with $\mathcal{W} = \mathcal{V}^{-1}B_1(a(t, x), \eta')$ there holds

$$\mathcal{W}(t, x, \eta')\mathcal{B}(a(t, x), \eta)\mathcal{V}(t, x, \eta') = \mathrm{i}\eta_1\mathrm{Id} + \mathcal{B}^\flat(\eta') := \mathcal{B}^\flat(\eta).$$

If the eigenvalues of $B_1^{-1}\mathcal{B}'(a(t, x), \eta')$ have constant multiplicity in (t, x) and $\eta' \neq 0$, the spectral projectors are smooth in (t, x, η') and homogeneous of degree 0 in η', yielding (ii). □

Concerning Assumption 4.13, the same proof applied to $\mathrm{i}\tau A_0 + \mathcal{B}(a, \eta)$ implies the following:

PROPOSITION 4.19. *With $E = 0$, the Assumption 4.12 is satisfied locally in (t, x) if:*

(i) *for all $\eta \in \mathbb{R}^m$, the eigenvalues of $A_0^{-1}(a(t, x))\mathcal{B}(a(t, x), \eta)$ are independent of (t, x),*
(ii) *the spectral projectors of $A_0^{-1}(a(t, x))\mathcal{B}(a(t, x), \eta)$ are smooth in (t, x), and belong to a bounded set in C^∞ when $\eta \in \mathbb{R}^m$.*

In particular, the second condition is satisfied if the multiplicities of the eigenvalues of $A_0^{-1}(a(t, x))\mathcal{B}(a(t, x, \eta))$ are constant in (t, x, η) for $\eta \neq 0$.

In Section 7, we will use the following extensions:

PROPOSITION 4.20. *Suppose that*

$$\widetilde{\mathcal{B}}(a, \tilde{\eta}) = \sum_{j=1}^{\tilde{m}} \mathrm{i}\tilde{\eta}_j \widetilde{B}_j(a) \tag{4.47}$$

satisfies the assumption of Proposition 4.18 [resp. Proposition 4.19], and that there is a $\tilde{m} \times m$ matrix M such that

$$\mathcal{B}(a, \eta) = \widetilde{\mathcal{B}}(a, M\eta). \tag{4.48}$$

Then, $\mathcal{B}$ satisfies the Assumption 4.12 [resp. Assumption 4.13].

There are analogous results when $E \neq 0$, but we omit them here for the sake of brevity.

4.2. *Equations with rapidly varying coefficients*

In this subsection we consider systems

$$A_0(a, u)\partial_t u + \sum_{j=1}^{d} A_j(a, u)\partial_{x_j} u + \frac{1}{\varepsilon} E(a)u = F(a, u) \tag{4.49}$$

with the idea that a and u have rapid oscillations. More precisely, we assume that $a \in L^\infty([0, T] \times \mathbb{R}^d)$ and that its derivative satisfy $0 < |\alpha| \leq s$:

$$\varepsilon^{|\alpha|-1} \|\partial_{t,x}^\alpha a\|_{L^\infty([0,T]\times\mathbb{R}^d)}. \leq C_1, \tag{4.50}$$

Following the general existence theory, it is assumed that the integer s satisfies $s > \frac{d}{2} + 1$. We suppose that $u = 0$ is almost a solution, that is that there is a real number $M > 0$ such that $f = F(a, 0)$ is satisfied when $|\alpha| \leq s$

$$\varepsilon^{|\alpha|} \|\partial_x^\alpha f^\varepsilon(t)\|_{L^2(\mathbb{R}^d)} \leq \varepsilon^M C_2. \tag{4.51}$$

How large M must be chosen is part of the analysis. Similarly, we consider initial data which satisfy $|\alpha| \leq s$

$$\varepsilon^{|\alpha|} \|\partial_x^\alpha h\|_{L^2(\mathbb{R}^d)} \leq \varepsilon^M C_3. \tag{4.52}$$

Note that these assumptions allow for families of data a, which have amplitude $O(\varepsilon)$ at frequencies $|\xi| \approx \varepsilon^{-1}$. In applications, a will be an approximate solution of the original equation. The equation (4.49) with coefficients $A_j(a, 0)$ is the linearized equation near this approximate solution. Its well-posedness accounts for the stability of the approximate solution and (4.49) can be regarded as an equation for a corrector.

We always assume that the equations are symmetric hyperbolic, that is that the matrices A_j are self-adjoint with A_0 positive definite and that E is skew symmetric.

THEOREM 4.21. *Under the assumptions above, if $M > 1 + d/2$, there are $\varepsilon_1 > 0$ and C_4, depending only on the constants C_1, C_2, C_3 and the coefficients A_j and F, such that for $\varepsilon \in]0, \varepsilon_1]$, the Cauchy problem for (4.49) with initial data h, has a unique solution $u \in C^0([0, T]; H^s(\mathbb{R}^d))$ which satisfies*

$$\varepsilon^{|\alpha|} \|\partial_x^\alpha u(t)\|_{L^2(\mathbb{R})} \leq \varepsilon^M C_4. \tag{4.53}$$

PROOF (*See [59]*). Write $u = \varepsilon^M v$, $f = \varepsilon^M g$ and $F(a, u) = f + \varepsilon^M G(a, u)v$. The equation for v reads

$$L(a, u, \partial)v := A_0(a, u)\partial_t v + \sum_{j=1}^{d} A_j(a, u)\partial_{x_j} v + \frac{1}{\varepsilon} E(a)v - G(a, u)v = g. \tag{4.54}$$

Introduce the norms

$$\|u\|_{H^s_\varepsilon} = \sup_{|\alpha| \le s} \varepsilon^{|\alpha|} \|\partial_x^\alpha u\|_{L^2(\mathbb{R}^d)}. \tag{4.55}$$

The main new ingredient is the weighted Sobolev inequality:

$$\|\varepsilon^M v\|_{L^\infty(\mathbb{R}^d)} \le K \varepsilon^{M - \frac{d}{2}} \|v\|_{H^{s-1}_\varepsilon}, \tag{4.56}$$

$$\|\varepsilon^M \partial_x v\|_{L^\infty(\mathbb{R}^d)} \le K \varepsilon^{M - \frac{d}{2} - 1} \|v\|_{H^s_\varepsilon}. \tag{4.57}$$

Since $M > \frac{d}{2} + 1$, this implies that, if the H^s_ε-norm of v is bounded by some constant C_4, then the Lipschitz norms of the coefficients $A_j(a, \varepsilon^M v)$ are bounded by some constant independent of C_4, if ε is small enough.

Therefore the symmetry implies the following L^2 estimate: there are constants C and K, which depend only on C_1, C_2, C_3, and $\varepsilon_0 > 0$ which depends in addition on C_4, such that for $\varepsilon \le \varepsilon_0$ and u satisfying (4.53) on $[0, T']$, there holds for $t \in [0, T']$:

$$\|v(t)\|_{L^2} \le C e^{Kt} \|v(0)\|_{L^2} + C \int_0^t e^{K(t-t')} \left\| L(a, u, \partial)v(t')\right\|_{L^2} dt'. \tag{4.58}$$

Next, one commutes the equation (pre-mulitplied by A_0^{-1}) with the weighted derivatives $\varepsilon\partial_{x_j}$. One proves that for $|\alpha| \le s$,

$$\left\| [\varepsilon^{|\alpha|}\partial_x^\alpha, A_0^{-1} L(a, u, \partial)]v(t) \right\|_{L^2} \le K \|v(t)\|_{H^s_\varepsilon} \tag{4.59}$$

with K depending only on C_1, C_2, C_3, provided that u satisfies (4.53) and $\varepsilon \le \varepsilon_0$ where $\varepsilon_0 > 0$ depends on C_1, C_2, C_3 and C_4. Indeed, by homogeneity, Lemma 3.5 implies that for $\sigma > \frac{d}{2}$ and $l \ge 0$, $m \ge 0$ with $l + m \le \sigma$, there holds

$$\|uv\|_{H^{\sigma-l-m}_\varepsilon} \le C\varepsilon^{-\frac{d}{2}} \|u\|_{H^{\sigma-l}_\varepsilon} \|v\|_{H^{\sigma-m}_\varepsilon}. \tag{4.60}$$

The commutator $[\varepsilon^{|\alpha|}\partial_x^\alpha, A_0^{-1} A_j \partial_j]v$ is a linear combination of terms

$$\varepsilon^{|\alpha|}\partial_x^{\beta^1} a \ldots \partial_x^{\beta^q} a \partial_x^{\alpha^1} u \ldots \partial_x^{\alpha^p} u \partial_x^\gamma v$$

with $0 < |\beta^j| \le s$, $|\alpha^k| \le s$, $|\gamma| \le s$ and $\sum|\beta^j| + \sum|\alpha^k| + |\gamma| \le |\alpha| + 1 \le s + 1$. These terms belong to L^2, as explained in the proof of Proposition 3.4. More precisely, using the assumptions (4.52) and (4.53) and the product rule (4.60), one obtains that the L^2 norm of such terms is bounded by

$$\varepsilon^{\mu} C C_1^q C_4^p \|v\|_{H^s_\varepsilon}$$

for some numerical constant C and

$$\mu = |\alpha| + pM - \sum(|\beta^j| - 1)_+ - \sum|\alpha^k| - |\gamma| - p\frac{d}{2}.$$

If $\sum|\beta^j| > 0$, then $\sum(|\beta^j| - 1)_+ \le (\sum|\beta^j|) - 1$ and $\mu = 0$ when $p = 0$ and $\mu > 0$ when $p > 0$ since $M > \frac{d}{2}$. If $\sum|\beta^j| = 0$, then $p > 0$ and $\mu > 0$ since $M > \frac{d}{2} + 1$. This implies that this term satisfies (4.59).

The commutator $[\varepsilon^{|\alpha|}\partial_x^\alpha, \varepsilon^{-1}A_0^{-1}E]v$ is a linear combination of terms

$$\varepsilon^{|\alpha|-1}\partial_x^{\beta^1} a \ldots \partial_x^{\beta^q} a \partial_x^\gamma v$$

with $\sum|\beta^j| + |\gamma| \le |\alpha| \le s$ and $\sum|\beta^j| > 0$. With (4.50), this term is also dominated as in (4.59).

Hence, there are constants C and K, which depend only on C_1, C_2, C_3, and $\varepsilon_0 > 0$ which depends in addition on C_4, such that for $\varepsilon \le \varepsilon_0$ and u satisfying (4.53) on $[0, T']$, there holds for $t \in [0, T']$:

$$\|v(t)\|_{H^s_\varepsilon} \le C\mathrm{e}^{Kt}\|v(0)\|_{H^s_\varepsilon} + C\int_0^t \mathrm{e}^{K(t-t')}\left\|L(a, u, \partial)v(t')\right\|_{H^s_\varepsilon} \mathrm{d}t'. \tag{4.61}$$

Choose $C_4 \ge 2C\mathrm{e}^{KT}(C_3 + C_2)$ and ε_0 accordingly, and assume that $\varepsilon \le \varepsilon_0$. Therefore, (4.61) shows that if u satisfies $\|u(t)\|_{H^s_\varepsilon} \le C_4$ on $[0, T']$, then it also satisfies $\|u(t)\|_{H^s_\varepsilon} \le \frac{1}{2}C_4$ on this interval. By continuation, this implies that the local solution u of (4.49) can be continued on $[0, T]$ and satisfies (4.53). □

5. Geometrical Optics

In this section we present the WKB method for two scale asymptotic expansions, applied to the construction of high frequency wave packet solutions. This method rapidly leads to the geometric optics equations, of which the main features are the eikonal equations, the polarization conditions and the transport equation most often along rays. The formal asymptotic solutions can be converted into approximate solutions by truncating the expansion. The next main step is to study their stability, with the aim of constructing exact solutions close to them. We first review the linear case and next we consider the general regime of *weakly nonlinear geometric optics* where general results of existence and stability of oscillating solutions are available. The transport equations are in general nonlinear. In particular Burger's equation appears as the generic transport equations for quasi-linear non-dispersive systems. However, for some equations, because of their special structure, the general results do not allow one to reach nonlinear regimes. In these cases one is led to increase the intensity of the amplitudes. There are cases of such large solutions where the construction of WKB solutions can be carried out. However, the stability analysis is much more delicate and strong instabilities exist. We end this section with several remarks about caustics and the focusing of rays, which is a fundamental feature of multi-dimensional geometric optics.

5.1. *Linear geometric optics*

We sketch below the main outcomes of Lax' analysis (see P. Lax [92]), applied to linear equations of the form

$$L(a, \partial)u := A_0(a)\partial_t u + \sum_{j=1}^{d} A_j(a)\partial_{x_j} u + \frac{1}{\varepsilon} E(a)u = 0 \tag{5.1}$$

where $a = a(t, x)$ is given, with values belonging to some domain $\mathcal{O} \subset \mathbb{R}^M$.

ASSUMPTION 5.1 (*Symmetry*). The matrices A_j and E are smooth functions on $\mathcal{O}$. The A_j are self-adjoint, with A_0 positive definite and E skew-adjoint.

5.1.1. An example using Fourier synthesis Consider a constant coefficient system (5.1) and assume that the eigenvalues of $A_0^{-1}(\sum \xi_j A_j + E)$, which are real by Assumption 5.1, have constant mutliplicity. We denote them by $\lambda_p(\xi)$ and call $\Pi_p(\xi)$ the corresponding eigenprojectors. Then the solution of (5.1) with initial data h is given by

$$u(t, x) = \frac{1}{(2\pi)^d} \sum_p \int e^{i(\varepsilon\xi x - t\lambda_p(\varepsilon\xi))/\varepsilon} \Pi_p(\varepsilon\xi)\hat{h}(\xi) d\xi. \tag{5.2}$$

For oscillating initial data

$$u^\varepsilon{}_{|t=0} = h(x) e^{ikx/\varepsilon} \tag{5.3}$$

the solution is

$$\begin{aligned} u^\varepsilon(t, x) &= \frac{1}{(2\pi)^{d/2}} \sum_p \int e^{i\{x(k+\varepsilon\xi) - t\lambda_k(k+\varepsilon\xi)\}/\varepsilon}\, \Pi_p(k + \varepsilon\xi)\hat{h}(\xi) d\xi \\ &= \sum_p e^{i(kx - t\omega_p)/\varepsilon} a_p^\varepsilon(t, x) \end{aligned} \tag{5.4}$$

with an obvious definition of a_p^ε and $\omega_p = \lambda_p(k)$. Expanding the phases to first order in ε yields

$$\lambda_p(k + \varepsilon\xi) = \lambda(k) + \varepsilon\xi \cdot \mathbf{v}_p + O(\varepsilon^2|\xi|^2), \qquad \mathbf{v}_p = \nabla_\xi \lambda_p(k)$$

and $\Pi_p(k+\varepsilon\xi) = \Pi_p(k) + O(\varepsilon|\xi|)$ so that

$$a_p^\varepsilon(t,x) = a_{p,0}(t,x) + O(\varepsilon t) + O(\varepsilon) \tag{5.5}$$

with

$$\begin{aligned} a_{p,0}(t,x) &= \frac{1}{(2\pi)^{d/2}} \int \mathrm{e}^{\mathrm{i}(x\xi - t\mathbf{v}_k)} \Pi_p(k) \hat{h}(\xi)\, \mathrm{d}\xi \\ &= \Pi_p(k) h(x - t\mathbf{v}_p). \end{aligned}$$

Indeed, using the estimates $\mathrm{e}^{\mathrm{i}tO(\varepsilon|\xi|^2)} - 1 = tO(\varepsilon|\xi|^2)$, $\Pi_p(k+\varepsilon\xi) - \Pi_p(k) = O(\varepsilon|\xi|)$ implies the precise estimate

$$\left\| a^\varepsilon(t,\cdot) - a_{p,0}(t,\cdot) \right\|_{H^s(\mathbb{R}^d)} \le C\,\varepsilon(1+t)\, \|h\|_{H^{s+2}(\mathbb{R}^d)}\,. \tag{5.6}$$

Together with (5.4), this gives an asymptotic description of u^ε with error $O(\varepsilon t)$. The amplitudes $a_{p,0}$ satisfy the *polarization condition*:

$$a_{p,0} = \Pi_p(k) a_{p,0} \tag{5.7}$$

and the simple *transport equation*:

$$(\partial_t + \mathbf{v}_p \cdot \nabla_x) a_{p,0} = 0. \tag{5.8}$$

The analysis just performed can be carried out without fundamental change for initial oscillations with nonlinear phase $\psi(x)$ (see the beginning of the paragraph about caustics below) and also for variable coefficient operators (see e.g. [92,114,65]). The typical results of linear geometric optics are contained in the description (5.4) (5.5), with the basic properties of polarization (5.7) and transport (5.8).

This method based on an explicit writing of the solutions using Fourier synthesis (or more generally Fourier Integral Operators) is limited to linear problems, because no such representation is available for nonlinear problems. On the contrary, the BKW method, which is presented below, is extendable to nonlinear problems and we now concentrate on this approach. Of course, in the linear case we will recover the same properties as those presented in the example above.

5.1.2. The BKW method and formal solutions In the BKW method one looks *a priori* for solutions which have a *phase-amplitude* representation:

$$u^\varepsilon(t,x) = \mathrm{e}^{\mathrm{i}\varphi(t,x)/\varepsilon} \sigma^\varepsilon(t,x), \qquad \sigma^\varepsilon(t,x) \sim \sum_{n\ge 0} \varepsilon^n \sigma_n(t,x). \tag{5.9}$$

We will consider here only *real* phase functions φ. That the parameter in front of L_0 has the same order as the inverse of the wave length, has already been discussed in Section 2: in this scaling, the zeroth order term L_0 comes in the definition of the dispersion relation and in all the aspects of the propagation, as shown in the computations below. Examples are various versions of Maxwell's equations coupled with the Lorentz model, or anharmonic oscillators, or with Bloch's equations.

Introduce the symbol of the equation:

$$\mathcal{L}(a,\tau,\xi)=\tau A_0(a)+\sum_{j=1}^{d}\xi_j A_j(a)-\mathrm{i}E(a)=A_0(a)\,(\tau\mathrm{Id}+\mathcal{G}(a,\xi))\,. \quad (5.10)$$

Plugging the expansion (5.9) into the equation, and ordering in powers of ε, yields the following cascade of equations:

$$\mathcal{L}(a,d_{t,x}\varphi)\sigma_0=0 \quad (5.11)$$

$$\mathrm{i}\mathcal{L}(a,d_{t,x}\varphi)\sigma_{n+1}+L_1(a,\partial)\sigma_n=0,\qquad n\geq 0 \quad (5.12)$$

with $L_1(a,\partial)=A_0\partial_t+\sum A_j\partial_{x_j}$.

DEFINITION 5.2. A formal solution is a formal series (5.9) which satisfies the equation in the sense of formal series, that is, it satisfies the equations (5.11) and (5.12) for all n.

5.1.3. The dispersion relation and phases The first Eq. (5.11) has a nontrivial solution $\sigma_0\neq 0$ if and only if

– φ solves the *eikonal equation*

$$\det\left(\mathcal{L}\left(a(t,x),d\varphi(t,x)\right)\right)=0 \quad (5.13)$$

– σ_0 satisfies the *polarization condition*

$$\sigma_0(t,x)\in\ker\left(\mathcal{L}\left(a(t,x),d\varphi(t,x)\right)\right). \quad (5.14)$$

DEFINITIONS 5.3. (i) The equation $\det\mathcal{L}(a,\tau,\xi)=0$ is called the *dispersion relation*. We denote by $\mathcal{C}$ the set of its solutions (a,τ,ξ).

(ii) A point $(\underline{a},\underline{\tau},\underline{\xi})\in\mathcal{C}$ is called *regular* if, on a neighborhood of this point, $\mathcal{C}$ is given by an equation $\tau+\lambda(a,\xi)=0$, where λ is a smooth function near $(\underline{a},\underline{\xi})$. We denote by $\mathcal{C}_{reg}$ the manifold of regular points.

(iii) Given a C^k function $a(t,x)$, $k\geq 2$, φ is called a *characteristic phase* if is satisfies the eikonal equation (5.13).

(iv) A characteristic phase is said to be of constant multiplicity m if the dimension of $\ker\mathcal{L}\left(a(t,x),d\varphi(t,x)\right)$ is equal to m for all points (t,x).

(v) A characteristic phase is said to be regular if $(a(t,x),d\varphi(t,x))\in\mathcal{C}_{reg}$ for all (t,x).

REMARKS 5.4. (1) A point (a,τ,ξ) belongs to $\mathcal{C}$ if and only if $-\tau$ is an eigenvalue of $\mathcal{G}(a,\xi)$.

(2) Because of the symmetry Assumption 5.1, $\mathcal{G}(a,\xi)$ has only real and semi-simple eigenvalues and the dispersion relation has real coefficients. In particular the kernel and image of $\tau \mathrm{Id} + \mathcal{G}(a,\xi)$ satisfy:

$$\ker(\tau \mathrm{Id} + \mathcal{G}(a,\xi)) \cap \mathrm{im}(\tau \mathrm{Id} + \mathcal{G}(a,\xi)) = \{0\}. \tag{5.15}$$

(3) If $(\underline{a}, \underline{\tau}, \underline{\xi}) \in \mathcal{C}_{reg}$, then for (a,ξ) in a neighborhood of $(\underline{a}, \underline{\xi})$, $\lambda(a,\xi)$ is the only eigenvalue of $\mathcal{G}(a,\xi)$ close to $-\underline{\tau}$, proving that $\lambda(a,\xi)$ has constant multiplicity. In particular, a regular phase has constant multiplicity.

(4) If φ is a characteristic phase of constant multiplicity, then the kernel $\ker \mathcal{L}(a(t,x), d\varphi(t,x))$ and the image $\mathrm{im}\mathcal{L}(a(t,x), d\varphi(t,x))$ are vector bundles of constant dimension, as smooth as $d\varphi$ with respect to (t,x).

EXAMPLES 5.5. (1) (*Planar phases*) If L has constant coefficients, for instance if $a = \underline{a}$ is fixed, then any solution of the dispersion relation, that is any eigenvalue $-\tau$ of $\mathcal{G}(\underline{a},\xi)$, yields a phase

$$\varphi(t,x) = \tau t + \xi \cdot x. \tag{5.16}$$

This phase has constant multiplicity since $\mathcal{L}\left(\underline{a}, d\varphi\right)$ is independent of (t,x).

(2) (*Regular phases*) If $\mathcal{C} = \{\tau + \lambda(a,\xi) = 0\}$ near a regular point, then one can solve, locally, the Hamilton–Jacobi equation

$$\partial_t \varphi + \lambda(a(t,x), \partial_y \varphi) = 0 \tag{5.17}$$

using the method of characteristics (see e.g. [43,65]). The graph of $d\varphi$ is a union of integral curves, called bi-characteristic curves, of the *Hamiltonian field*

$$H = \partial_t + \sum_{j=1}^{d} \partial_{\xi_j} \lambda(a(t,x),\xi)\partial_{x_j} - \sum_{k=0}^{d} \partial_{x_k} \lambda(a(t,x),\xi)\partial_{\xi_k}. \tag{5.18}$$

Starting from an initial phase, $\varphi_0(x)$, one determines the initial manifold $\Lambda_0 = \{0, x, -\lambda(0, x, d_x\varphi_0(x), d_x\varphi_0(x))\}$,and next, the manifold Λ which is the union of the bi-characteristic curves launched from Λ_0. As long as the projection $(t,x,\tau,\xi) \mapsto (t,x)$ from Λ to $\mathbb{R}^{1+d}$ remains invertible, Λ is the graph of the differential of a function φ, that is $\Lambda = \{(t, x, \partial_t\varphi, \partial_x\varphi)\}$, and φ is a solution of (5.17).

The projections on the (t,x) space of the bi-characteristic curves drawn in Λ are called the *rays*. They are integral curves of the field

$$X = \partial_t + \sum_{j=1}^{d} \partial_{\xi_j} \lambda(a(t,x), d_x\varphi(t,x))\partial_{x_j}. \tag{5.19}$$

5.1.4. The propagator of amplitudes The Eq. (5.12) for $n = 1$ reads

$$\mathrm{i}\mathcal{L}(a, d_{t,x}\varphi)\sigma_1 + L_1(a,\partial)\sigma_0 = 0.$$

A necessary and sufficient condition for the existence of solutions is that $L_1(a, \partial)\sigma_0$ belongs to the range of $\mathcal{L}(a, d\varphi)$. Therefore, σ_0 must satisfy:

$$\sigma_0 \in \ker \mathcal{L}(a, d\varphi), \qquad L_1(a, \partial)\sigma_0 \in \mathrm{im}\mathcal{L}(a, d\varphi). \tag{5.20}$$

Suppose that a is smooth and φ is a smooth characteristic phase of constant multiplicity m. In this case $\mathcal{N}(t, x) = \ker \mathcal{L}(a(t, x), d\varphi(t, x))$ and $\mathcal{I}(t, x) = \mathrm{im}\mathcal{L}(a(t, x), d\varphi(t, x))$ define smooth vector bundles $\mathcal{N}$ and $\mathcal{I}$ with fiber of dimension m and $N - m$ respectively. Introduce the quotient bundle $\mathcal{N}' := \mathbb{C}^N/\mathcal{I}$ and the natural projection $\pi : \mathbb{C}^N \mapsto \mathbb{C}^N/\mathcal{I}$. The profile equations (5.20) read

$$\sigma_0 \in C^\infty_{\mathcal{N}}, \qquad L_\varphi \sigma_0 = 0 \tag{5.21}$$

where $C^\infty_{\mathcal{N}}$ denotes the C^∞ sections of the bundle $\mathcal{N}$ and

$$L_\varphi = \pi L_1(a, \partial) \tag{5.22}$$

is a first order system from $C^\infty_{\mathcal{N}}$ to $C^\infty_{\mathcal{N}'}$. We will prove below that the Cauchy problem for (5.21) is well-posed and thus determines σ_0. The operator L_φ, acting from sections of $\mathcal{N}$ to sections of $\mathcal{N}'$, is the *intrinsic formulation of the propagation operator*.

In order to make computations and proofs, it is convenient to use more practical forms of L_φ. We give two of them.

• *Formulation using projectors*. We still assume that φ is a smooth characteristic phase of constant multiplicity m. Thus, there are smooth projectors $P(t, x)$ and $Q(t, x)$ on $\ker \mathcal{L}(a(t, x), d\varphi(t, x))$ and $\mathrm{im}\mathcal{L}(a(t, x), d\varphi(t, x))$, respectively, (see Remarks 5.4 4). They satisfy at each point (t, x):

$$(I - Q)\mathcal{L}(a, d\varphi) = 0, \quad \mathcal{L}(a, d\varphi)P = 0. \tag{5.23}$$

With these notations, the conditions (5.20) are equivalent to

$$\sigma_0(t, x) = P(t, x)\sigma_0(t, x), \tag{5.24}$$

$$(I - Q)L_1(a, \partial)\sigma_0 = 0. \tag{5.25}$$

In this setup, the transport operator reads

$$L_\varphi \sigma := (I - Q)L_1(a, \partial)(P\sigma) = A_{0,\varphi}\partial_t + \sum_{j=1}^{d} A_{j,\varphi}\partial_{x_j} + E_\varphi \tag{5.26}$$

where $A_{j,\varphi} = (\mathrm{Id} - Q)A_j P$ and $E_\varphi = (I - Q)L_1(a, \partial)(P)$.

REMARK 5.6 (*About the choice of projectors*). The choice of the projectors P and Q is completely free. This does not mean a lack of uniqueness in the determination of σ_0 since all the formulations are equivalent to (5.20). Moreover, if P' and Q' are other projectors on $\ker \mathcal{L}$ and $\mathrm{im}\mathcal{L}$, respectively, there are matrices α and β such that $P' = P\alpha$ and $(\mathrm{Id} - Q') = \beta(\mathrm{Id} - Q)$, since P and P' have the same image and $(\mathrm{Id} - Q)$ and $(\mathrm{Id} - Q')$ have the same kernel. Therefore the operators $L'_\varphi = (\mathrm{Id} - Q')A_j P'$ and L_φ are obviously conjugated and thus share the same properties.

Of course, the symmetry assumption suggests **a natural choice**: one can take $P(t,x)$ to be the spectral projector on $\ker(\partial_t\varphi + \mathcal{G}(a, d_x\varphi))$, that is the projector on the kernel along the image of $\partial_t\varphi + \mathcal{G}(a, d_x\varphi)$. By the symmetry assumption, it is an orthogonal pojector for the scalar product induced by $A_0(a(t,x))$ so that $A_0 P = P^* A_0$, and the natural associated projector on the image is $Q = A_0(\mathrm{Id} - P)A_0^{-1} = \mathrm{Id} - P^*$. This choice is natural and sufficient for a mathematical analysis and the reader can assume that this choice is made all along theses notes.

• *Formulation using parametrization of the bundles*. Alternately, one can parametrize (at least locally) the kernel as

$$\ker \mathcal{L}(a(t,x), d\varphi(t,x)) = \rho(t,x)\mathbb{R}^m \tag{5.27}$$

where $\rho(t,x)$ is a smooth and injective $N \times m$ matrix. For instance, in the case of Maxwell's equations, the kernel can be paramametrized by the E component (see Example 5.11 below). Similarly, one can parametrize the co-kernel of $\mathcal{L}(a, d\varphi)$ and introduce a smooth $m \times N$ matrix ℓ of rank m such that

$$\ell(t,x)\mathrm{im}\mathcal{L}(a(t,x), d\varphi(t,x)) = \{0\}. \tag{5.28}$$

With these notations, the conditions (5.20) are equivalent to

$$\sigma_0 = \rho\sigma^\flat, \tag{5.29}$$

$$\ell L_1(a, \partial)\rho\sigma^\flat = 0. \tag{5.30}$$

For instance, $\mathcal{L}(a, d\varphi)$ being self-adjoint, one can choose $\ell = \rho^*$. In this setup, the transport operator reads

$$L^\flat\sigma^\flat := \ell L_1(a, \partial)(\rho\sigma^\flat) = A_0^\flat\partial_t + \sum_{j=1}^{d} A_j^\flat\partial_{x_j} + E^\flat \tag{5.31}$$

where $A_j^\flat = \ell A_j \rho$ and $E^\flat = \ell L_1(a, \partial)(\rho)$. It is clear that changing ρ [resp. ℓ] is just a linear change of unknowns [resp. change of variables in the target space].

For instance, let $(r_1, \ldots, r_m)$ be a smooth basis of $\ker \mathcal{L}(a, d\varphi)$ (the columns of the matrix ρ) and let $(\ell_1, \ldots, \ell_m)$ be a smooth dual basis of left null vectors of $\mathcal{L}(a, d\varphi)$ (the rows of ℓ). They can be normalized so that

$$\ell_p(t,x)A_0(a(t,x)r_q(t,x)) = \delta_{p,q}, \qquad 1 \le p, q \le m. \tag{5.32}$$

(One can choose the $\{r_p\}$ forming an orthonormal basis for the Hermitian scalar product induced by A_0 and thus $\ell_p = r_p^*$). The polarization condition (5.29) reads

$$\sigma(t,x) = \sum_{p=1}^{m} \sigma_p(t,x) r_p(t,x), \tag{5.33}$$

with scalar functions σ_p. The equation (5.30) is

$$\partial_t \sigma_p(t,x) + \sum_{j=1}^{d} \sum_{q=1}^{m} \ell_p(t,x) \partial_{x_j} \big(A_j(a(t,x)) \sigma_q(t,x) r_q(t,x) \big) = \ell_p f. \tag{5.34}$$

With $\vec{\sigma}$ denoting the column m-vector with entries σ_p, the operator in the left-hand side reads

$$L^\flat \vec{\sigma} := \partial_t \vec{\sigma} + \sum_{j=1}^{d} A_j^\flat \partial_{x_j} \vec{\sigma} + E^\flat \vec{\sigma} \tag{5.35}$$

where $A_j^\flat$ is the $m \times m$ matrix with entries $\ell_p A_j r_q$.

LEMMA 5.7 (*Hyperbolicity of the propagation operator*). *Suppose that φ is a smooth characteristic phase of constant multiplicity. Then the first order system L_φ on the smooth fiber bundle $\mathcal{N} = \ker \mathcal{L}(a, d\varphi)$ is symmetric hyperbolic, in the sense that*

(i) *for any choice of projectors P and Q as above, there is a smooth matrix $S_\varphi(t,x)$ such that the matrices $S_\varphi A_{j,\varphi}$ occurring in (5.26) are self-adjoint and $S_\varphi A_{0,\varphi}$ is positive definite on $\mathcal{N}$,*
or equivalently,

(ii) *for any choice of matrices ρ and ℓ, the $m \times m$ system $L^\flat$ (5.31) is symmetric hyperbolic.*

PROOF. Using the Remark 5.6 it is sufficient to make the proof when P is the spectral projector and $(\mathrm{Id} - Q) = P^*$, in which case the result is immediate with $S = \mathrm{Id}$.

For a general direct proof, note that $f \in \ker(\mathrm{Id} - Q)$ when $f \in \mathrm{im}\mathcal{L}(a, d\varphi)$. Because of symmetry, $\mathrm{im}\mathcal{L}(a, d\varphi) = \ker \mathcal{L}(a, d\varphi)^\perp$. Therefore, this space is equal to $(\mathrm{im} P)^\perp = \ker P^*$. This shows that $\ker(\mathrm{Id} - Q) = \ker P^*$, implying that there is a smooth matrix S such that

$$S(\mathrm{Id} - Q) = P^*.$$

Therefore $S A_{j,\varphi} = P^* A_j P$ is symmetric and $S A_{0,\varphi} P = P^* A_0 P$ is positive definite on $\mathcal{N}$.

The proof for the $L^\flat$ representation is quite similar. □

The classical existence theory for symmetric hyperbolic systems can be transported to vector bundles, for instance using the existence theory for $L^\flat$. It can also be localized on domains of determinacy as sketched in Section 3.6. In the remaining part of these notes we generally use the formulation (5.26) of the equations.

THEOREM 5.8. *Suppose that $a \in C^\infty$ and that φ is a C^∞ characteristic phase of constant multiplicity near Ω of $(0, \underline{x})$. Given a neighborhood ω of $\underline{x}$ in $\mathbb{R}^d$, then, shrinking Ω if necessary, for all C^∞ section h over ω of $\ker \mathcal{L}(a(0,x), d\varphi(0,x))$ there is a unique solution $\sigma \in C^\infty(\Omega)$ of*

$$\sigma \in \ker \mathcal{L}(a, d\varphi), \qquad L_1(a, \partial)\sigma \in \mathrm{im}\mathcal{L}(a, d\varphi), \qquad \sigma_{|t=0} = h. \tag{5.36}$$

Equivalently, the equations can be written

$$\sigma = P\sigma, \qquad (\mathrm{Id} - Q)L_1(a, \partial)P\sigma = 0, \qquad P\sigma_{|t=0} = h \tag{5.37}$$

for *any* set of projectors P and Q.

LEMMA 5.9 (*Transport equations for regular phases*). *If φ is a regular phase associated with the eigenvalue $\lambda(a, \xi)$, then the principal part of the operator L_φ is the transport operator*

$$A_{0,\varphi} X(t, x, \partial_{t,x}) \tag{5.38}$$

where $A_{0,\varphi} = (I - Q)A_0 P$ and $X = \partial_t + \mathbf{v}_g \cdot \partial_x$ is the ray propagator (5.19).

DEFINITION 5.10 (*Group velocity*). Under the assumptions of the previous lemma, $\mathbf{v}_g = \nabla_\xi \lambda(a, \partial_x \varphi)$ is called the group velocity. It is constant when $a = \underline{a}$ is constant and a is a planar phase.

PROOF. Near a regular point $(\underline{a}, \underline{\tau}, \underline{\xi})$ there are smooth projectors $\mathcal{P}(a, \xi)$ and $\mathcal{Q}(a, \xi)$ on $\ker \mathcal{L}(a, -\lambda(a, \xi), \xi)$ and $\mathrm{im}\mathcal{L}(a, -\overline{\lambda}(a, \xi), \xi)$, respectively. Differentiating the identity

$$\Big(-\lambda(a, \xi)A_0(a) + \sum \xi_j A_j(a) - \mathrm{i}E(a)\Big)\mathcal{P}(a, \xi) = 0 \tag{5.39}$$

with respect to ξ and multiplying on the left by $(I - \mathcal{Q})$ implies that

$$-\partial_{\xi_j}\lambda(I - \mathcal{Q})A_0\mathcal{P} + (I - \mathcal{Q})A_j\mathcal{P} = 0. \tag{5.40}$$

Evaluating at $\xi = d_x\varphi$ implies that

$$(I - Q)A_j P = -\partial_{\xi_j}\lambda(a, d_x\varphi)(I - Q)A_0 P, \tag{5.41}$$

that is (5.38). □

EXAMPLE 5.11 (*Maxwell–Lorentz equations*). Consider the system

$$\partial_t B + \mathrm{curl}\, E = 0, \quad \partial_t E - \mathrm{curl}\, B = -\partial_t P, \quad \varepsilon^2\partial_t^2 P + P = \gamma E \tag{5.42}$$

and an optical planar phase $\varphi = \omega t + kx$ which satisfies the dispersion relation

$$|k|^2 = \omega^2 \left(1 + \frac{\gamma}{1-\omega^2}\right) := \mu^2(\omega) \tag{5.43}$$

(see (3.15)). The polarization conditions are

$$E \in k^\perp, \qquad B = -\frac{1}{\omega} k \times E, \qquad P = \frac{\gamma E}{1-\omega^2} \tag{5.44}$$

(see (3.16)). Thus the kernel $\ker \mathcal{L}$ is parametrized by $E \in k^\perp$ and the transport equation reads

$$\mu'(\omega)\partial_t E - \frac{k}{|k|} \cdot \partial_x E = 0. \tag{5.45}$$

In the general case, the characteristic determinant of $\mathcal{L}_\varphi$ can be related to the Taylor expansion of the dispersion relation:

PROPOSITION 5.12. *Suppose that φ is a characteristic phase with constant multiplicity m. Then the polynomial in (τ, ξ)*

$$\det \mathcal{L}(a(t,x), \partial_t\varphi(t,x) + \tau, \partial_x\varphi(t,x) + \xi) \tag{5.46}$$

vanishes at order m at the origin, and its homogeneous part of degree m is proportional to the characteristic determinant of L_φ.

PROOF. Fix (t,x). For small ξ, there is a smooth decomposition of $\mathbb{C}^N = \mathbb{E}_0(\xi) + \mathbb{E}_1(\xi)$, into invariant spaces $\mathbb{E}_l$ of $\mathcal{G}(a, \partial_x\varphi + \xi)$, such that $\mathbb{E}_0(0) = \ker(\partial_t\varphi + \mathcal{G}(a, \partial_x\varphi))$. In bases $\{r_p(\xi)\}_{1\le p\le m}$ for $\mathbb{E}_0$ and $\{r_p(\xi)\}_{m+1\le p\le N}$ for $\mathbb{E}_1$, $\mathcal{G}(a, \partial_x\varphi+\xi)$ has a block diagonal form:

$$\partial_t\varphi I + \mathcal{G}(a, \partial_x\varphi + \xi) = \begin{pmatrix} \mathcal{G}_0 & 0 \\ 0 & \mathcal{G}_1 \end{pmatrix}$$

with $\mathcal{G}_0 = 0$ and $\mathcal{G}_1$ invertible at $\xi = 0$. Therefore

$$\det \mathcal{L}(a, \partial_{t,x}\varphi + (\tau,\xi)) = c(\tau,\xi) \det(\tau I + \mathcal{G}_0(\xi))$$

where $c(0,0) \neq 0$. Let $\{\psi_p\}$ denote the dual basis of $\{r_p\}$. The entries of $\mathcal{G}_0$ are

$$\begin{aligned}\psi_p(\xi)\,(\partial_t\varphi I + \mathcal{G}(a, \partial_x\varphi + \xi))\, r_q(\xi) &= \psi_p(0)\mathcal{G}(a,\xi) r_q(0) + O(|\xi|^2) \\ &= \sum_j \xi_j \psi_p(0) A_j(a) r_q(0) + O(|\xi|^2)\end{aligned}$$

for $1 \le p, q \le m$. At $\xi = 0$, $\{r_1, \ldots, r_m\}$ is a basis of $\ker \mathcal{L}(a, d\varphi)$ and $\{\ell_p\} = \{\psi_p(A_0)^{-1}\}$ is a basis of left null vectors of $\ker \mathcal{L}(a, d\varphi)$ which satisfies (5.32). Therefore the $\psi_p(0)A_j r_q(0)$ are the entries of the matrix $A_j^\flat$ introduced in (5.35) and hence

$$\det(\tau I + \mathcal{G}_0(\xi)) = \det\left(\tau I + \sum \xi_j A_j^\flat\right) + O\left((|\tau| + |\xi|)^{m+1}\right).$$

The proposition follows. □

REMARK 5.13. When the phase is regular of multiplicity m, Lemma 5.9 asserts that the first order part of the propagator is $L^\flat = XI$, whose characteristic determinant is

$$\det L^\flat(\tau, \xi) = \left(\tau + \sum \xi_j \partial_{\xi_j} \lambda(a, \partial_x \varphi)\right)^m.$$

Therefore, we recover that

$$\begin{aligned}\det \mathcal{L}(a, \partial_{t,x}\varphi + (\tau, \xi)) &= c(\tau, \xi)\,(\tau + \partial_t \varphi + \lambda(a, \partial_x \varphi + \xi))^m \\ &= c \det L^\flat(\tau, \xi) + O\left((|\tau| + |\xi|)^{m+1}\right).\end{aligned}$$

From a geometrical point of view, this means that the characteristic manifold of $L^\flat$ is the tangent space of $\mathcal{C}_{reg}$ at $(a, d\varphi)$. In general, the geometrical meaning of Proposition 5.12 is that the characteristic manifold of $L^\flat$ is the tangent cone of $\mathcal{C}$ at $(a, d\varphi)$.

EXAMPLE 5.14 (*Conical refraction*). Consider Maxwell's equations in a bi-axial crystal. With notations as in see (3.19), the characteristic determinant satisfies

$$\det \mathcal{L}(\tau, \xi) = \frac{1}{4}\tau^2((2\tau^2 - \Psi)^2 - (P^2 + Q^2)).$$

There are exceptional points in the characteristic variety $\mathcal{C}$ which are *not* regular: this happens when $P^2 + Q = 0$ and $\tau = \pm\frac{1}{2}\Psi$, whose explicit solution is given in (3.20). Consider such a point $(\omega, k) \in \mathcal{C} \setminus \mathcal{C}_{reg}$ and the planar phase $\varphi = \omega t + kx$, which has constant multiplicity $m = 2$, see Examples 5.5 (1). Near this point,

$$\begin{aligned}\det \mathcal{L}(\partial_{t,x}\varphi + (\tau, \xi)) = c((4\tau - \xi \cdot \Psi'(k))^2 \\ - (\xi \cdot P'(k))^2 - (\xi \cdot Q'(k))^2) + \text{h.o.t.}\end{aligned}$$

Thus the relation dispersion of L_φ is

$$\left(4\tau - \xi \cdot \Psi'(k)\right)^2 - (\xi \cdot P'(k))^2 - (\xi \cdot Q'(k))^2 = 0.$$

Because $P'(k)$ and $Q'(k)$ are linearly independent, this cannot be factored and this is the dispersion relation of a *wave* equation, not of a transport equation. We refer the reader to [102,76], for instance, for more details.

5.1.5. Construction of WKB solutions

THEOREM 5.15 (*WKB solutions*). *Suppose that $a \in C^\infty$ and that φ is a C^∞ characteristic phase of constant multiplicity m on a neighborhood Ω of $(0, \underline{x})$. Given a neighborhood ω of $\underline{x}$, shrinking Ω is necessary, for all sequence of functions $h_n \in C^\infty(\omega)$ satisfying*

$P(0,x)h_n = h_n$, there is a unique sequence of functions $\sigma_n \in C^\infty(\Omega; \mathbb{C}^N)$ which satisfies (5.11) (5.12) and the initial conditions:

$$P\sigma_{n|t=0} = h_n. \tag{5.47}$$

In addition, $\sigma_0 = P\sigma_0$ is polarized.

SKETCH OF PROOF. The equation (5.12) is of the form

$$\mathcal{L}(a(t,x), d\varphi(t,x))\sigma = f. \tag{5.48}$$

A necessary condition for the existence of σ is that $f \in \mathrm{im}\mathcal{L}(a, d\varphi)$, and equivalently, that $(\mathrm{Id} - Q)f = 0$ where Q is a smooth projector on the range of $\mathcal{L}(a, d\varphi)$. When this condition is satisfied, this equation determines σ up to an element of the kernel. More precisely, the class $\tilde{\sigma}$ of σ in the quotient space $\mathbb{C}^N / \ker \mathcal{L}(a, d\varphi)$ is

$$\tilde{\sigma} = (\widetilde{\mathcal{L}}(a, d\varphi))^{-1} f \tag{5.49}$$

where $\widetilde{\mathcal{L}}$ is the natural isomorphism from $\mathbb{C}^N / \ker \mathcal{L}(a, d\varphi)$ to $\mathrm{im}\mathcal{L}(a, d\varphi)$ induced by $\mathcal{L}(a, d\varphi)$.

This relation can be lifted to $\mathbb{C}^N$, introducing a partial inverse of $\mathcal{L}(a, d\varphi)$. Given projectors P and Q on $\ker \mathcal{L}(a, d\varphi)$ and $\mathrm{im}\mathcal{L}(a, d\varphi)$ respectively, there is a unique partial inverse $R(t,x)$ such that for all (t,x):

$$R\mathcal{L}(a, d\varphi) = I - P, \qquad PR = 0, \qquad R(I - Q) = 0. \tag{5.50}$$

In particular, (5.48) is equivalent to

$$(\mathrm{Id} - Q)f = 0, \qquad \sigma = Rf + P\sigma. \tag{5.51}$$

Using these notations, the cascade of equations (5.11) (5.12) is equivalent to

$$P\sigma_0 = \sigma_0, \tag{5.52}$$

$$(I - Q)L_1(a, \partial)\sigma_n = 0, \quad n \geq 0, \tag{5.53}$$

$$\sigma_{n+1} = \mathrm{i}RL_1(a, \partial)\sigma_n + P\sigma_{n+1}, \quad n \geq 0 \tag{5.54}$$

and thus to

$$P\sigma_0 = \sigma_0, \qquad (I - Q)L_1(a, \partial)P\sigma_0 = 0, \tag{5.55}$$

and for $n \geq 1$:

$$(I - Q)L_1(a, \partial)P\sigma_n = -\mathrm{i}(I - Q)L_1(a, D)RL_1(a, \partial)\sigma_{n-1}, \tag{5.56}$$

$$(I - P)\sigma_n = \mathrm{i}RL_1(a, \partial)\sigma_{n-1}. \tag{5.57}$$

By (5.59), The Cauchy problems (5.55) and (5.56), with unknowns $P\sigma_n$, can be solved in a neighborhood of $(0, \underline{x})$ using Theorem 5.8. □

The initial conditions (5.47) can be replaced by conditions of the form

$$H\sigma_{n\,|t=0} = h_n \tag{5.58}$$

where H is an $m \times N$ matrix, which depends smoothly on x, and such that

$$\ker H(x) \cap \ker \mathcal{L}(a(0, x), d\varphi(0, x)) = \{0\}. \tag{5.59}$$

THEOREM 5.16. *Suppose that $a \in C^\infty$ and that φ is a C^∞ characteristic phase of constant multiplicity m on a neighborhood Ω of $(0, \underline{x})$. Given a neighborhood ω of $\underline{x}$, shrinking Ω is necessary, for all sequences of functions $h_n \in C^\infty(\omega; \mathbb{C}^m)$, there is a unique sequence of solutions $\sigma_n \in C^\infty(\Omega; \mathbb{C}^N)$ of (5.11) (5.12) and the initial conditions (5.58).*

Indeed, the nth initial condition reads

$$HP\sigma_{n\,|t=0} = h_n - H(\mathrm{Id} - P)\sigma_{n\,|t=0} = h_n - \mathrm{i}HRL_1\sigma_{n-1\,|t=0}.$$

Knowing σ_{n-1}, the assumption (5.59) implies that this equation can be solved and give the desired initial conditions for $P\sigma_{n\,|t=0}$.

REMARK 5.17 (*About the choice of projectors*). The projector P intervenes in the formulation of the Cauchy condition (5.47). We point out here that the coefficients σ_n do not depend on splitting the (5.53) (5.54) equations, where one could use as well other projectors P' and Q': the equations are always equivalent to (5.11) (5.12) and the theorem asserts uniqueness. Moreover, the set of asymptotic solutions obtained by this construction does not depend on the projector P used for the initial condition.

REMARK 5.18. Imposing initial conditions for the $P\sigma_n$ is natural from the proof. We follow this approach in the remaining part of these notes. However, in applications, the formulation (5.58) may be better adapted to physical considerations. For instance, when considering Maxwell's equations as in (5.42), it makes sense to impose initial conditions on the electric field, or on the electric induction (see Example 5.11).

5.1.6. Approximate solutions Given $(\sigma_0, \dots, \sigma_n) \in C^\infty$ solutions of (5.11) (5.12) for $0 \le k \le n$, the family of functions

$$u^\varepsilon_{\mathrm{app},n}(t, x) = \sum_{k=0}^{n} \varepsilon^k \sigma_k(t, x) \mathrm{e}^{\mathrm{i}\varphi(t,x)/\varepsilon} \tag{5.60}$$

satisfies the Eq. (5.1) with an error term of order $O(\varepsilon^n)$, that is

$$L(a, \partial)u^\varepsilon_{\mathrm{app},n} = \varepsilon^n f_n \mathrm{e}^{\mathrm{i}\varphi/\varepsilon} \tag{5.61}$$

where $f_n = QL_1(a, \partial)\sigma_n$ is smooth.

If $\{\sigma_n\}_{n\in\mathbb{N}}$ is a family of solutions of (5.11) (5.12), one can use Borel's Theorem to construct asymptotic solutions

$$u^\varepsilon_{\mathrm{app}}(t,x)=\sigma^\varepsilon(t,x)\mathrm{e}^{\mathrm{i}\varphi(t,x)/\varepsilon},\qquad \sigma^\varepsilon(t,x)\sim\sum_{k\geq 0}\varepsilon^k\sigma_k(t,x) \tag{5.62}$$

where the symbol $\sim$ means that for all n and all m:

$$\sigma^\varepsilon-\sum_{k\leq n}\varepsilon^k\sigma_k=O(\varepsilon^{n+1})\quad\text{in } C^m. \tag{5.63}$$

In this case

$$L(a,\partial)u^\varepsilon_{\mathrm{app}}=r^\varepsilon\mathrm{e}^{\mathrm{i}\varphi/\varepsilon} \tag{5.64}$$

where $r^\varepsilon=O(\varepsilon^\infty)$ in C^∞, meaning that for all n and all m,

$$r^\varepsilon=O(\varepsilon^n)\quad\text{in } C^m. \tag{5.65}$$

Note that this also implies that

$$r^\varepsilon\mathrm{e}^{\mathrm{i}\varphi/\varepsilon}=O(\varepsilon^\infty)\quad\text{in } C^\infty. \tag{5.66}$$

5.1.7. Exact solutions The construction of exact solutions u^ε of $L(a,\partial)u^\varepsilon=0$ close to the approximate solutions $u^\varepsilon_{\mathrm{app}}$ amounts to solving the equation for the difference

$$L(a,\partial)(u^\varepsilon-u^\varepsilon_{\mathrm{app}})=-r^\varepsilon\mathrm{e}^{\mathrm{i}\varphi/\varepsilon}. \tag{5.67}$$

Because this system is symmetric hyperbolic and linear, the Cauchy problem is locally well-posed. The question is how to obtain estimates for the difference $u^\varepsilon-u^\varepsilon_{\mathrm{app}}$, which are uniform in ε. The symmetry immediately implies uniform L^2 estimates. The proof of uniform Sobolev estimates is more delicate when the term $\varepsilon^{-1}E(a)$ is present in the Eq. (5.1), as discussed in Section 4. In all cases, the method of weighted estimates using H^s_ε norms can be applied. Combining Theorems 5.15 and 4.21, localized on domains of determinacy as in Section 3.6, implies the following.

THEOREM 5.19. *Suppose that $a\in C^\infty$ and that φ is a C^∞ characteristic phase of constant multiplicity near $(0,\underline{x})$, and suppose that $\{h_n\}_{n\in\mathbb{N}}$ is a sequence of C^∞ functions on a fixed neighborhood of $\underline{x}$, such that $P(0,x)h_n(x)=h_n(x)$. Then there are C^∞ solutions u^ε of $L(a,\partial)u^\varepsilon=0$ in a neighborhood of $(0,\underline{x})$, independent of ε, such that*

$$u^\varepsilon=\sigma^\varepsilon\mathrm{e}^{\mathrm{i}\varphi/\varepsilon},\qquad \sigma^\varepsilon\sim\sum_{n\geq 0}\varepsilon^n\sigma_n$$

where the $\{\sigma_n\}_{n\in\mathbb{N}}$ *are the unique solutions of* (5.11) (5.12) *such that* $P\sigma_{n\,|t=0} = h_n$. *In particular*

$$Pu^\varepsilon{}_{|t=0} = h^\varepsilon e^{i\varphi(0,\,\cdot\,)/\varepsilon}, \qquad h^\varepsilon \sim \sum_{n\geq 0} \varepsilon^n h_n.$$

5.2. *Weakly nonlinear geometric optics*

To fix the notations, consider here the first order system

$$\begin{aligned} L(a,u,\partial)u &:= \left(\varepsilon A_0(a,u)\partial_t + \sum_{j=1}^{d} \varepsilon A_j(a,u)\partial_{x_j} + E(a)\right) u \\ &= f(a,u), \end{aligned} \tag{5.68}$$

where f is a smooth function with $f(a,0) = 0, \nabla_u f(a,0) = 0$. The A_j are smooth and symmetric matrices and E is skew symmetric. To model high frequency oscillatory nonlinear waves, one looks for solutions of the form

$$u^\varepsilon(t,x) \sim \varepsilon^p \sum_{n\geq 0} \varepsilon^n U_n(t,x,\varphi(t,x)/\varepsilon). \tag{5.69}$$

There are two differences with (5.9):

– because of the nonlinearity, *harmonics* $e^{ik\varphi/\varepsilon}$ are expected, yielding to general periodic functions of φ/ε and thus profiles $U_n(t,x,\theta)$ that are periodic in θ, see Section 7 for an elementary example.
– a prefactor ε^p, which measures the amplitude of the oscillations. When p is large, the wave is small and driven by the linear part $L(a,0,\partial)$, the nonlinear effects appearing only as perturbations of the principal term. Decreasing p, the *weakly nonlinear regime* is reached when the nonlinear effects are present in the propagation of the principal term U_0.

For general quasi-linear equations (5.68) and quadratic nonlinearities, this corresponds to $p = 1$. When the nonlinearities are cubic, then the natural scaling is $p = \frac{1}{2}$ (see the general discussion in [41]). In the remaining part of this section, we mainly concentrate on the most general case with $p = 1$ (see however the third example in Remarks 5.26).

Other regimes of *strongly nonlinear geometric optics* will be briefly discussed in the subsequent sections.

REMARK 5.20 (*Other frameworks*). In the framework presented here, $u_0 = 0$ is a solution of (5.68) and we study perturbations of this particular solution. The analysis applies as well to slightly different equations. For instance, one can replace the condition $f = O(|u|^2)$ by $f = \varepsilon \tilde{f}$ and consider perturbations of nonconstant solutions. Examples are systems

$$L(u,\partial)u := \left(A_0(u)\partial_t + \sum_{j=1}^{d} A_j(u)\partial_{x_j} + \frac{1}{\varepsilon}E\right) u = f(u). \tag{5.70}$$

Note that the prefactor ε^p has to be adapted to this setting. We leave it to the reader to write down the corresponding modifications. Some of the examples below concern this class of equations.

5.2.1. Asymptotic equations With $p = 1$, plug the expansion (5.69) into the equation, expand in a power series of ε and equate to zero the coefficients. We obtain a cascade of equations:

$$\mathcal{L}_0(a, d\varphi, \partial_\theta)U_0 = 0, \tag{5.71}$$

$$\mathcal{L}_0(a, d\varphi, \partial_\theta)U_{n+1} + \mathcal{L}_1(a, U_0, \partial_{t,x,\theta})U_n = F_n, \qquad n \geq 0, \tag{5.72}$$

where

$$\mathcal{L}_0(a, d\varphi, \partial_\theta) = \left(\sum_{j=0}^{d} \partial_j \varphi A_j(a, 0)\right) \partial_\theta + E(a),$$

$$\mathcal{L}_1(a, v, \partial_{t,x,\theta}) = L_1(a, 0, \partial_{t,x}) + B(a, d\varphi, v)\partial_\theta,$$

with

$$L_1(a, 0, \partial_{t,x}) = \sum_{j=0}^{d} A_j(a, 0)\partial_j, \qquad B(a, d\varphi, v) = \sum_{j=0}^{d} \partial_j \varphi \, v \cdot \nabla_u A_j(a, 0).$$

Moreover,

$$F_0 = \nabla^2 f(a, 0)(U_0, U_0),$$

$$F_n = 2\nabla^2 f(a, 0)(U_0, U_n) - \sum_{j=0}^{d} \partial_j \varphi \, U_n \cdot \nabla_u A_j(a, 0)\partial_\theta U_0 + G_{n-1}$$

where G_{n-1} depends only on $(U_0, \ldots, U_{n-1})$.

The operator $\mathcal{L}_0$ has constant coefficients in θ. Using Fourier series,

$$U(t, x, \theta) = \sum_{\alpha \in \mathbb{Z}} \hat{U}_\alpha(t, x) e^{i\alpha\theta}, \tag{5.73}$$

$$\mathcal{L}_0(a, d\varphi, \partial_\theta)U = \sum_{\alpha} i\mathcal{L}(a, \alpha d\varphi)\hat{U}_\alpha(t, x) e^{i\alpha\theta}$$

where $\mathcal{L}(a, \tau, \xi)$ is the symbol (5.10) associated with the operator $L(a, 0, \partial)$. For $\mathcal{L}_0$ to have a nontrivial kernel, at least one phase $\alpha\varphi$ must be characteristic. Changing φ is necessary, we assume that this occurs for $\alpha = 1$. Next, the analysis is quite different depending on whether or not $E = 0$. In the former case, all the harmonics are characteristic; in the latter, the harmonics are not expected to be characteristic in general, except $\alpha = -1$ if one considers real-valued solutions. The following assumptions are satisfied in virtually all cases of application. They give a convenient framework for the construction of asymptotic solutions.

ASSUMPTION 5.21. $\varphi \in C^\infty(\overline{\Omega})$ is a characteristic phase of constant multiplicity for $L(a, 0, \partial)$ on a neighborhood Ω of $(0, \underline{x})$.

If $E \neq 0$, we further assume that $\det(\sum \partial_j \varphi A_j)$ does not vanish on Ω and that for all α, either $\alpha\varphi$ is a characteristic phase of constant multiplicity for $L(a, 0, \partial)$, or $\det \mathcal{L}(a, 0, \alpha d\varphi)$ does not vanish on $\overline{\Omega}$.

EXAMPLE 5.22. In the constant coefficient case, $a(t, x) = \underline{a}$ constant, for planar phase $\varphi = \tau t + \xi x$, the conditions when $E \neq 0$ read

$$\det\Big(\tau A_0 + \sum \xi_j A_j - \mathrm{i}E\Big) = 0, \qquad \det\Big(\tau A_0 + \sum \xi_j A_j\Big) \neq 0. \tag{5.74}$$

Denote by $Z \subset \mathbb{Z}$ the set of indices α such that $\alpha\varphi$ is a characteristic phase. If $E = 0$, then $Z = \mathbb{Z}$ and by homogeneity, $\mathcal{L}(a, 0, \alpha d\varphi) = \alpha \mathcal{L}(a, 0, d\varphi)$. On the other hand, if $E \neq 0$, by continuity, the assumption above implies that $\det(\sum \partial_j \varphi A_j - \mathrm{i}\frac{1}{\alpha} E) \neq 0$ on $\overline{\Omega}$ if $|\alpha|$ is large. Therefore the set $Z \subset \mathbb{Z}$ of indices α, such that $\alpha\varphi$ is a characteristic phase, is finite. In both cases, there are projectors P_α, Q_α and partial inverse R_α such that for all $\alpha \in \mathbb{Z}$.

$$\begin{aligned} &(I - Q_\alpha)\mathcal{L}(a, 0, \alpha d\varphi) = 0, \qquad \mathcal{L}(a, 0, \alpha d\varphi) P_\alpha = 0, \\ &R_\alpha \mathcal{L}(a, 0, \alpha d\varphi) = I - P_\alpha, \qquad P_\alpha R_\alpha = 0, \qquad R_\alpha (I - Q_\alpha) = 0. \end{aligned}$$

In particular, if $\alpha \notin Z$, $P_\alpha = 0$, $Q_\alpha = \mathrm{Id}$ and $R_\alpha = (\mathcal{L}(a, 0, \alpha d\varphi))^{-1}$.

REMARK 5.23. When $E = 0$, then for $\alpha = 0$, one has $P_0 = \mathrm{Id}$, $Q_0 = 0$ and $R_0 = 0$. When $\alpha \neq 0$, by homogeneity one can choose $P_\alpha = P_1$, $Q_\alpha = Q_1$ and $R_\alpha = \alpha^{-1} R_1$. This is systematically assumed in the exposition below.

In both cases (Z finite or $E = 0$ with the particular choice above), Assumption 5.21 implies that the matrices P_α, Q_α and R_α are uniformly bounded for $(t, x) \in \Omega$ and $\alpha \in \mathbb{Z}$. Thus one can define projectors $\mathcal{P}$ and $\mathcal{Q}$ on the kernel and on the image of $\mathcal{L}_0(a, \partial_\theta)$, respectively, and a partial inverse $\mathcal{R}$. In the Fourier expansion (5.73), $\mathcal{P}$ is defined by the relation

$$\mathcal{P}U = \sum_\alpha P_\alpha \hat{U}_\alpha(t, x) \mathrm{e}^{\mathrm{i}\alpha\theta} \tag{5.75}$$

with similar definitions for $\mathcal{Q}$ and $\mathcal{R}$.

With these notations, the cascade (5.71) (5.72) is analyzed as in the linear case. It is equivalent to

$$\mathcal{P}U_0 = U_0, \qquad (I - \mathcal{Q})\mathcal{L}_1(a, U_0, \partial)\mathcal{P}U_0 = (I - \mathcal{Q})F_0, \tag{5.76}$$

and for $n \geq 1$:

$$(I - \mathcal{Q})\mathcal{L}_1(a, U_0, \partial)\mathcal{P}U_n = (I - \mathcal{Q})\left(F_n - \mathcal{L}_1(a, U_0, D)(I - \mathcal{P})U_n\right), \tag{5.77}$$

$$(I - \mathcal{P})U_n = \mathcal{R}\left(F_{n-1} - \mathcal{L}_1(a, \partial)U_{n-1}\right). \tag{5.78}$$

REMARK 5.24. As in Remark 5.6, the choice of projectors P_α, Q_α is free. The natural choice is to take spectral projectors, that is $P_\alpha(t,x)$ orthogonal with respect to its scalar product defined by $A_0(a(t,x))$ and $Q_\alpha = A_0(\mathrm{Id} - P_\alpha)A_0^{-1} = \mathrm{Id} - P_\alpha^*$.

5.2.2. The structure of the profile equations I: the dispersive case We consider here the case where $E \neq 0$. In this case the set Z of characteristic harmonics is finite, and the polarization condition in (5.76) reads

$$U_0(t,x,\theta) = \sum_{\alpha\in Z} \hat{U}_{0,\alpha}(t,x)\mathrm{e}^{\mathrm{i}\alpha\theta}, \qquad P_\alpha \hat{U}_{0,\alpha} = \hat{U}_{0,\alpha} \tag{5.79}$$

and the equation is a coupled system for the $\{\hat{U}_{0,\alpha}\}_{\alpha\in Z}$:

$$(I - Q_\alpha)L_1(a,0,\partial)P_\alpha\hat{U}_{0,\alpha} = (I - Q_\alpha)\sum_{\beta+\gamma=\alpha} \hat{\Gamma}_{\alpha,\beta,\gamma}\left(\hat{U}_{0,\beta},\hat{U}_{0,\gamma}\right) \tag{5.80}$$

with

$$\hat{\Gamma}_{\alpha,\beta,\gamma}(U,V) = -\mathrm{i}\gamma B(a,d\varphi,U)V + \nabla^2 f(a)(U,V). \tag{5.81}$$

The linear analysis applies to the left-hand side in (5.80). It is hyperbolic and is a transport equation if $\alpha\varphi$ is a regular phase, as explained in Lemmas 5.7 and 5.9. Choosing bases in the ranges of the P_α, it can be made explicit as in (5.34). In general, the quadratic term in the right-hand side of (5.80) does not vanish, so that the system for U_0 appears as a *semi-linear* hyperbolic system.

EXAMPLE 5.25 (*Maxwell – anharmonic Lorentz equations*). Consider the system

$$\begin{aligned} &\partial_t B + \mathrm{curl}\, E = 0, \qquad \partial_t E - \mathrm{curl}\, B = -\partial_t P,\\ &\qquad \varepsilon^2\partial_t^2 P + P + V(P) = \gamma E \end{aligned} \tag{5.82}$$

where V is at least quadratic. We apply the general computations above, that is, to asymptotic expansions of order $O(\varepsilon)$. The dispersion relation is the same as for Maxwell–Lorentz. Consider an optical planar phase $\varphi = \omega t + kx$ which satisfies the dispersion relation

$$|k|^2 = \omega^2\left(1 + \frac{\gamma}{1-\omega^2}\right) := \mu^2(\omega). \tag{5.83}$$

The set of the harmonics which satisfies the dispersion relations is $Z = \{-1, 0, +1\}$. The polarization conditions for harmonic $+1$ and -1 are

$$\hat{E}_{\pm 1} \in k^\perp, \qquad \hat{B}_{\pm 1} = -\frac{1}{\omega}k \times \hat{E}_{\pm 1}, \qquad \hat{P}_{\pm 1} = \frac{\gamma \hat{E}_{\pm 1}}{1-\omega^2}. \tag{5.84}$$

Physical solutions correspond to real valued fields, that is, they satisfy the conditions $\hat{E}_{-1} = \overline{\hat{E}_1}$. The polarization conditions for the harmonic 0 (the mean field) reduce to

$$\hat{P}_0 = \gamma \hat{E}_0. \tag{5.85}$$

The profile equations read

$$\begin{cases} \mu'(\omega)\partial_t \hat{E}_1 - \dfrac{k}{|k|} \cdot \partial_x \hat{E}_1 = \mathrm{i}\dfrac{\gamma\omega^2}{2|k|(1-\omega^2)}(V_2(\hat{E}_0, \hat{E}_1))_\perp, \\ \partial_t \hat{B}_0 + \operatorname{curl} \hat{E}_0 = 0, \qquad (1+\gamma)\partial_t \hat{E}_0 - \operatorname{curl} \hat{B}_0 = 0 \end{cases} \tag{5.86}$$

where V_2 is the quadratic part of V at the origin and $A_\perp$ is the projection of A on $k^\perp$.

REMARKS 5.26. (1) (Rectification). When only the harmonics $-1, 0, +1$ are present, one expects from equation (5.80) a quadratic coupling of $\hat{U}_1$ and $\hat{U}_{-1}$ as a source term for the propagation of $\hat{U}_0$, implying that a mean field $\hat{U}_0$ can be created by oscillatory waves. This phenomenon is called *optical rectification*. It is not present in the case of (5.86) where, in addition, the propagation equation for $\hat{U}_0$ is linear. Rectification occurs in optics, but for different equations.

(2) (Generation of harmonics). For non-isotropic crystals (see (2.18), (2.19)), the dispersion relation is non-isotropic in k, and for special values of k the harmonic $(2\omega, 2k)$ can be characteristic, see [39]. In this case, the analogue of (5.86) couples $\hat{E}_1$ and $\hat{E}_2$. This *nonlinear* phenomenon is used in physical devices for *doubling frequencies*.

EXAMPLE 5.27 (*Generic equations for cubic nonlinearities*). When V is cubic, the system (5.86) is linear. In this case, the regime of weakly nonlinear optics is not reached for amplitudes $O(\varepsilon)$, but for amplitudes $O(\varepsilon^{\frac{1}{2}})$. All the computations can be carried out with the prefactor $\sqrt{\varepsilon}$ in front of the sum in (5.69) in place of ε (see [39,41]). For instance, consider the cubic model (2.21) with no damping and oscillations with amplitudes of order $\varepsilon^{\frac{1}{2}}$. With mean field $\hat{E}_0 = 0$, the transport equations for $\hat{E}_1$ are of the form:

$$\partial_t \hat{E}_1 + v_g \partial_x \hat{E}_1 = \mathrm{i}\delta\left(|\hat{E}_1|^2 \hat{E}_1 + \frac{1}{2}(\hat{E}_1 \cdot \hat{E}_1)\overline{\hat{E}_1}\right). \tag{5.87}$$

The solutions of this system can be computed explicitly using the conservations

$$(\partial_t + v_g\partial_x)|\hat{E}_1|^2 = 0, \quad (\partial_t + v_g\partial_x)(\hat{E}_1 \times \overline{\hat{E}_1}) = 0.$$

The solutions are determined from their initial data $\hat{E}_{1|t=0} = \hat{A} \in k^\perp$:

$$\hat{E}_1(t,x) = \mathrm{e}^{\mathrm{i}tI(x-v_gt)} R(t\phi(x-v_gt))\hat{A}(x-v_gt)$$

where $R(s)$ is the rotation of angle s in the plane $k^\perp$, oriented by the direction of k and

$$I = \frac{3\delta}{2}|\hat{A}|^2, \qquad \phi = \frac{\mathrm{i}\delta}{2|k|} k \cdot (A \times \overline{A}).$$

Two physical *nonlinear* phenomena are described by this formula:

- *the polarization of the electric field rotates* at the speed ϕ in the plane $k^\perp$;
- when incorporated in the phase $(\omega t + kx)/\varepsilon$, the term tI corresponds to a *self-modulation of the phase, depending on the intensity* of the field.

5.2.3. The structure of the profile equation II: the non-dispersive case; the generic Burger's equation We now assume that $E = 0$. In this case the set of characteristic frequencies is $Z = \mathbb{Z}$. By homogeneity, there are only two different possibilities:

$$\begin{cases} \alpha = 0: & P_0 = \mathrm{Id}, \qquad Q_0 = 0, \qquad R_0 = 0, \\ \alpha \neq 0: & P_\alpha = P_1, \qquad Q_\alpha = Q_1, \qquad R_\alpha = R_1. \end{cases}$$

It is convenient to split a periodic function $U(\theta)$ into its mean value $\underline{U} = \mathbb{M}U = \hat{U}_0$ and its oscillating part $U^* = \mathbb{O}sU = U - \underline{U}$. Dropping several subscripts 1 in P_1 etc, the Eq. (5.76) reads

$$\begin{cases} PU_0^* = U_0^*, \\ (I - Q)L_1(a, 0, \partial)PU_0^* = (I - Q)\mathbb{O}s\left(\Gamma(U_0, U_0)\right), \\ L_1(a, 0, \partial)\underline{U}_0 = \mathbb{M}\left(\Gamma(U_0, U_0)\right) \end{cases}$$

with

$$\Gamma(U, V) = \nabla^2 f(a)(U_0, U_0) - B(a, d\varphi, U)\partial_\theta V.$$

REMARKS 5.28. **(1) The case of conservation laws.** For balance laws (3.2), the matrices $A_j(u)$ are the Jacobian matrices $\nabla f_j(u)$ of the fluxes f_j. Thus

$$\begin{aligned} B(d\varphi, U)\partial_\theta U = &\ \frac{1}{2}\partial_\theta \left(\sum \partial_j\varphi \nabla_u^2 f_j(0)(U, U)\right) \\ := &\ \frac{1}{2}\partial_\theta \left(b(d\varphi)(U, U)\right) \end{aligned} \tag{5.88}$$

implying that

$$\mathbb{M}\left(B(d\varphi, U)\partial_\theta U\right) = 0.$$

For conservation laws, the source term f vanishes and the equation for $\underline{U}_0$ decouples and reduces to the linearized equation $L_1(0, \partial_{t,x})\underline{U}_0 = 0$. In particular, $\underline{U}_0 = 0$ if its initial value vanishes and the equation for U_0^* reduces to

$$\begin{cases} PU_0^* = U_0^*, \\ (I - Q)L_1(a, 0, \partial)PU_0^* + \dfrac{1}{2}(I - Q)\partial_\theta \left(b(d\varphi)(U_0^*, U_0^*)\right) = 0. \end{cases} \tag{5.89}$$

PROPOSITION 5.29 (*Self-interaction coefficient for simple modes*). *Suppose that*

(a) $\lambda(a, u, \xi)$ *is a simple eigenvalue of* $A_0^{-1}(a, u) \sum \xi_j A_j(a, u)$ *for* (a, u, ξ) *close to* $(\underline{a}, 0, \underline{\xi})$, *with eigenvector* $R(a, u, \xi)$ *normalized by the condition* ${}^t R A_0 R = 1$ *(which means that* R *is unitary for the scalar product defined by* A_0*);*

(b) *the phase* φ *is a a solution of the eikonal equation*

$$\partial_t \varphi + \lambda(a(t, x), 0, \partial_x \varphi) = 0 \tag{5.90}$$

for (t, x) *close to* $(0, \underline{x})$, *with* $\partial_x \varphi(0, \underline{x}) = \underline{\xi}$.

Then,

(i) *the polarization condition* $PU_0^* = U_0^*$ *reads*

$$\begin{aligned} U_0^*(t, x, \theta) &= \sigma(t, x, \theta) r(t, x), \\ r(t, x) &= R(a(t, x), 0, \partial_x \varphi(t, x)), \end{aligned} \tag{5.91}$$

(ii) *the equation for* σ *deduced from the equation for* U_0^* *reads*

$$\partial_t \sigma + \mathbf{v}_g \cdot \partial_x \sigma + b_0(\underline{U}_0) \partial_\theta \sigma + \gamma \partial_\theta \sigma^2 = \mathbb{O}s F(\underline{U}_0, \sigma) \tag{5.92}$$

where $\mathbf{v}_g = \partial_\xi \lambda(a, 0, \partial_x \varphi)$ *is the group velocity as in (5.19), the self-interaction coefficient* γ *is*

$$\gamma = \frac{1}{2} r \cdot \nabla_u \lambda(a, 0, \partial_x \varphi). \tag{5.93}$$

The term b_0 *is linear in* $\underline{U}_0$ *and* F *is at most quadratic.*

In particular, for systems of balance laws, $b_0 = 0$*. For systems of conservation laws and* $a = \underline{a}$ *constant, with a planar phase* $\varphi = \omega t + \xi \cdot x$, *the equation reduces to*

$$\partial_t \sigma + \mathbf{v}_g \cdot \partial_x \sigma + \gamma \partial_\theta \sigma^2 = 0. \tag{5.94}$$

The coefficient γ *vanishes exactly when the eigenvalue* $\lambda(\underline{a}, \cdot, \xi)$ *is linearly degenerate at* $u = 0$.

PROOF. We proceed as in Remark 5.6, noticing, that by symmetry, the right kernel of $-\lambda A_0 + \sum \xi_j A_j$ is generated by ${}^t R$. The equation is therefore

$${}^t r L_1(a, 0, \partial_{t,x})(\sigma r) = \mathbb{O}s {}^t r \Gamma(\underline{U}_0 + \sigma r, \underline{U}_0 + \sigma r).$$

The left-hand side is $(\partial_t + \mathbf{v}_g \cdot \partial_x)\sigma + c\sigma$, as explained in Lemma 5.9. Similarly, differentiating in u the identity $(-\lambda A_0 + \sum \xi_j A_j) r = 0$, and multiplying on the left by ${}^t r$ implies that

$${}^t r B(a, d\varphi, v) = v \cdot \nabla_u \lambda(a, 0, \partial_x \varphi) {}^t r A_0.$$

This implies the proposition. □

REMARK 5.30 (*Generic equations*). Under the assumptions of Proposition 5.29, the equations for $U_0 = \underline{U}_0 + \sigma r$ consist of a hyperbolic system for $\underline{U}_0$ coupled with a transport equation for σ. In the genuinely nonlinear case, the transport equation is a *Burger's equation* which therefore appears as the *generic model for the propagation of the profile of nonlinear oscillations*.

5.2.4. Approximate and exact solutions

PROPOSITION 5.31. *The Cauchy problem for (5.76) is locally well-posed.*

PROOF. The symmetry of the original problem reflects into a symmetry of the profile equation

$$(I - Q)\mathcal{L}_1(a, V, \partial_{t,x,\theta})PU = (I - Q)F. \tag{5.95}$$

This has been seen in Lemma 5.7 for the $\partial_{t,x}$ part. The proof is similar for $(I - Q)BP\partial_\theta$. Thus all the machinery of symmetric hyperbolic systems can be used, implying that the Cauchy problem is locally well-posed. □

In the cascade of equations for U_n, only the the first one for U_0 is nonlinear. This implies that the domain of existence of solutions for the subsequent equations can be chosen independent of n. Therefore:

THEOREM 5.32 (*WKB solutions*). *Suppose that $a \in C^\infty$ and that φ is a C^∞ characteristic phase of constant multiplicity near $(0, \underline{x})$ of $L(a, 0, \partial)$. Let $\mathcal{P}$ be the projector (5.75). Given a neighborhood ω of $\underline{x}$ and a sequence of C^∞ functions H_n on a $(\omega \times \mathbb{T})$ such that $\mathcal{P}_{|t=0}H_n = H_n$, there is a neighborhood Ω of $(0, \underline{x})$ and a unique sequence U_n of C^∞ solutions of (5.76) (5.77) (5.78) on $\Omega \times \mathbb{T}$ such that*

$$\mathcal{P}U_{n\,|t=0} = H_n. \tag{5.96}$$

We suppose below that such solutions of the profile equations are given. One can construct approximate solutions

$$u^\varepsilon_{\mathrm{app},n}(t, x) = \varepsilon \sum_{k=0}^{n} \varepsilon^k U_k(t, x, \varphi(t, x)/\varepsilon) \tag{5.97}$$

which satisfy the Eq. (5.68) with an error term of order $O(\varepsilon^{n+1})$. Using Borel's summation, one can also construct approximate solutions at order $O(\varepsilon^\infty)$. More precisely, $err^\varepsilon_n = L(a, u^\varepsilon_{\mathrm{app},n}, \partial)u^\varepsilon_{\mathrm{app},n} - f(a, u^\varepsilon_{\mathrm{app},n})$ satisfies

$$\varepsilon^{|\alpha|}\|\partial^\alpha_{t,x} err^\varepsilon_n\|_{L^\infty(\Omega)} \le \varepsilon^{n+1}C_\alpha. \tag{5.98}$$

We look for exact solutions

$$u^\varepsilon = u^\varepsilon_{\mathrm{app},n} + v^\varepsilon. \tag{5.99}$$

The equation for v^ε has the form (4.49):

$$\tilde{A}_0(b, v)\partial_t v + \sum_{j=1}^{d} \tilde{A}_j(b, v)\partial_{x_j} v + \frac{1}{\varepsilon} E(a)v = \tilde{F}(b, v) \tag{5.100}$$

where $b = (a, u^\varepsilon_{\mathrm{app},n})$ is uniformly bounded and satisfies estimates (4.50) for $\alpha > 0$:

$$\varepsilon^{|\alpha|-1} \|\partial^\alpha_{t,x} b\|_{L^\infty(\Omega)} \leq C_\alpha. \tag{5.101}$$

The derivatives of $F(b, 0) = \varepsilon^{-1} err^\varepsilon_n$ are controlled by (5.98). Moreover, $A_j(b, v) = A_j(a, u^\varepsilon_{\mathrm{app},n} + v)$. Therefore, we are in a position to apply Theorem 4.21, localized on a suitable domain of determinacy, shrinking Ω if necessary. Consider initial data h^ε which satisfy for $|\alpha| \leq s$

$$\varepsilon^{|\alpha|} \|\partial^\alpha_x h\|_{L^2(\omega)} \leq \varepsilon^n C_\alpha. \tag{5.102}$$

THEOREM 5.33. *If $n > 1 + d/2$, there are $\varepsilon_1 > 0$ and a neighborhood Ω of $(0, \underline{x})$, such that for $\varepsilon \in]0, \varepsilon_1]$, the Cauchy problem for* (5.68) *with initial data $u^\varepsilon_{\mathrm{app},n|t=0} + h^\varepsilon$, has a unique solution $u^\varepsilon \in C^0 H^s(\Omega)$ for all s, and which satisfies*

$$\varepsilon^{|\alpha|} \|\partial^\alpha_x (u^\varepsilon(t) - u^\varepsilon_{\mathrm{app},n}(t))\|_{L^2} \leq \varepsilon^n C_\alpha. \tag{5.103}$$

In particular, this implies that

$$\|\partial^\alpha_x (u^\varepsilon(t) - u^\varepsilon_{\mathrm{app},n}(t))\|_{L^\infty} \leq \varepsilon^{n-\frac{d}{2}-|\alpha|} C_\alpha. \tag{5.104}$$

REMARK 5.34 (*About domains of determinacy*). All the analysis above is local and we have not been very careful about large domains Ω where they would apply. This is a technical and difficult question, we just give some hints.

– The final comparison between u^ε and $u^\varepsilon_{\mathrm{app},n}$ must be performed on a domain of determinacy Ω of ω for the system (5.68). For quasi-linear equations, the domains of determinacy depend on the solution, as explained in Section 3.6. However, for conical domains of the form (3.38), it is sufficient that the slope λ_* is strictly larger than $\lambda_*(M)$ where M only involves a L^∞ bound of $u^\varepsilon_{\mathrm{app}}$, since for small ε, the L^∞ norm of u^ε will remain smaller than M' with $\lambda(M') \leq \lambda_*$.

– If Ω is contained in the domain of determinacy of ω for the full system, then it is also contained in the domain of determinacy of the propagator L_φ if φ is smooth on Ω. For domains of the form (3.38), it easily follows from the explicit formula (3.39). Thus, if φ is defined on Ω, one can solve the profile equations as well as the equation for the residual on Ω and Theorem 5.33 extends to such domains.

– There are special cases where one can improve the general result above. Consider for instance a semi-linear wave equation in $\mathbb{R}^3$ and the phase $\varphi = t + |x|$. Consider $\Omega = \{(t, x) : 0 \le t \le \min\{T, R - |x|\}\}$. Ω is contained in the domain of determinacy of the ball $\omega = \{x : |x| \le R\}$ but φ is *not* smooth on Ω. The transport operator associated with φ is $\partial_t - \partial_r - 2/r$, with $r = |x|$. If the initial oscillatory data are supported away from 0, in the annulus $\{R_1 \le |x| \le R\}$, then the profiles can be constructed for $t \le R_1$, since the rays launched from the support of the initial profiles do not reach the singular set $\{x = 0\}$ before this time. Therefore, if $T \le R_1$, and decreasing it if necessary because of the nonlinearity of the first profile equation, one can construct profiles and approximate solutions on Ω, whose support do not meet the set $\{x = 0\}$ where the phase is singular. Next one can compare approximate and exact solutions on Ω.

5.3. *Strongly nonlinear expansions*

In the previous section we discussed the standard regime of weakly nonlinear geometric optics. No assumptions on the structure of the nonlinear terms were made. There are cases where these general theorems do not provide satisfactory results. Typically, this happens when interaction coefficients vanish because of the special structure of the equations. This implies that the transport equations are linear instead of being nonlinear. This phenomenon is called *transparency* in [41]. This happens, for instance, in the following two cases:

- for systems of conservation laws and oscillations polarized in a linearly degenerate mode, since then the self interaction coefficient γ in (5.94) vanishes.
- for semi-linear dispersive systems with quadratic nonlinearity $f(u, u)$ when the polarization projectors P_α and Q_α associated with the characteristic harmonic phases $\alpha\varphi$, $\alpha \in Z$, satisfy

$$(I - Q_\alpha) f(P_\beta, P_\gamma) = 0, \qquad (\alpha, \beta, \gamma) \in Z^3, \qquad \alpha = \beta + \gamma. \tag{5.105}$$

The transparency condition is closely related to the *null condition* for quadratic interaction introduced to analyze the global existence of smooth small solutions, as it means that some interactions of oscillations are not present. It happens to be satisfied in many examples from Physics. Note also, that the transparency condition is completely different from the degeneracy condition evoked in Remarks 5.26 (3) for cubic nonlinearities: there the nonlinear terms were absent in the main profile equation for *all* polarizations, because the nonlinearity was too weak. In the present case, the quadratic term is not identically zero; only certain quadratic interactions are killed.

To deal with nonlinear regimes, one idea is to consider waves of larger amplitude or, equivalently, of higher energy. This program turns out to be very delicate. We do not give a complete report on the available results here but just give several hints and references. Two questions can be raised:

- What are the conditions for the construction of BKW solutions?

- When they exist, what are the conditions for their stability, i.e. when are they close to exact solutions?

It turns out that there is no general answer to the second question: there are (many) cases where one can construct BKW solutions, which define approximate solutions of the equation at any order ε^n, but which are strongly unstable, due to a *supercritical* nonlinearity. This emphasizes the importance of the stability results obtained in the weakly nonlinear regimes.

5.3.1. An example: two levels Maxwell–Bloch equations We first give an example showing that the particular form of the equations plays a fundamental role in the setting of the problem. Consider the system

$$\begin{cases} \partial_t B + \operatorname{curl} E = 0, & \partial_t E - \operatorname{curl} B = -\partial_t P, \\ \varepsilon^2 \partial_t P + P = \gamma_1 \, (N_0 + N) E, & \partial_t N = -\gamma_2 \partial_t P \cdot E. \end{cases} \tag{5.106}$$

Introducing $Q = \varepsilon \partial_t P$ and $u = (B, E, P, Q, N)$, it falls into the general framework of semi-linear dispersive equations (5.68), with quadratic nonlinearity:

$$L(\varepsilon\partial)u == \varepsilon\partial_t u + \sum_{j=1}^{d} \varepsilon A_j u + Eu = q(u, u).$$

Consider BKW solutions (5.69) associated with a planar phase φ. The general theory for quadratic interaction concerns solutions of amplitude $O(\varepsilon)$ ($p = 1$ in (5.69)). Recall that the wave number $\beta = d\varphi$ satisfies the eikonal equation $\det L(\mathrm{i}\beta) = 0$ and the Fourier coefficients of the principal term $U_0 = \sum \hat{U}_{0,\nu} \mathrm{e}^{\mathrm{i}\nu\theta}$ satisfy the polarization condition $\hat{U}_{0,\nu} = P(\nu\beta)\hat{U}_{0,\nu}$, where $P(\xi)$ is the orthogonal projector on $\ker L(\mathrm{i}\xi)$. The propagation equations for $\hat{U}_{0,\nu}$ are a coupled system of hyperbolic equations:

$$\hat{L}_\nu \hat{U}_{0,\nu} = P(\nu\beta) \sum_{\nu_1+\nu_2=\nu} q(\hat{U}_{0,\nu_1}, \hat{U}_{0,\nu_2}). \tag{5.107}$$

When the characteristic phases $\nu\varphi$ are regular with group velocity $\mathbf{v}_\nu$, $\hat{L}_\nu = \partial_t + \mathbf{v}_\nu \cdot \partial_x$.

For the Maxwell–Bloch equations (5.106), this analysis is unsatisfactory for two reasons. First, for physically relevant choices of U_0, the interaction terms vanish. Thus the transport equations are linear and the nonlinear regime is not reached. Second, Maxwell–Bloch equations are supposed to be a refinement of cubic models in nonlinear optics, such as the anharmonic oscillator model which is discussed in Remarks 5.26 (3). Both facts suggest that solutions of amplitude $O(\sqrt{\varepsilon})$ could exist. Actually, BKW solutions of the equation with $p = 1/2$, are constructed in [41]. Indeed, for Eq. (5.106), there is an easy trick, *a change of unknowns, which reduces the quadratic system to a cubic one*. Introduce the following inhomogeneous scaling of the amplitudes:

$$(B, E, P, Q) = \sqrt{\varepsilon}(\widetilde{B}, \widetilde{E}, \widetilde{P}, \widetilde{Q}), \qquad N = \varepsilon \widetilde{N}.$$

Then, the Maxwell–Bloch equations read

$$\begin{cases} \varepsilon\partial_t \widetilde{B} + \varepsilon\mathrm{curl}\widetilde{E} = 0, & \varepsilon\partial_t \widetilde{E} - \varepsilon\mathrm{curl}\widetilde{B} = -\widetilde{Q}, \\ \varepsilon\partial_t \widetilde{P} - \widetilde{Q} = 0, & \varepsilon\partial_t \widetilde{Q} + \Omega^2 P = \gamma_1 N_0 \widetilde{E} + \varepsilon\gamma_1 \widetilde{N}\widetilde{E}, \\ \varepsilon\partial_t \widetilde{N} = -\gamma_2 \widetilde{Q}\cdot\widetilde{E}. \end{cases}$$

This scaling agrees with the BKW solutions of [41]. The question is to construct oscillatory solutions $\tilde{u} = (\widetilde{B}, \widetilde{E}, \widetilde{P}, \widetilde{Q}, \widetilde{N})$ of amplitude $O(1)$. The difficulty is that the source term in the last equation is nonlinear and of amplitude $O(1)$. A tricky argument gives the answer. Consider the change of unknowns

$$n = \widetilde{N} + \frac{\gamma_2}{\gamma_1 N_0}(\widetilde{Q}^2 + \Omega^2 \widetilde{P}^2).$$

Then the last equation is transformed into

$$\varepsilon\partial_t n = \varepsilon\frac{\gamma_2}{N_0}\widetilde{N}\widetilde{Q}\cdot\widetilde{E} = \varepsilon(c_1 n - c_2(\widetilde{Q}^2 + \Omega^2\widetilde{P}^2))\widetilde{Q}\cdot\widetilde{E}$$

and the source term is now $O(\varepsilon)$. Introducing $u^\sharp := (\widetilde{B}, \widetilde{E}, \widetilde{P}, \widetilde{Q}, n)$, the system is transformed to a system

$$L^\sharp(\varepsilon\partial_x)u^\sharp = \varepsilon f(u^\sharp)$$

where the key point is that the right hand side is $O(\varepsilon)$. For this equation, the *standard* regime of nonlinear geometric optics concerns $O(1)$ solutions and thus one can construct BKW solutions and prove their stability.

This algebraic manipulation yields a result with physical relevance. The transport equations for the amplitude obtained in this are cubic, in accordance with the qualitative properties given by other isotropic models of nonlinear optics such as the anharmonic model or the Kerr model (see [39]).

5.3.2. A class of semi-linear systems The analysis above has been extended to more general Maxwell–Bloch equations. In [73], one considers the more general framework:

$$\begin{cases} L(\varepsilon\partial)u + \varepsilon f(u, v) = 0, \\ M(\varepsilon\partial)v + q(u, u) + \varepsilon g(u, v) = 0, \end{cases} \tag{5.108}$$

where f and g are smooth polynomial functions of their arguments and vanish at the origin, q is bilinear and

$$L(\varepsilon\partial) := \varepsilon\partial_t + \sum \varepsilon A_j\partial_{x_j} + L_0 := \varepsilon L_1(\partial) + L_0,$$
$$M(\varepsilon\partial) := \varepsilon\partial_t + \sum \varepsilon B_j\partial_{x_j} + M_0 := \varepsilon M_1(\partial) + M_0$$

are symmetric hyperbolic, meaning that the A_j and B_j are Hermitian symmetric while L_0 and M_0 are skew-adjoint. The main feature of this system is that the principal nonlinearity $q(u, u)$ appears only on the second equation and depends only in the first set of unknowns u. The goal is to construct solutions satisfying

$$u^\varepsilon(t, x) \sim \sum_{n\geq 0} \varepsilon^n U_n(t, x, \varphi/\varepsilon), \qquad v^\varepsilon(x) \sim \sum_{n\geq 0} \varepsilon^n V_n(t, x, \varphi/\varepsilon) \tag{5.109}$$

where $\varphi = \omega t + kx$ is a linear phase. The profiles $U_n(t, x, \theta)$ and $V_n(t, x, \theta)$ are periodic in θ. The difficulty comes from the $O(1)$ interaction term in the second equation. We do not write the precise results of [73] but we point out the different levels of compatibility that are needed to carry out the analysis:

Level 1: The transparency condition. The eikonal equation is that $\det L(\mathrm{i}d\varphi) = 0$. Denote by $P(\beta)$ the orthogonal projector on $\ker L(\mathrm{i}\beta)$. Similarly introduce $Q(\beta)$, the orthogonal projector on $\ker M(\mathrm{i}\beta)$. Assume

$$\det L(\mathrm{i}\nu d\varphi) \neq 0 \quad \text{and} \quad \det M(\mathrm{i}\nu d\varphi) \neq 0 \quad \text{for } \nu \text{ large.}$$

The transparency condition states that polarized oscillations U_0 in $\ker L(d\varphi\partial_\theta)$ produce oscillations $q(U_0, U_0)$ in the range of $M(d\varphi\partial_\theta)$: For all integers ν_1 and ν_2 in $\mathbb{Z}$ and for all vectors u and v,

$$Q((\nu_1 + \nu_2)d\varphi)q\left(P(\nu_1 d\varphi)u, P(\nu_2 d\varphi)v\right) = 0. \tag{5.110}$$

When this condition is satisfied, one can write a triangular cascade of equations for the (U_n, V_n). Using notations $\mathcal{P}$ and $\mathcal{Q}$ similar to (5.75) for the projectors on $\ker L(d\varphi\partial_\theta)$ and $\ker M(d\varphi\partial_\theta)$ respectively, and $\mathcal{M}^{-1}$ for a partial inverse of $M(d\varphi\partial_\theta)$, the equations for U_0 and $\widetilde{V}_0 = V_0 + \mathcal{M}^{-1}q(U_0, U_0)$ are

$$\begin{aligned} &\mathcal{P}L_1(\partial)\mathcal{P}U_0 = \mathcal{P}f(U_0, V_0), \qquad \mathcal{P}U_0 = U_0, \\ &\mathcal{Q}M_1(\partial)\mathcal{Q}\widetilde{V}_0 + \mathcal{D}(U_0, \partial_x)U_0 = \mathcal{Q}G(U_0, V_0), \qquad \mathcal{Q}\widetilde{V}_0 = \widetilde{V}_0 \end{aligned} \tag{5.111}$$

where G is a nonlinear functional which involves the projectors $\mathcal{P}$, $\mathcal{Q}$ and the partial inverse $\mathcal{M}^{-1}$. There are similar equations for the terms (U_n, V_n).

Level 2: Hyperbolicity of the profile equations. This assumption can be made explicit (see again [73]). When it is satisfied, one can construct BKW solutions and approximate solutions at any order.

Level 3: Stability of BKW solutions. The linear and nonlinear, local in time, stability of the approximate solutions is proved under the following assumption:

For all $\nu \in \mathbb{Z}$, there is a constant C such that for all wave numbers β β' and all vectors u and u'

$$\left|Q(\beta')q\left(P(\nu d\varphi)u, P(\beta)u'\right)\right| \leq C|\beta'_0 - \beta_0 - \nu\partial_t\varphi||u||u'|$$

where β_0 denotes the first component of β. It is shown that this condition implies both the conditions of levels 1 (easy) and 2 (more intricate).

Level 4: Normal form of the equation. The condition above can be strengthened:
There is a constant C such that for all wave numbers β, β', β'' and all vectors u and u'

$$\left|Q(\beta'')q\left(P(\beta)u, P(\beta')u'\right)\right| \le C|\beta_0'' - \beta_0 - \beta_0'||u||u'|.$$

When this condition is satisfied, the system (5.108) is conjugate, via a nonlinear pseudo-differential change of unknowns, to a similar system with $q = 0$. Exceptionally, the change of variables may be local and not involve pseudo-differential operators, this is the case of Maxwell–Bloch equations.

5.3.3. Field equations and relativity The construction of waves using asymptotic expansions is used in many areas of physics. In particular, in general relativity it is linked to the construction of nonlinear gravity waves (see [26,27,1]). This also concerns field equations. An important new difficulty is that these equations are *not* hyperbolic, due to their gauge invariance. We mention the very important paper [71], where this difficulty is examined in detail, providing BKW expansions *and* rigorous justification of the asymptotic expansion for exact solutions, with applications to Yang–Mills equations and relativistic Maxwell's equations. For these equations, the transparency condition is satisfied so that the relevant nonlinear analysis concerns oscillations of large amplitude. An important feature of the paper [71] is that the compatibility condition is used for the construction of both asymptotic and exact solutions. The rigorous justification of the asymptotic expansion given by Y.Choquet-Bruhat for Einstein's equations, seems to remain an open problem.

5.3.4. Linearly degenerate oscillations For systems of conservation laws, the interaction coefficient γ in the Burger's transport equation (5.94) vanishes when the mode is linearly degenerate. We briefly discuss in this paragraph the construction and the stability of oscillatory waves

$$u^\varepsilon(t,x) \sim \underline{u}_0(t,x) + \sum_{n\ge p} (\sqrt{\varepsilon})^n U_n\left(t, x, \frac{\varphi_\varepsilon(t,x)}{\varepsilon}\right) \tag{5.112}$$

for symmetric hyperbolic systems of conservation laws

$$\partial_t f_0(u) + \sum_{j=1}^{d} \partial_{x_j} f_j(u) = 0. \tag{5.113}$$

The unperturbed state $\underline{u}_0$ is a solution of (5.113) and the principal term of the phase $\varphi_\varepsilon = \varphi_0 + \sqrt{\varepsilon}\varphi_1$ satisfies the eikonal equation

$$\partial_t\varphi_0 + \lambda(\underline{u}_0, \partial_x\varphi_0) = 0 \tag{5.114}$$

where $\lambda(u,\xi)$ is a *linearly degenerate* eigenvalue of constant multiplicity of the system.

The standard regime of weakly nonlinear geometric optics corresponds to $p = 2$. Because of the linear degeneracy, we consider now the case of larger amplitudes corresponding to the cases $p = 1$ and $p = 0$.

- In space dimension $d = 1$ and for $p = 0$, a complete analysis is given in [47] for gas dynamics, and in [62,120,34] in a general framework providing exact solutions which admit asymptotic expansions of the form.
- In any dimension, associated with linearly degenerate modes, there always exist simple waves

$$u(t,x) = v(h(\xi \cdot x - \sigma t)),$$

that are exact solutions of (5.113), with $v(s)$ a well-chosen curve defined for $s \in I \subset \mathbb{R}$, ξ and h *an arbitrary* $C^1(\mathbb{R}; I)$ function (cf [103]). Choosing h periodic with period ε, yields exact solutions satisfying (5.112) with $p = 0$ and associated with the phase $\varphi = \xi \cdot x - \sigma t$. The question is to study the stability of such solutions: if one perturbs the initial data, do the solutions resemble to the unperturbed solutions? We give below partial answers, mainly taken from [23,24], but we also refer to [22] for extensions.

1. *Existence of BKW solutions for* $p = 1$. It is proved in [23] that in general, one can construct asymptotic solutions (5.112) with $p = 1$, the main oscillation U_1^* being polarized along the eigenspace associated with λ. The average $\underline{u}_1$ of U_1 is an arbitrary solution of the linearized equation from (5.113) at $\underline{u}_0$. When $\underline{u}_1 \neq 0$, a correction $\sqrt{\varepsilon}\varphi_1$ must be added to the phase φ_0, satisfying

$$\partial_t \varphi_1 + \partial_x \varphi_1 \cdot \nabla_\xi \lambda(\underline{u}_0, \partial_x \varphi_0) + \underline{u}_1 \cdot \nabla_u \lambda(\underline{u}_0, \partial_x \varphi_0) = 0 \tag{5.115}$$

so that $\varphi_\varepsilon = \varphi_0 + \sqrt{\varepsilon}\varphi_1$ satisfies the eikonal equation at the order $O(\varepsilon)$, i.e.:

$$\partial_t \varphi_\varepsilon + \lambda(\underline{u}_0 + \sqrt{\varepsilon}\underline{u}_1, \partial_x \varphi_\varepsilon) = O(\varepsilon). \tag{5.116}$$

Moreover, the evolution of the main evolution U_1^* is nonlinear (in general) and coupled to the evolution of the average of U_2. The nonlinear regime is reached.

2. *Stability / Instability of BKW solutions.* In space dimension $d = 1$, the linear and nonlinear stability is proved using the notion of a *good symmetrizer* cf [34,62,110,119]. When $d > 1$, the example of gas dynamics shows that the existence of a good symmetrizer does not suffice to control oscillations which are transversal to the phase and which can provoke *strong instabilities of Rayleigh type*, cf. [54,57].

As before, the question is to know whether the approximate solutions constructed by the BKW method are close to exact solutions, on some time interval independent of ε. The main difficulty can be seen from two different angles:

– with $u^\varepsilon_{\text{app}} = \underline{u}_0 + \sqrt{\varepsilon}U_1(t, x, \varphi_\varepsilon/\varepsilon) + \ldots$, the coefficients $f'_j(u^\varepsilon_{\text{app}})$ are not uniformly Lipschtiz continuous, so that the usual energy method does not provide solutions on a uniform domain,

– the linearized equations involve a singular term in $\varepsilon^{-\frac{1}{2}} D$, which has no reason to be skew symmetric.

The first difficulty is resolved if one introduces the fast variable and looks for exact solutions of the form

$$u^\varepsilon(t,x) = U^\varepsilon(t,x,\varphi_\varepsilon/\varepsilon). \tag{5.117}$$

The equation for U^ε reads

$$\sum_{j=0}^{d} A_j(U^\varepsilon)\,\partial_j U^\varepsilon + \frac{1}{\varepsilon}\sum_{j=0}^{d}\partial_j\varphi_\varepsilon\; A_j(U^\varepsilon)\partial_\theta U^\varepsilon = 0. \tag{5.118}$$

The construction of BKW solutions yields approximate solutions $U^\varepsilon_{\text{app}}$. The linearized equation is of the form

$$\mathcal{L}_c^\varepsilon + \frac{1}{\varepsilon} G_0 \partial_\theta + \frac{1}{\sqrt{\varepsilon}}(G_1\partial_\theta + H_1) + C^\varepsilon \tag{5.119}$$

where $\mathcal{L}^\varepsilon$ is *symmetric hyperbolic, with smooth coefficients in the variables* (t,x,θ). Moreover, G_0 is symmetric and independent of θ. On the other hand, G_1 is symmetric but *does depend* on θ, as it depends on U_1. Therefore the energy method after integration by parts reveals the singular term

$$\frac{1}{\sqrt{\varepsilon}}\left(D\dot{U},\dot{U}\right) \quad \text{with} \quad D := \frac{1}{2}\left(-\partial_\theta G_1 + H_1 + H_1^*\right). \tag{5.120}$$

This matrix can be computed explicitly: denoting by $S(u)$ a symmetrizer of the system (5.113) and by $\Sigma(u,\xi) := \sum \xi_j S(u) f_j' u$,

$$\left(D\dot{U},\dot{U}\right) = D'(\partial_\theta U_1, \dot{U}, \dot{U}) + D''(\partial_\theta U_1, \dot{U}, \dot{U}) \tag{5.121}$$

with

$$2D'(u,v,w) = -\left((u\cdot\nabla_u\Sigma)v, w\right) + \left((v\cdot\nabla_u\Sigma)u, w\right) + \left((w\cdot\nabla_u\Sigma)u, v\right),$$
$$2D''(u,v,w) = \left((v\cdot\nabla_u\lambda)Su, w\right) + \left((w\cdot\nabla_u\lambda)Su_1, v\right)$$

where $\nabla_u\Sigma$, $\nabla_u\lambda$ and S are taken at $\underline{u}_0$ and $\xi = \partial_x\varphi_0$. Note that $D(t,x,\theta)$ is a symmetric matrix. Since its average in θ vanishes, it must vanish if it is nonnegative. Therefore:

the energy method provides uniform L^2 estimates for the linearized equations if and only if $D(t,x,\theta) \equiv 0$.

Conversely, this condition is sufficient for the existence of Sobolev estimates and for the nonlinear stability (see [23]):

THEOREM 5.35. *If $U^\varepsilon_{\rm app}$ is a BKW approximate solution of order ε^n with n sufficiently high and if the matrix D vanishes, there are exact solutions U^ε of (5.118) such that $U^\varepsilon - U^\varepsilon_{\rm app} = O(\varepsilon^n)$.*

The condition $D = 0$ is very strong and unrealistic in general, except if $U_1^* = 0$, in which case we recover the standard scaling studied before. In gas dynamics, it is satisfied when the oscillations of U_1 concern only the entropy but neither the velocity nor the density: these are what we call entropy waves, but in this case, one can go further and construct oscillations of amplitude $O(1)$ as explained below. Conversely, the Rayleigh instabilities studied in [54] for gas dynamics seem to be the general behavior that one should expect when $D \neq 0$.

3. *Entropy waves* In contrast to the one dimensional case, D. Serre [119] has shown that the construction of $O(1)$ solutions ($p = 0$) for isentropic gas dynamics yields ill-posed equations for the cascade of profiles. The case of full gas dynamics is quite different: one can construct $O(1)$ oscillations provided that the main term concerns only the entropy and not the velocity. Consider the complete system of gas dynamics expressed in the variables (p, v, s), pressure, velocity and entropy:

$$\begin{cases} \rho\,(\partial_t v + (v\cdot\nabla_x)v) + \nabla_x p = 0,\\ \alpha\,(\partial_t p + (v\cdot\nabla_x)p) + \operatorname{div}_x v = 0,\\ \partial_t s + (v\cdot\nabla_x)s = 0 \end{cases} \tag{5.122}$$

with $\rho(p, s) > 0$ and $\alpha(p, s) > 0$. In [24] (see also [22] for extensions), we prove the existence and the stability of non-trivial solutions $u^\varepsilon = (v^\varepsilon, p^\varepsilon, s^\varepsilon)$ of the form

$$\begin{aligned} v^\varepsilon(t,x) &= v_0(t,x) + \varepsilon V^\varepsilon(t,x,\varphi(t,x)/\varepsilon),\\ p^\varepsilon(t,x) &= p_0 + \varepsilon P^\varepsilon(t,x,\varphi(t,x)/\varepsilon),\\ s^\varepsilon(t,x) &= S^\varepsilon\,(t,x,\varphi(t,x)/\varepsilon) \end{aligned} \tag{5.123}$$

with V, P and S functions of (t, x, θ), periodic in θ and admitting asymptotic expansions $\sum \varepsilon^n V_n$ etc. The unperturbed velocity v_0 satisfies the over-determined system

$$\partial_t v_0 + (v_0\cdot\nabla_x)v_0 = 0, \qquad \operatorname{div}_x v_0 = 0,$$

for instance v_0 can be a constant. Moreover, p_0 is a constant and the phase φ is a smooth real-valued function satisfying the eikonal equation $\partial_t\varphi + (v_0\cdot\nabla_x)\varphi = 0$. The evolution of the principal term is governed by polarization conditions and a coupled system of propagation equations.

For instance, consider the important example where v_0 is constant (and we can take $v_0 = 0$ by Galilean transformation). We choose a linear phase function solution of the eikonal equation $\partial_t\varphi = 0$, and by rotating the axis we have $\varphi(t, x) \equiv x_1$. The components of the velocity are accordingly split into $v = (v_1, w)$. For the principal term, the polarization condition requires that the oscillations of the first components of V_0 and P_0, denoted by $V^*_{0,1}$ and P^*_0, vanish. Therefore

$$
\begin{cases}
v_1^\varepsilon(t,x) = \varepsilon \underline{V}_{0,1}(t,x) + O(\varepsilon^2),\\
w^\varepsilon(t,x) = \varepsilon W_0(t,x,x_1/\varepsilon) + O(\varepsilon^2),\\
p^\varepsilon(t,x) = p_0 + \varepsilon \underline{P}_0(t,x) + O(\varepsilon^2),\\
s^\varepsilon(t,x) = S(t,x,x_1/\varepsilon) + O(\varepsilon).
\end{cases}
$$

The profiles $V_1(t,x)$, $V_2(t,x,\theta)$, $P(t,x)$ and $S(t,x,\theta)$ satisfy

$$
\begin{cases}
\langle \rho(p_0,S)\rangle \partial_t \underline{V}_{0,1} + \partial_1 \underline{P} = 0, & V_{0,1\,|t=0} = a_1(x),\\
\rho(p_0,S)\,(\partial_t W_0 + \underline{V}_{0,1}\partial_\theta W_0) + \nabla' \underline{P} = 0, & W_{0|t=0} = a_2(x,\theta),\\
\langle \alpha(p_0,S)\rangle \partial_t \underline{P} + \partial_1 \underline{V}_{0,1} + \langle \mathrm{div}' W\rangle = 0, & \underline{P}_{|t=0} = a_3(x),\\
\partial_t S + \underline{V}_{0,1}\partial_\theta S = 0, & S_{|t=0} = a_4(x,\theta)
\end{cases}
$$

where $\langle U\rangle$ denotes the average in θ of the periodic function $U(\theta)$, and ∇' and div' denote the gradient and the divergence in the variables $(x_2,\dots,x_d)$, respectively.

5.3.5. Fully nonlinear geometric optics In this paragraph, we push the analysis one step further and consider an example where the amplitude of the oscillation is so large that the eikonal equation for the phase and the propagation equation for the profile are coupled. This example is taken from [95] (see also [127]) and concerns the Klein–Gordon equation

$$
\varepsilon^2 \partial_t^2 u - \varepsilon^2 \partial_x^2 u + f(u) = 0 \tag{5.124}
$$

with $f(u) = u^p$, $p \in \mathbb{N}$, $p > 1$ and odd. As in [95] one can consider a much more general f, which may also depend on (t,x), but for simplicity we restrict the exposition to this case. For such equations, the weakly-nonlinear regime concerns solutions of amplitude $O(\varepsilon^{\frac{2}{p-1}})$. Here we look for solutions of amplitude $O(1)$:

$$
u^\varepsilon(t,x) = U^\varepsilon\left(t,x,\frac{\varphi(t,x)}{\varepsilon}\right), \qquad U^\varepsilon(t,x,\theta) \sim \sum_{k\geq 0} \varepsilon^k U_k(t,x,\theta) \tag{5.125}
$$

with $U_k(t,x,\theta)$ periodic in θ. The *compatibility or transparency condition* which is necessary for the construction is stated in Proposition 5.37 below.

Plugging (5.125) into (5.124) yields the following equations

$$
\begin{aligned}
&\sigma^2 \partial_\theta^2 U_0 + f(U_0) = 0,\\
&\sigma^2 \partial_\theta^2 U_1 + \partial_u f(U_0) U_1 + T\partial_\theta U_0 = 0,\\
&\sigma^2 \partial_\theta^2 U_k + \partial_u f(U_0) U_k + T\partial_\theta U_{k-1} + \Box U_{k-2} + R_k(x,U_0,\dots,U_{k-1}) = 0
\end{aligned}
$$

with $\Box = \partial_t^2 - \partial_x^2$,

$$
\sigma^2 = (\partial_t\varphi)^2 - (\partial_x\varphi)^2, \qquad T := 2\varphi_t'\partial_t - 2\varphi_x'\partial_x + \Box\varphi. \tag{5.126}
$$

Moreover, the R_k denote smooth functions of their arguments.

THEOREM 5.36. *The profile equations, together with initial conditions, admit solutions* $\{U_k\}$.

We sketch the first part of the proof taken from [95].

(1) Since σ is independent of θ, the first equation is an o.d.e. in θ depending on the parameter σ. For σ fixed, the solutions depend on two parameters. They are periodic in θ, the period depending on the energy. Imposing the period equal to 2π determines one of the parameters. Summing up, we see that

the 2π *periodic in* θ *solutions of the first equations are*

$$U_0(t,x,\theta) = V_0(t,x,\theta+\Theta(t,x)) = K\left(\sigma(t,x),\theta+\Theta(t,x)\right) \tag{5.127}$$

with Θ an arbitrary phase shift and $K(\sigma,\theta) = \sigma^{\frac{2}{p-1}} G(\theta)$, where $G(\theta)$ is the unique 2π-periodic solution of $\partial_\theta^2 G + G^p = 0$ satisfying $G'(0) = 0$, $G(0) > 0$.

(2) The second equation is a linear o.d.e.

$$\mathcal{L}U_1 = -T\partial_\theta U_0 \tag{5.128}$$

where $\mathcal{L}$ is the linearized operator at U_0 from the first equation. The translation invariance implies that the kernel of $\mathcal{L}$ is not trivial. Indeed, for fixed (t,x), $\mathcal{L}$ is self adjoint and ker $\mathcal{L}$ is a one-dimensional space generated by $\partial_\theta V_0(t,x,\cdot+\Theta)$. Therefore, the o.d.e. in U_1 has a periodic solution if and only if the right-hand side is orthogonal to the kernel. There holds

$$-T\partial_\theta U_0(t,x,\theta) = -\left\{T\partial_\theta V_0 + X(\Theta)\partial_\theta^2 V_0\right\}(t,x,\theta+\Theta)$$

where $X = 2\varphi_t'\partial_t - 2\varphi_x'\partial_x$. Therefore the integrability condition reads

$$\begin{aligned}\int_0^{2\pi} ((T\partial_\theta V_0)\partial_\theta V_0 + Z(\Theta)\partial_\theta^2 V_0 \partial_\theta V_0)\,\mathrm{d}\theta &= \frac{1}{2}T\int_0^{2\pi} (\partial_\theta V_0)^2 \mathrm{d}\theta \\ &= 0,\end{aligned} \tag{5.129}$$

that is

$$\partial_t\varphi\partial_t J - \partial_x\varphi\partial_x J + \Box\varphi J = 0, \tag{5.130}$$

where

$$\begin{aligned}J(t,x) &:= \int_0^{2\pi} (\partial_\theta K)^2 \mathrm{d}\theta \\ &= \sigma^{\frac{4}{p-1}} \int_0^{2\pi} (\partial_\theta G(\theta))^2\,\mathrm{d}\theta \\ &= c\left((\partial_t\varphi)^2 - (\partial_x\varphi)^2\right)^{\frac{4}{p-1}}.\end{aligned}$$

Therefore, (5.130) is a second order nonlinear equation in φ:

$$\alpha\partial_t^2\varphi - 2\beta\partial_t\partial_x\varphi - \gamma\partial_x^2\varphi = 0 \tag{5.131}$$

with

$$\alpha = \frac{p-1}{2}\sigma^2 + (\partial_t\varphi)^2, \quad \beta = \partial_t\varphi\,\partial_x\varphi, \quad \gamma = \frac{p-1}{2}\sigma^2 - (\partial_x\varphi)^2.$$

Note that this equation is strictly hyperbolic for $\sigma > 0$. It is a substitute for the eikonal equation, but its nature is completely different:

φ is determined by the nonlinear wave equation (5.131)

The compatibility condition (5.129) being satisfied, the solutions of (5.128) are

$$U_1(t,x,\theta) = \lambda_1\partial_\theta U_0 - \mathcal{L}^{-1}T(\partial_\theta U_0) = \lambda_1\partial_\theta U_0 + W_1 \tag{5.132}$$

where λ_1 is an arbitrary function of (t,x) and $\mathcal{L}^{-1}$ denotes the partial inverse on to $\ker\mathcal{L}^\perp$ of the operator $\mathcal{L} = \sigma^2\partial_\theta^2 + f'(U_0)$. Note that $U_1(t,x,\theta)$ is also of the form $V_1(t,x,\theta+\Theta(t,x))$.

(3) The third equation is of the form

$$\mathcal{L}U_2 = F_2 = T\partial_\theta U_1 + \frac{1}{2}f''(U_0)U_1^2 + \Box U_0. \tag{5.133}$$

For this o.d.e in U_2, the integrability condition reads

$$\int_0^{2\pi} F_2(t,x,\theta)\partial_\theta U_0(t,x,\theta)\mathrm{d}\theta = 0. \tag{5.134}$$

Using (5.132), it is an equation for λ_1 and Θ.

PROPOSITION 5.37 (*[95]*). *If the phase equation* (5.130) *is satisfied, then the condition* (5.134) *is independent of λ_1 and reduces to a second order hyperbolic equation for Θ.*

PROOF. The term in λ_1^2 is $\frac{1}{2}\int_0^{2\pi} f''(U_0)(\partial_\theta U_0)^3\mathrm{d}\theta$. The integral vanishes, since the translation invariance implies that

$$\left(\sigma^2\partial_\theta^2 + f'(U_0)\right)\partial_\theta^2 U_0 = -f''(U_0)(\partial_\theta U_0)^2 \tag{5.135}$$

implying that the right hand side is orthogonal to $\partial_\theta U_0$.

The linear term in λ_1 is

$$\int_0^{2\pi} T(\lambda_1\partial_\theta^2 U_0)\partial_\theta U_0\mathrm{d}\theta + \int_0^{2\pi}\lambda_1 f''(U_0)(\partial_\theta U_0)^2 W_1\mathrm{d}\theta.$$

Using (5.135), the second integral is

$$-\int_0^{2\pi} \mathcal{L}(\partial_\theta^2 U_0) W_1 \mathrm{d}\theta = \int_0^{2\pi} \partial_\theta^2 U_0 \, (T \partial_\theta U_0) \mathrm{d}\theta$$

so that the linear term is

$$T\left(\int_0^{2\pi} \lambda_1 \partial_\theta^2 U_0 \partial_\theta U_0 \mathrm{d}\theta\right) = T(0) = 0.$$

Therefore the condition (5.134) reduces to

$$\int_0^{2\pi} \left(T(\partial_\theta W_1) + \frac{1}{2} f''(U)(W_1)^2 + \Box U_0\right) \partial_\theta U_0 \mathrm{d}\theta = 0. \tag{5.136}$$

There holds

$$W_1(t,x,\theta) = -\left\{\mathcal{L}_0^{-1}(T\partial_\theta V_0) + X(\Theta)\mathcal{L}_0^{-1}(\partial_\theta^2 V_0)\right\}(t,x,\theta+\Theta),$$
$$\Box U_0(t,x,\theta) = \{\Box\Theta \, \partial_\theta V_0 + \partial_t \Theta A + \partial_x \Theta B + \Box V_0\}(t,x,\theta+\Theta)$$

where $\mathcal{L}_0^{-1}$ denotes the partial inverse of $\mathcal{L}_0 = \sigma^2 \partial_\theta^2 + f'(V_0)$, and A and B depend only on V_0. Therefore, (5.136) is a second order nonlinear equation in Θ, whose principal part is

$$\Box\Theta \int_0^{2\pi} (\partial_\theta V_0)^2 \mathrm{d}\theta + X^2\Theta \int_0^{2\pi} \mathcal{L}_0^{-1}(\partial_\theta^2 V_0)\partial_\theta^2 V_0 \mathrm{d}\theta.$$

$\square$

(4) Construction of BKW solutions. The phase φ is constructed from (5.131) and suitable initial conditions. The phase shift Θ is given by the nonlinear hyperbolic equation and initial conditions. This completely determines the principal term U_0, U_1 up to the choice of λ_1, and U_2 is also determined from (5.133) up to a function $\lambda_2 \partial_\theta U_0$ in the kernel of $\mathcal{L}$. The fourth equation in the cascade is $\mathcal{L}U_3 = F_3$ and the compatibility condition is $\int_0^{2\pi} F_3 \partial_\theta U_0 \mathrm{d}\theta = 0$. It is independent of λ_2 and gives an equation for λ_1. The procedure goes on to any order and provides asymptotic solutions (5.125).

The bad point is that these asymptotic solutions are *unstable*. This has been used by G. Lebeau to prove that the Cauchy problem $\Box u + u^p$ is ill-posed in $H^s(\mathbb{R}^d)$ in the supercritical case, that is when $1 < s < \frac{d}{2} - \frac{d}{p+1}$.

THEOREM 5.38 (*[95]*). *The BKW solutions constructed in Theorem* 5.36 *are unstable.*

We just give below an idea of the mechanism of instability. The linearized equation is

$$\varepsilon^2(\partial_t^2 - \partial_x^2)\dot{u} - f'(u_{\mathrm{app}}^\varepsilon)\dot{u} = \dot{f}. \tag{5.137}$$

The potential $f'(u_{\mathrm{app}}^\varepsilon)$ is a perturbation of $\sigma^2 f'(G(\varphi/\varepsilon + \Theta))$. Taking φ as a new time variable or restricting the evolution to times $t \ll \sqrt{\varepsilon}$, one can convince oneself that the relevant model is the equation:

$$\varepsilon^2(\partial_t^2 - \partial_x^2)\dot{u} + f'(G(t/\varepsilon)\dot{u}) = \dot{f}. \tag{5.138}$$

Setting $s = t/\varepsilon$ and performing a Fourier transform in x, reduces to the analysis of the Hill operator

$$\mathcal{M}_\lambda \dot{u} := \left(\partial_s^2 + f'(G(s)) + \lambda\right)\dot{u}, \qquad \lambda = \varepsilon^2\xi^2. \tag{5.139}$$

Let $E_\lambda(s)$ denote the 2×2 matrix which describes the evolution of $(\dot{u}, \dot{u}')$ for the solutions of $\mathcal{M}_\lambda \dot{u} = 0$. Since the potential $f'(G)$ is 2π-periodic, the evolution for times $s > 2\pi$ is given by

$$E_\lambda(s) = E_\lambda(s')\,(E_\lambda(2\pi))^k\,, \qquad s = s' + 2k\pi. \tag{5.140}$$

Therefore, the behavior as $s \to \infty$ is given by the iterates M_λ^k, and depends only on the spectrum of $\mathbb{M}_\lambda$.

PROPOSITION 5.39 ([95]). *There are $\mu_0 > 0$ and $\lambda_0 > 0$ such that $e^{\mu_0} > 1$ is an eigenvalue of $\mathbb{M}_{\lambda_0}$, and for all $\lambda \geq 0$ the real part of the eigenvalues of $\mathbb{M}_\lambda$ are less than or equal to* e^{μ_0}.

This implies that there are initial data such that the homogeneous equation (5.138) with $\dot{f} = 0$ has solutions which grow as $\mathrm{e}^{\mu_0 t/\varepsilon}$, which are thus larger than any given power of ε in times $t = O(\varepsilon|\ln\varepsilon|)$. This implies that expansion in powers of ε is not robust under perturbations, and eventually this implies the nonlinear instability of the approximate solutions. For the details, see [95].

5.4. *Caustics*

A major phenomenon in multidimensional propagation is *focusing*: this occurs when the rays of geometric optics accumulate and form an envelope. The phases satisfy eikonal equations,

$$\partial_t\varphi + \lambda(t, x, \partial_x\varphi) = 0 \tag{5.141}$$

and are given by Hamilton–Jacobi theory. Generically, that is for non-planar planar phases, the rays are not parallel (and curved) and thus have an envelope which is called *the caustic set*.

When rays focus, amplitudes grow and, even in the linear case, one must change the asymptotic description. In the nonlinear case, the large amplitudes can be amplified by nonlinearities and therefore strongly nonlinear phenomena can occur.

5.4.1. Example: spherical waves Consider for instance the following wave equation in space dimension d

$$\Box\, u^\varepsilon + F(\nabla_{t,x}u^\varepsilon) = 0 \tag{5.142}$$

and weakly nonlinear oscillating solutions $u^\varepsilon \sim \varepsilon U(t, x, \varphi/\varepsilon)$ associated with one of the two phases $\varphi_\pm = t \pm |x|$. The propagation of the oscillation of the principal amplitude U_0, is given by the following equation for $V = \partial_\theta U_0$:

$$2\left(\partial_t \mp \frac{1}{|x|} x \cdot \partial_x \mp \frac{d-1}{|x|}\right) V + F(V\nabla\varphi_\pm) = 0. \tag{5.143}$$

- The rays of geometric optics are the integral curves of $\partial_t \mp \partial_r$, that is the lines $x = y \mp t\frac{y}{|y|}$. For φ_+, all the rays issuing from the circle $|y| = a$ cross, at time $t = a$, at $x = 0$. *This is the focusing case* (for positive times). The caustic set is $\mathcal{C} = \{x = 0\}$. For φ_-, the rays diverge and do not intersect, *this is the defocusing case* (for positive times).
- For linear equations ($F = 0$), *the local density of energy is preserved along the rays by the linear propagation*. This is a general phenomenon when the linear equations have conserved energy. This means that if $F = 0$, $e(t, x, \theta) := |x|^{d-1}|V(t, x, \theta)|^2$ satisfies

$$(\partial_t \mp \partial_r)e = 0. \tag{5.144}$$

In the focusing case, $|x| \to 0$ when one approaches the caustic set along a ray, and therefore the conservation of e implies that the intensity $|V| \to \infty$.

- For nonlinear equations, $\tilde{V} = r^{\frac{d-1}{2}} V$ satisfies

$$(\partial_t \mp \partial_r)\tilde{V} + r^{\frac{d-1}{2}} F\left(r^{-\frac{d-1}{2}} V\nabla\varphi_\pm\right) = 0. \tag{5.145}$$

From here, it is clear that the nonlinearity can amplify or decrease the growth of the amplitude as $r \to 0$ along the rays. We give several examples.

Example 1: Blow up mechanisms

Consider in space dimension $d = 3$, Eq. (5.142) with $F(\nabla u) = -|\partial_t u|^2 \partial_t u$. For the solution of the focussing equation (5.143) the local energy $e(t, x, \theta) := |x|^2|V(t, x, \theta)|^2$ is satisfied on the ray $x = y - t\frac{y}{|y|}$:

$$e(t, x, \theta) = \frac{e(0, y, \theta)}{1 - \frac{te(0,y,\theta)}{|x||y|}}, \qquad y := x + t\frac{x}{|x|}. \tag{5.146}$$

Therefore, if $V(0, y, \theta) \neq 0$, even if it is very small, e and V blow up on the ray starting at y, *before reaching the caustic set* $x = 0$.

Two mechanisms are conjugated in this example. The first, is due to the nonlinearity $-|V|^2 V$, as in the ordinary differential equation $V' - |V|^2 V = 0$. But for this equation, the blow up time depends on the size of the data. It is very large when the data are small. The second mechanism is the amplification caused by focusing, as in the linear transport equation $\partial_t V - \partial_{|x|} V - \frac{1}{|x|} V = 0$. Even if the data is small, it forces V to be very large before one reaches the caustic. Then the first mechanism is launched and the solution blows up very quickly.

Note that the blow up occurs not only in L^∞ but also in L^1, *before the first time of focusing*. It proves that for solutions u^ε of (5.142), the principal term in the geometric optics expansion of ∇u^ε blows up. This does not prove that ∇u^ε itself becomes infinite for a fixed ε, but at least that it becomes arbitrarily large. This shows that focusing is an essential part of the mechanism which produces large solutions in a finite time bounded independently of the smallness of the data.

One can modify the example to provide an example where *the exact solution is not extendable, even in the weak sense, after the first time of focusing*. Consider in dimension $d = 5$, Eq. (5.142) with $F(\nabla u) = (\partial_t u^\varepsilon)^2 - |\nabla_x u^\varepsilon|^2$ and initial data supported in the annulus $\{1 < |x| < 2\}$:

$$u^\varepsilon(0, x) = 0, \qquad \partial_t u^\varepsilon(0, x) = U_1(x, |x|/\varepsilon),$$

Nirenberg's linearization, $v^\varepsilon := 1 - \exp(-u^\varepsilon)$, transforms the equation into the linear initial value problem

$$\Box v^\varepsilon = 0, \qquad v^\varepsilon(0, x) = 0, \qquad \partial_t v^\varepsilon(0, x) = U_1(x, |x|/\varepsilon).$$

The solution v^ε is defined for all time and vanishes in the cone $\mathcal{C} := \{|x| \le 1 - t\}$, but the domain of definition of u^ε is Ω_ε, defined as the connected open subset of $\{v^\varepsilon < 1\}$, which contains $\mathcal{C}$. Focusing make v^ε arbitrarily large on the focus line $\{x = 0\}$, or times t arbitrarily close to 1 as $\varepsilon \to 0$. One can choose data such that for all $\delta > 0$, if ε is small enough, then the sphere B_δ, of radius δ and centered at $(t = 1, x = 0)$ (the first focusing point), is not contained in Ω_ε for $\varepsilon < \varepsilon_0$, and $(\partial_t u^\varepsilon)^2 - |\nabla_x u^\varepsilon|^2$ is not integrable on $\Omega_\varepsilon \cap B_\delta$ (up to the boundary $\partial\Omega_\varepsilon$). In particular, u^ε *cannot* be extended to B_δ, even as a *weak* solution. Thus the first focusing time is the largest common lifetime of the (weak) solutions u^ε.

These examples taken from [76,79] illustrate that focusing and blow up can be created in the principal oscillations themselves. This is called *direct focusing* in [76]. But nonlinear interactions make the problem much harder. Focusing and blow up can be created by phases not present in the principal term of the expansion, but which are generated after several interactions. This phenomenon is explored in detail in [76] where it is called *hidden focusing*.

Example 2: Absorption of oscillations

When combined with strongly dissipative mechanisms, focusing can lead to a complete absorption of oscillations, in finite time. The oscillations disappear when they reach the caustic set. The following example of such a behaviour is taken from [78]. Consider in dimension $d = 3$, the dissipative wave equation

$$\Box u + |\partial_t u|^2 \partial_t u = 0,$$

with oscillating initial data

$$u^\varepsilon(0, x) = \varepsilon U_0(x, |x|/\varepsilon), \qquad \partial_t u^\varepsilon(0, x) = U_1(x, |x|/\varepsilon),$$

There are global weak solutions $u^\varepsilon \in C^0([0,\infty[; H^1)$ with $\partial_t u^\varepsilon \in C^0([0,\infty[; L^2) \cap L^4([0,\infty[\times\mathbb{R}^4)$ (see [99]). When U_0 is even in θ and $U_1 = \partial_\theta U_0$, the asymptotic expansion of u^ε is given by

$$u^\varepsilon(t,x) = \varepsilon U_0(t,x,(|x|+t)/\varepsilon) + \varepsilon^2 \dots .$$

and $V := \partial_\theta U_0$ satisfies the transport equation (5.145), which implies that the density of energy $e(t,x,\theta) := |x|^2 |V(t,x,\theta)|^2$ is also satisfied (compare with (5.146)):

$$e(t,x,\theta) = \frac{e(0,y,\theta)}{1 + \frac{te(0,y,\theta)}{|x||y|}}, \quad y := x + t\frac{x}{|x|}.$$

If $V(0,y,\theta) \neq 0$, then along the ray $x = y - t\frac{y}{|y|}$,

$$|V| \to +\infty, \quad \text{but} \quad e \to +0 \quad \text{as} \quad |x| \to 0.$$

This means that the amplitude V is infinite at the caustic, but the density of energy transported by the oscillations tends to zero, that is, it is entirely dissipated. This suggests that the oscillations are absorbed when they reach the caustic set. This is rigourously proved in [78].

5.4.2. Focusing before caustics Example 1 of the preceding section illustrates a general phenomenon. Consider a semi-linear equation and a phase φ solution of the eikonal equation (5.141) associated with a simple eigenvalue $\lambda(t,x,\xi)$ with initial value φ_0. The solution φ is constructed by the Hamilton–Jacobi method: the graph of $\partial_x\varphi$ at time t is the flow out Λ of $\partial_x\varphi_0$ by H_λ, the Hamiltonian field of λ. The phase φ is defined and smooth as long as the projection $(x,\xi) \mapsto x$ from $\Lambda \subset \mathbb{R}^d \times \mathbb{R}^d$ to $\mathbb{R}^d$ is a diffeomorphism. This is equivalent to requiring that the flow of the X_φ, the projection of $H_{\lambda|\Lambda}$, is a diffeomorphism. This ceases to be true at the envelope of the integral curves of X_φ, which is called the caustic set $\mathcal{C}$. The transport equation for amplitudes given by geometric optics is of the form

$$\partial_t U_0 + \sum \partial_{\xi_j}(t,x,\partial_x\varphi)\partial_{x_j} U_0 + c(t,x)U_0 = F(U_0). \tag{5.147}$$

The coefficient c is precisely singular on the caustic set $\mathcal{C}$ and becomes singular along the rays when one approaches $\mathcal{C}$. Therefore the discussion presented in the example (5.143) can be repeated in this more general context.

5.4.3. Oscillations past caustics

The linear case

In the linear theory, the exact solutions are smooth and bounded. This means that for high frequency waves, focusing creates large but not infinite intensities. This is a classical device to prove that mutli-dimensional hyperbolic equations do not preserve L^p norms except for $p = 2$. One idea for studying the behaviour of oscillations near caustics, is to replace phase-amplitude expansions like (5.9) by *Lagrangian integrals* (see [43,64,65]). For constant coefficient equations, they are of the form

$$(2\pi\varepsilon)^{-d} \iint e^{i\Phi(t,x,y,\xi)/\varepsilon} a(t,y)\, dy\, d\xi, \tag{5.148}$$

where the phase

$$\Phi(t,x,y,\xi) := -t\lambda(\xi) + (x-y)\cdot\xi + \psi(y)$$

is associated with a smooth piece of the characteristic variety given by the equation $\tau + \lambda(\xi) = 0$ (see the representation (5.2)). These integrals are associated with Lagrangian manifolds, $\Lambda \subset T^*\mathbb{R}^{1+d}$, which are the union of the bi-characteristics

$$x = y + t\nabla_\xi\lambda(\xi), \qquad \xi = d\psi(y), \qquad \tau = -\lambda(\xi). \tag{5.149}$$

The projections in (t,x) of these lines are the rays of optics. For small times and as long as it makes sense, Hamilton–Jacobi theory tells us that Λ is the graph of $d\varphi$ where φ is the solution of the eikonal equation

$$\partial_t\varphi + \lambda(\partial_x\varphi) = 0, \qquad \varphi_{|t=0} = \psi.$$

The projection from Λ to the base is nonsingular at points where the Hessian $\partial^2_{y,\xi}\Phi$ of Φ with respect to the variables (y,ξ) is nonsingular. Therefore the caustic set $\mathcal{C}$ associated with Λ is

$$\mathcal{C} := \{(t,x) : \exists y : x = y + tg(y) \quad \text{and} \quad \Delta(t,y) = 0\} \tag{5.150}$$

where $\Delta(t,y) = \left|\det \partial^2_{y,\xi}\Phi\right|^{\frac{1}{2}}$ and $g(y) = \nabla_\xi\lambda(d\psi(y))$.

The local behaviour of integrals (5.148) is given by stationary phase expansions. Outside the caustic set, one recovers geometric optics expansions (5.9). The principal term is

$$\sum_{\{y|y+tg(y)=x\}} \frac{1}{\Delta(t,y)} i^{m(t,y)} a(t,y) e^{i\psi(y)/\varepsilon} \tag{5.151}$$

where $2m(t,y)$ is the signature of $\partial^2_{y,\xi}\Phi$. For small times, for each (t,x) there is one critical point y, $\psi(y) = \varphi(t,x)$ and $m = 0$. When one approaches the caustic set along the ray, the amplitude tends to infinity, as expected, since $\Delta \to 0$, but *after the caustic along that ray, one recovers a similar expansion, where the phase has experienced a shift equal to* $m\frac{\pi}{2}$.

This explains how the geometric optics expansions (5.9) can be recovered from the integral representation (5.148) and how this representation resolves the singularity present in the geometric optics description. It also explains how geometric optics expansions are recovered *after* the caustic.

The detailed behavior of oscillatory integrals near the caustic is a complicated and delicate problem (see [43]). In particular, one issue is to get sharp L^∞ bounds, that is the maximum value of intensities, depending on ε. In the simplest case when the singularity of the projection of the Lagrangian Λ is a fold, there is good nonsingular asymptotic description using *Airy functions* (see e.g. [101,65]):

$$u^\varepsilon = \Big\{\varepsilon^{-\frac{1}{6}} a_\varepsilon(t,x) \mathrm{Ai}(\sigma/\varepsilon^{\frac{2}{3}}) + \varepsilon^{\frac{1}{6}} b_\varepsilon(t,x) \mathrm{Ai}'(\sigma/\varepsilon^{\frac{2}{3}})\Big\} \mathrm{e}^{\mathrm{i}\rho/\varepsilon} \tag{5.152}$$

where σ and ρ are two phase functions and the amplitudes a_ε and b_ε have asymptotic expansions $\sum \varepsilon^n a_n$ and $\sum \varepsilon^n b_n$, respectively. Moreover, the caustic set is $\{\sigma = 0\}$, and using the asymptotic expansions of the Airy function Ai and its derivative Ai′, one finds that $u^\varepsilon = O(\varepsilon^\infty)$ in any compact subset of the shadow region $\{\sigma > 0\}$, and that u^ε is the superposition of two wave trains (5.9) in any compact subset of the illuminated region $\{\sigma < 0\}$ associated with the phases $\rho \pm \frac{2}{3}(-\sigma)^{\frac{3}{2}}$.

The nonlinear case

• The extension of the previous results to nonlinear equations is not easy and most questions remain open. An important difficulty is that products of an oscillatory integral like (5.148) involve integrals with more and more variables, hinting that a similar representation of solutions of nonlinear equations would require an infinite number of variables. Thus this path seems difficult to follow, except if one has some reason to restrict the analysis to the principal term of the integrals.

This has been done for semi-linear equations with constant coefficients in two cases: for *dissipative* nonlinearities or when the nonlinear source term *is globally Lipschitzean*. In these two cases, the structure of the equation implies that one can analyze the exact solution in low regularity spaces such as L^2 or L^p and provide asymptotic expansions using oscillatory integrals with small errors *in these spaces only* (see [79,80]). The natural extension of (5.148) is

$$I^\varepsilon(A) = \frac{1}{(2\pi\varepsilon)^d} \iint \mathrm{e}^{-\mathrm{i}(t\lambda(\xi)+(x-y)\xi)} A(t, y, \psi(y)/\varepsilon)\, \mathrm{d}y\, \mathrm{d}\xi,$$

associated with profiles

$$A(t, y, \theta) = \sum_{n\neq 0} a_n(t, y) \mathrm{e}^{\mathrm{i}n\theta}.$$

There are *transport equations for the profile* A, which are *o.d.e along the lifted rays* $y + tg(y)$ *in the Lagrangian* Λ. The phenomenon of absorption is interpreted as the property that $A = 0$ past the caustic. It occurs if the dissipation is strong enough. For globally Lipchitzean nonlinearities or weak dissipation, the profile A is continued past the caustic, providing geometric optics approximations at the leading order. Note that in the analogue of (5.151), the multiplication by i^m of the amplitude a has to be replaced by $\mathcal{H}^m\mathcal{A}$, where $\mathcal{H}$ is the Hilbert transform of the Fourier series

$$\mathcal{H}\left(\sum_{n\neq 0} a_n \mathrm{e}^{\mathrm{i}n\theta}\right) := \sum_{n\neq 0} i^{\operatorname{sign} n} a_n \mathrm{e}^{\mathrm{i}n\theta}.$$

The very weak point of this analysis is that is leaves out completely the case of *conservative* non-dissipative equations.

• In the same spirit, the use of representations such as (5.152) using Airy functions is not obvious, since a product of Airy functions is not a Airy function. Such an approach is suggested by J. Hunter and J. Keller in [69] as a technical device to match the geometric optics expansions before and after the caustics. Another output of their analysis is a formal classification of the qualitative properties of weakly nonlinear geometric optics, separating linear and nonlinear propagation, and linear and nonlinear effects of the caustic. R. Carles has rigorously explained this classification, mainly for spherical waves, for the wave equations

$$(\partial_t^2 - \Delta_x)u + a|\partial_t u|^{p-1}\partial_t u = 0, \quad p > 1; a \in \mathbb{C}, \tag{5.153}$$

($a > 0, a < 0, a \in \mathrm{i}\mathbb{R}$ corresponding to the dissipative, accretive and conservative case, respectively) or the semiclassical nonlinear Schrödinger equation (NLS),

$$\mathrm{i}\varepsilon\partial_t u + \frac{1}{2}\varepsilon^2\Delta_x u = \varepsilon^\alpha |u|^{2\sigma}u, \qquad u_{|t=0} = f(x)\mathrm{e}^{-|x|^2/2\varepsilon} \tag{5.154}$$

with $\alpha \geq 1, \sigma > 0$.

In [12], for $1 < p < 2$, the reader can find an L^∞ *description near* $x = 0$ of radial waves in $\mathbb{R}^3$ for (5.153). As expected, the profiles and the solutions are unbounded (uniformly in ε) and *new amplitudes* (of size ε^{1-p}) must be added to the one predicted in [69] as correctors near the caustic in order to have a uniform approximation in L^∞. In particular, this gives the correct evaluation of the exact solutions at the caustic set $\{x = 0\}$.

Concerning NLS, R. Carles has investigated all the behaviors with different powers $\alpha \geq 1$ and $\sigma > 0$.

- When $\alpha > d\sigma$ and $\alpha > 1$, the propagation of the main term ignores the nonlinearity, outside the caustic and at the caustic.
- When $\alpha = 1 > d\sigma$, the propagation before and after the caustic follows the rule of weakly nonlinear optics, and the matching at the focal point is like in the linear case.
- When $\alpha = d\sigma > 1$, the nonlinear effects take place only near the focal point. A remarkable new idea is that *the transition between the amplitudes before and after the focal point is given a scattering operator*.

We refer the reader to [13–15] for details.

6. Diffractive optics

The regime of geometric optics is not adapted to the propagation of laser beams in distances that are much larger than the width of the beam. The standard descriptions that can be found in physics text books involve Schrödinger equations, used to model the diffraction of light in the direction transversal to the beam along long distances. The goal of this section is to clarify the nature and range of validity of the approximations leading to Schrödinger-like equations, using simple models and examples.

6.1. *The origin of the Schrödinger equation*

6.1.1. A linear example Consider again a symmetric hyperbolic linear system (5.1) with constant coefficients and eigenvalues of constant multiplicities. Consider oscillatory initial data (5.3). The solution is computed by Fourier synthesis, as explained in (5.4), revealing integrals

$$\begin{aligned} u_p^\varepsilon(t,x) &= \frac{1}{(2\pi)^{d/2}} \int \mathrm{e}^{\mathrm{i}\{y(k+\varepsilon\xi)-t\lambda_p(k+\varepsilon\xi)\}/\varepsilon} \Pi_p(k+\varepsilon\xi)\hat{h}(\xi)\mathrm{d}\xi \\ &= \mathrm{e}^{\mathrm{i}(kx-t\omega_p)/\varepsilon} a_p^\varepsilon(t,x). \end{aligned} \tag{6.1}$$

The error estimates (5.6) show that the approximation of geometric optics is valid as long as $t = o(1/\varepsilon)$. To study times and distances of order $1/\varepsilon$, insert the second order Taylor expansion

$$\lambda_p(k+\varepsilon\xi) = \omega_p + \varepsilon\xi\mathbf{v}_p + \varepsilon^2 S_p(\xi) + O(\varepsilon^3|\xi|^3),$$

with

$$S_p(\xi) := \frac{1}{2}\sum_{j,k} \partial^2_{\xi_j,\xi_k}\lambda_p(k)\xi_j\xi_k, \tag{6.2}$$

into the integral (6.1) to find

$$a_p^\varepsilon(t,x) = \frac{1}{(2\pi)^{d/2}} \int \mathrm{e}^{\mathrm{i}(x\xi - t\xi\mathbf{v}_p)} \mathrm{e}^{-\mathrm{i}\varepsilon t S_p(\xi)} \Pi_p(k)\hat{h}(\xi)\mathrm{d}\xi + O(\varepsilon^2 t) + O(\varepsilon).$$

Introduce the slow variable $T := \varepsilon t$ and

$$a_{p,0}(T,t,y) := \frac{1}{(2\pi)^{d/2}} \int \mathrm{e}^{\mathrm{i}(x\xi - t\xi\mathbf{v}_p)} \mathrm{e}^{-\mathrm{i}T S_p(\xi)} \Pi_p(k)\hat{h}(\xi)\mathrm{d}\xi. \tag{6.3}$$

Then

$$\left\| a_p^\varepsilon(t,\cdot) - a_{p,0}(\varepsilon t, t, \cdot) \right\|_{H^s(\mathbb{R}^d)} \le C(\varepsilon^2 t + \varepsilon)\, \|h\|_{H^{s+3}(\mathbb{R}^d)}\,. \tag{6.4}$$

The profile $a_{p,0}$ satisfies the polarization conditions $a_{p,0} = \Pi_p(k)a_{p,0}$ and a pair of partial differential equations:

$$(\partial_t + \mathbf{v}_p \cdot \nabla_x)a_{p,0} = 0, \quad \text{and} \quad \big(\mathrm{i}\partial_T + S_p(\partial_x)\big)\, a_{p,0} = 0. \tag{6.5}$$

The first is the transport equation of geometric optics and the second is the Schrödinger equation, with *evolution governed by the slow time T*, which we were looking for. These, together with the initial condition

$$a_{p,0}(0, 0, y) = f(y),$$

suffice to uniquely determine $a_{p,0}$.

Note that for $t = o(1/\varepsilon)$ one has $T \to 0$, and setting $T = 0$ in (6.3) one recovers the approximation of geometric optics. A typical solution of the Schödinger equation has spatial width which grows linearly in T (think of Gaussian solutions). Thus the typical width of our solution u^ε grows linearly in εt, which is consistent with the geometric observation that the wave vectors comprising $u_{|t=0} = h\mathrm{e}^{\mathrm{i}kx/\varepsilon}$ make an angle $O(\varepsilon)$ with k. The approximation (6.4) clearly presents three scales; the wavelength ε, the lengths of order 1 on which f varies, and, the lengths of order $1/\varepsilon$ traveled by the wave on the time scales of the variations of a_p with respect to the slow time T.

In contrast to the case of the geometric optics expansion, the analysis above does not extend to non-planar phases $\varphi(t, x)/\varepsilon$. Note that the rays associated with nonlinear phases focus or spread in times $O(1)$. The case of focusing of rays and its consequences for nonlinear waves was briefly discussed in Section 5. On the other hand, with suitable convexity hypotheses on the wavefronts, nonlinear transport equations along spreading rays yield nonlinear geometric optics approximations valid globally in time (see [56]). In the same way, one does not find such Schrödinger approximations for linear phases when the geometric approximations are not governed by transport equations. The classical example is conical refraction.

However, the analysis does extend to phases $(kx - \omega t + \psi(\varepsilon t, \varepsilon x))/\varepsilon$ with *slowly varying* differentials, observing that for times $t \lesssim \varepsilon^{-1}$ the rays remain almost parallel. We refer to [67,45] for these extensions.

CONCLUSION 6.1. *The Schrödinger approximations provide diffractive corrections for times $t \sim 1/\varepsilon$ to solutions of wavelength ε which are adequately described by geometric optics with parallel rays for times $t \sim 1$.*

These key features, parallel rays and three scales, are commonly satisfied by laser beams. The beam is comprised of virtually parallel rays. A typical example with three scales would have wavelength $\sim 10^{-6}$ m, the width of the beam $\sim 10^{-3}$ m, and the propagation distance ~ 1 m.

6.1.2. Formulating the Ansatz For simplicity, we study solutions of semi-linear symmetric hyperbolic systems with constant coefficients and nonlinearity which is or order J

near $u = 0$. The quasi-linear case can be treated similarly. Consider

$$L(\varepsilon\partial) = \varepsilon A_0 \partial_t + \sum_{j=1}^{d} \varepsilon A_j \partial_{x_j} + E \tag{6.6}$$

with self-adjoint matrices A_j, A_0 being positive definite, and with E skew-adjoint. The nonlinear differential equation to solve is

$$L^{\varepsilon}(\partial_x) u^{\varepsilon} + F(u^{\varepsilon}) = 0 \tag{6.7}$$

where u^{ε} is a family of $\mathbb{C}^N$ valued functions. The nonlinear interaction term F is assumed to be a smooth nonlinear function of u and of order $J \geq 2$ at the origin, in the sense that the Taylor expansion at the origin satisfies

$$F(u) = \Phi(u) + O(|u|^{J+1}) \tag{6.8}$$

where Φ is a homogeneous polynomial of degree J in $u, \overline{u}$.

Time of nonlinear interaction and amplitude of the solution. The amplitude of nonlinear waves in crucially important. Our solutions have amplitude $u^{\varepsilon} = O(\varepsilon^p)$ where the exponent $p > 0$ is chosen so that the nonlinear term $F(u) = O(\varepsilon^{pJ})$ affects the principal term in the asymptotic expansion for times of order $1/\varepsilon$. The time of nonlinear interaction is comparable to the times for the onset of diffractive effects.

The time of nonlinear interaction is estimated as follows. Denote by $S(t)$ the propagator for the linear operator L^{ε}. Then in $L^2(\mathbb{R}^d)$, $\|S(t)\| \leq C$. The Duhamel representation

$$u(t) = S(t)u(0) - \int_0^t S(t-\sigma) F(u(\sigma))\, \mathrm{d}\sigma$$

suggests that the contribution of the nonlinear term at time t is of order $t\varepsilon^{pJ}$. For the onset of diffraction, $t \sim 1/\varepsilon$ so the accumulated effect is expected to be $O(\varepsilon^{pJ-2})$ (note that the coefficient of ∂_t is ε). For this to be comparable to the size of the solution *we choose p so that $pJ - 2 = p$*:

$$p = \frac{2}{J-1}. \tag{6.9}$$

The basic Ansatz. It has three scales

$$u^{\varepsilon}(x) = \varepsilon^p a(\varepsilon, \varepsilon t, \varepsilon x, t, x, \varphi(t,x)/\varepsilon) \tag{6.10}$$

where $\varphi(t,x) = kx - t\omega$ and $a(\varepsilon, T, X, t, x, \theta)$ is periodic in θ and has an expansion:

$$a = a_0 + \varepsilon a_1 + \varepsilon^2 a_2 + \dots. \tag{6.11}$$

Since this expansion is used for times $t \sim 1/\varepsilon$, in order for the correctors εa_1 and $\varepsilon^2 a_2$ to be smaller than the principal term for such times, one must control the growth of the profiles in t. The most favorable case would be that they are uniformly bounded for $\varepsilon t \leq T_0$, but this is not always the case as will be shown in the analysis below. Therefore we *impose the weaker condition of sublinear growth*:

$$\lim_{t\to\infty} \frac{1}{t} \sup_{T,X,y,\theta} |a_1(T, X, t, x, \theta)| = 0. \tag{6.12}$$

This condition plays a central role in the analysis to follow.

6.1.3. First equations for the profiles Plugging (6.10) into the equation and ordering in powers of ε yields the equations

$$\mathcal{L}(\beta\partial_\theta)a_0 = 0, \tag{6.13}$$

$$\mathcal{L}(\beta\partial_\theta)a_1 + L_1(\partial_{t,x})a_0 = 0, \tag{6.14}$$

$$\mathcal{L}(\beta\partial_\theta)a_2 + L_1(\partial_{t,x})a_1 + L_1(\partial_{T,X})a_0 + \Phi(a_0) = 0 \tag{6.15}$$

with $\beta = \mathrm{d}\varphi = (-\omega, k)$ and $\mathcal{L}(\beta\partial_\theta) = L_1(\beta)\partial_\theta + E$.

Analysis of Eqs (6.13) *and* (6.14). The first two equations are those met in the geometric optics regime. They are analyzed using Fourier series expansions in θ. We make the following assumptions, satisfied in many applications:

ASSUMPTION 6.2. (i) β satisfies the dispersion relation

$$\det(L_1(\beta) - \mathrm{i}E) = 0. \tag{6.16}$$

Denote by Z the set of harmonics $n \in \mathbb{Z}$ such that $\det(L_1(n\beta) - \mathrm{i}E) = 0$.

(ii) If $E \neq 0$, $\det(L_1(\beta)) \neq 0$, so that Z is finite.

(iii) For $n \in Z$, $n \neq 0$, the point $n\beta$ is a regular point in the characteristic set in the sense of Definitions 5.3.

As in Section 5, we denote by $\mathcal{P}$ and $\mathcal{Q}$ projectors on the kernel and the image of $\mathcal{L}(\beta\partial_\theta)$, respectively, and by $\mathcal{R}$ the partial inverse such that

$$\mathcal{R}\mathcal{L}(\beta\partial_\theta) = \mathrm{Id} - \mathcal{P}, \qquad \mathcal{P}\mathcal{R} = 0, \qquad \mathcal{R}(\mathrm{Id} - \mathcal{Q}) = 0. \tag{6.17}$$

They are defined by projectors P_n and Q_n and partial inverses R_n acting on the Fourier coefficients $\hat{a}_n$ of Fourier series $\sum \hat{a}_n \mathrm{e}^{\mathrm{i}n\theta}$.

The first equation asserts that the principal profile is polarized along the kernel of $\mathcal{L}(\beta\partial_\theta)$:

$$a_0 = \mathcal{P}a_0 \Leftrightarrow \forall n, \hat{a}_{0,n} = P_n \hat{a}_{0,n}. \tag{6.18}$$

The second equation is equivalent to

$$(\mathrm{Id} - \mathcal{Q})L_1(\partial_{t,x})\mathcal{P}a_0 = 0, \tag{6.19}$$

$$(\mathrm{Id} - \mathcal{P})a_1 = -\mathcal{R}L_1(\partial_{t,x})\mathcal{P}a_0. \tag{6.20}$$

Analysis of Eq. (6.15). This is the new part of the analysis. Applying the projectors $\mathcal{Q}$ and $\mathrm{Id} - \mathcal{Q}$ and using (6.20) the equation is equivalent to

$$-\mathcal{T}(\partial_{t,x})a_1 = \mathcal{S}(\partial_{T,X}, \partial_{t,x})a_0 + (\mathrm{Id} - \mathcal{Q})\Phi(a_0), \tag{6.21}$$

$$(\mathrm{Id} - \mathcal{P})a_2 = -\mathcal{R}\left(L_1(\partial_{t,x})a_1 + L_1(\partial_{T,X})a_0 + \Phi(a_0)\right) \tag{6.22}$$

with

$$\mathcal{T}(\partial_{t,x}) = (\mathrm{Id} - \mathcal{Q})L_1(\partial_{t,x})\mathcal{P} \tag{6.23}$$

$$\mathcal{S}(\partial_{T,X}, \partial_{t,x}) = (\mathrm{Id} - \mathcal{Q})L_1(\partial_{T,X})\mathcal{P} - (\mathrm{Id} - \mathcal{Q})L_1(\partial_{t,x})\mathcal{R}L_1(\partial_{t,x})\mathcal{P}. \tag{6.24}$$

CONCLUSION 6.3. *One has to determine the range of $\mathcal{T}(\partial_{t,x})$, acting in a space of functions satisfying the sublinear growth condition (6.12), so that a_0 will be determined by the polarization condition (6.19), the propagation equation (6.21) in the intermediate variables, and the additional equation:*

$$\mathcal{S}(\partial_{T,X}, \partial_{t,x})a_0 - (I - \mathcal{Q})\Phi(a_0) \in \mathrm{Range}\,\mathcal{T}(\partial_{t,x}). \tag{6.25}$$

In Section 6.2 we investigate this question in different situations with increasing order of complexity. Before that, we make several general remarks about the operator $\mathcal{S}$.

6.1.4. The operator $\mathcal{S}$ as a Schrödinger operator The linear operators $\mathcal{T}(\partial_{t,x})$ and $\mathcal{S}(\partial_{T,X}, \partial_{t,x})$ act term by term on Fourier series in θ. On the nth coefficient, we obtain the operators

$$\widehat{\mathcal{T}}_n(\partial_{t,x}) = (\mathrm{Id} - Q_n)L_1(\partial_{t,x})P_n, \tag{6.26}$$

$$\begin{aligned}\widehat{\mathcal{S}}_n(\partial_{T,X}, \partial_{t,x}) = {} & (\mathrm{Id} - Q_n)L_1(\partial_{T,X})P_n \\ & -(\mathrm{Id} - Q_n)L_1(\partial_{t,x})R_nL_1(\partial_{t,x})P_n.\end{aligned} \tag{6.27}$$

Following Assumption 6.2, for $n \neq 0$, $n\beta$ is a regular point in the characteristic variety $\mathcal{C}$, meaning that near $n\beta$, $\mathcal{C}$ is given by the equation

$$\tau + \lambda_n(\xi) = 0 \tag{6.28}$$

where λ_n is an eigenvalue of constant multiplicity. In particular $-n\omega + \lambda_n(nk) = 0$.

PROPOSITION 6.4. (i) *Let $X_n(\partial_{t,x}) = \partial_t + \mathbf{v}_n \cdot \nabla_x$ with $\mathbf{v}_n = \nabla_\xi \lambda_n(nk)$. Then*

$$\widehat{\mathcal{T}}_n(\partial_{t,x}) = (\partial_t + \mathbf{v}_n \cdot \nabla_x)J_n, \qquad J_n := (\mathrm{Id} - Q_n)A_0P_n. \tag{6.29}$$

(ii) *Let*

$$S_n(\partial_x) := \frac{1}{2}\sum_{j,k} \partial^2_{\xi_j\xi_k}\lambda_n(nk)\partial_{x_j}\partial_{x_k}. \tag{6.30}$$

Then

$$\widehat{\mathcal{S}}_n(\partial_{T,X}, \partial_{t,x}) = (\partial_T + \mathbf{v}_n \cdot \nabla_X - \mathrm{i}S_n(\partial_x))\, J_n + \widehat{\mathcal{R}}_n(\partial_{t,x})X_n(\partial_{t,x}) \tag{6.31}$$

where $\widehat{\mathcal{R}}_n$ *is a first order operator which is specified below.*

PROOF. Consider $n = 1$ and drop the indices n. The first part is proved in Lemma 5.9. Recall that differentiating the identity

$$\left(-\mathrm{i}\lambda(\xi)A_0 + \sum_j \mathrm{i}\xi_j A_j + E\right) P(\xi) = 0$$

once, yields

$$\left(-\mathrm{i}\lambda(\xi)A_0 + \sum_j \mathrm{i}\xi_j A_j + E\right) \partial_{\xi_p} P(\xi) = -\mathrm{i}(-\partial_{\xi_p}\lambda(\xi)A_0 + A_p)P(\xi).$$

Multiplying on the left by $\mathrm{Id} - Q(k)$ implies

$$(\mathrm{Id} - Q(k))(-\partial_{\xi_p}\lambda(k)A_0 + A_p)P(k) = 0, \tag{6.32}$$

$$(\mathrm{Id} - P(k))\partial_{\xi_p}P(k) = -\mathrm{i}R(k)(-\partial_{\xi_p}\lambda(k)A_0 + A_p)P(k). \tag{6.33}$$

The first identity implies (6.29).

Differentiating the identity again and multiplying on the left by $\mathrm{Id} - Q(k)$ implies that

$$\begin{aligned}&(\mathrm{Id} - Q)(-\partial_{\xi_p}\lambda A_0 + A_p)\partial_{\xi_q}P + (\mathrm{Id} - Q)(-\partial_{\xi_q}\lambda A_0 + A_q)\partial_{\xi_p}P\\ &\quad = \partial^2_{\xi_p,\xi_q}\lambda(\mathrm{Id} - Q)A_0P\end{aligned}$$

where the functions are evaluated at $\xi = k$. Using (6.32), one can replace $\partial_{\xi_p}P$ by $(\mathrm{Id} - P)\partial_{\xi_p}p$ in the identity above. With (6.33), this implies that

$$\begin{aligned}&(\mathrm{Id} - Q)(-\partial_{\xi_p}\lambda A_0 + A_p)R(-\partial_{\xi_q}\lambda A_0 + A_q)P\\ &\quad + (\mathrm{Id} - Q)(-\partial_{\xi_q}\lambda A_0 + A_q)R(-\partial_{\xi_p}\lambda A_0 + A_p)P\\ &\qquad = \mathrm{i}\partial^2_{\xi_p,\xi_q}\lambda\ (\mathrm{Id} - Q)A_0P.\end{aligned}$$

Together with the identity

$$L_1(\partial_t, \partial_x) = A_0(\partial_t + \mathbf{v}\cdot\nabla_x) + \sum_{p=1}^{d}(A_p - \partial_{\xi_p}\lambda A_0)\partial_{x_p} = A_0X(\partial_{t,x}) + L_1'(\partial_x)$$

this implies that

$$\begin{aligned}(\mathrm{Id}-Q)L_1(\partial_{t,x})RL_1(\partial_{t,x})P &= (\mathrm{Id}-Q)X(\partial_{t,x})RL_1(\partial_{t,x})P\\ &\quad+(\mathrm{Id}-Q)L_1'(\partial_x)RX(\partial_{t,x})P\\ &\quad+\frac{\mathrm{i}}{2}\lambda\,(\mathrm{Id}-Q)A_0P\sum\partial^2_{\xi_p,\xi_q}\partial_{x_p}\partial_{x_q}\end{aligned}$$

implying (6.31). □

COROLLARY 6.5. *If $a(T,X,t,x)$ satisfies the polarization condition $a=P_na$ and the transport equation $X(\partial_{t,x}a)=0$, then*

$$\widehat{\mathcal{S}}_n(\partial_{T,X},\partial_{t,x})a=J_n\left(\partial_T+\mathbf{v}\cdot\nabla_X+\mathrm{i}S_n(\partial_x)\right)a.$$

The conclusion is that for all frequency n, the operator $\hat{\mathcal{S}}_n$ acting on the polarized vectors a is the Shrödinger operator $\partial_T+\mathbf{v}\cdot\nabla_X+\mathrm{i}S_n(\partial_x)$, that is, for functions independent of X, the operator (6.5).

6.2. *Construction of solutions*

We consider first the simplest case:

6.2.1. The non-dispersive case with odd nonlinearities and odd profiles We assume here that in Eq. (6.7), $E=0$ and $F(u)$ *is an odd function* of u with a nontrivial cubic term, that is with $J=3$, so that we choose $p=1$ according to the rule (6.9). We look for solutions

$$u^\varepsilon(t,x)\sim\varepsilon\sum_{n\geq 0}\varepsilon^n a_n(\varepsilon t,t,x,\varphi(t,x)/\varepsilon), \tag{6.34}$$

where $\varphi(t,x)=kx-t\omega$ and the profiles $a_n(T,t,x,\theta)$ are *periodic and odd in* θ.

We assume that φ is a characteristic phase and, more precisely, we assume that $d\varphi$ is a *regular point* of the characteristic variety $\det L(\tau,\xi)=0$. We introduce projectors P and Q on the kernel and on the image of $L(d\varphi)$, respectively, and we denote by R the partial inverse of $L(d\varphi)$ with properties as above. We denote by $X(\partial_{t,x})=\partial_t+\mathbf{v}\cdot\nabla_x$ the transport operator and similarly by $S(\partial_{t,x})$ the second order operator (6.2) or (6.30) associated with $d\varphi$. By homogeneity, all the harmonics $n\varphi$, $n\neq 0$ are characteristic and one can choose the projectors P_n to be equal to P, $Q_n=Q$. Moreover, all the transport fields have the same speed and $X_n(\partial_{t,x})=X(\partial_{t,x})$. Similarly, $S_n(\partial_x)=S(\partial_x)$.

Because we *postulate* that a_0 is odd in θ, its zeroth harmonic (its mean value) vanishes. Therefore, the polarization condition (6.20) reduces to

$$a_0(T,t,x,\theta)=Pa_0(T,t,x,\theta), \tag{6.35}$$

and the first condition Eq. (6.19) reads

$$(\partial_t+\mathbf{v}\nabla_x)a_0(T,t,x,\theta)=0. \tag{6.36}$$

Because we also postulate that a_1 is odd, (6.20) is equivalent to

$$(\mathrm{Id} - P)a_1(T, t, x, \theta) = -\partial_\theta^{-1} R L_1(\partial_{t,x}) P a_0 \tag{6.37}$$

where ∂_θ^{-1} is the inverse of ∂_θ acting on functions with a vanishing mean value. Next, Eq. (6.21) is

$$\begin{aligned} -(\partial_t + \mathbf{v}\nabla_x)(\mathrm{Id} - Q)A_0 P a_1 = {} & (\mathrm{Id} - Q)\Phi(a_0) + \partial_T(\mathrm{Id} - Q)A_0 P a_0 \\ & - (\mathrm{Id} - Q)L_1(\partial_t, x)\partial_\theta^{-1} R L_1(\partial_{t,x})P, \end{aligned} \tag{6.38}$$

where we have used the following remark:

$$\text{if } \Phi \text{ is odd in } u \text{ and if } a_0 \text{ is odd in } \theta, \text{ then } \Phi(a_0) \text{ is odd in } \theta, \tag{6.39}$$

implying that the mean value of $\Phi(a_0)$ vanishes and that $\mathcal{Q}\Phi(a_0) = Q\Phi(a_0)$.

Next we note that the *scalar* operator $X(\partial_{t,x})$ commutes with all the constant coefficient operators in the right-hand side of (6.38) so that (6.36) implies that the right-hand side f_0 of (6.38) satisfies

$$X(\partial_{t,x}) f_0 = 0, \tag{6.40}$$

and the equation for $\tilde{a}_1 = (\mathrm{Id} - Q)A_0 P a_1$ reduces to

$$X(\partial_{t,x})\tilde{a}_1 = f_0. \tag{6.41}$$

By (6.40), f_0 is constant along the rays of X, and thus $\tilde{a}_1$ is affine on these rays. Therefore,

LEMMA 6.6. *If f_0 satisfies (6.40), Eq. (6.41) has a solution with sublinear growth in t, if and only if $f_0 = 0$. In this case, the solutions $\tilde{a}_1$ are bounded in time.*

Therefore, we see that Eq. (6.38) has a solution a_1 with sublinear growth in t, if and only if the following *two* equations are satisfied:

$$\begin{aligned} & (\mathrm{Id} - Q)\Phi(a_0) + \partial_T(\mathrm{Id} - Q)A_0 P a_0 \\ & \quad - (\mathrm{Id} - Q)L_1(\partial_t, x)\partial_\theta^{-1} R L_1(\partial_{t,x})P = 0, \end{aligned} \tag{6.42}$$

$$(\partial_t + \mathbf{v}\nabla_x)P a_1 = 0. \tag{6.43}$$

Moreover, using Proposition 6.4, we see that for a_0 satisfying (6.36), the first equation is equivalent to

$$\partial_T a_0 + \partial_\theta^{-1} S(\partial_x) a_0 = R_0 \Phi(a_0) \tag{6.44}$$

where R_0 is the inverse of $(\mathrm{Id} - Q)A_0 P$ from the image of $(\mathrm{Id} - Q)$ to the image of P. On the harmonic 1, $\partial_\theta^{-1} = -\mathrm{i}$, and we recover a Schrödinger equation as in (6.5), now with a source term.

PROPOSITION 6.7 (*Existence of the principal profile*). *Given a Cauchy data $h_0(x, \theta)$, in $H^s(\mathbb{R}^d \times \mathbb{T})$, with $s > \frac{d+1}{2}$, odd in θ and satisfying $h_0 = Ph_0$, there is $T_* > 0$ such that the two times problem (6.36), (6.44) has a unique solution $a_0 \in C^0([0, T_*] \times \mathbb{R}; H^s(\mathbb{R}^d \times \mathbb{T}))$, odd in θ and satisfying $a_0 = Pa_0$ and*

$$a_0(0, 0, x, \theta) = h_0(x, \theta). \tag{6.45}$$

PROOF. By the first equation we look for a_0 as a function of $x - t\mathbf{v}$

$$a_0(T, t, x, \theta) = \mathrm{a}_0(T, x - t\mathbf{v}, \theta). \tag{6.46}$$

Then the equation for $\mathrm{a}_0(T, y, \theta)$ is:

$$\partial_T \mathrm{a}_0 + \partial_\theta^{-1} S(\partial_y)\mathrm{a}_0 + R_0 \Phi(\mathrm{a}_0) = 0, \quad \mathrm{a}_0(0, y, \theta) = h(y, \theta). \tag{6.47}$$

This equation is quite similar to classical nonlinear Schrödinger equations, ∂_θ^{-1} being bounded and skew-adjoint. Thus it has a local solution in H^s, if s is large enough. The only point is to check that all the terms in the equations preserve oddness. □

Next we show that (6.37), (6.46) and (6.43) imply that a_1 is of the form

$$a_1(T, t, x, \theta) = \mathrm{a}_1(T, x - t\mathbf{v}, \theta) \tag{6.48}$$

with $(I - P)\mathrm{a}_1$ determined by a_0 and thus odd in θ.

In the expansion in powers of ε of the equation, the term coming after (6.15) is

$$\mathcal{L}(\beta\partial_\theta)a_3 + L_1(\partial_{t,x})a_2 + L_1(D_{T,X})a_1 + \Phi'(a_0)a_1 = 0$$

(Recall that F is odd, so $F(u) = \varepsilon^3 \Phi(a_0) + \varepsilon^4 \Phi_0'(a_0)a_1 + O(\varepsilon^5)$). Projecting with $(\mathrm{Id} - Q)$ and Q, using the facts that $(\mathrm{Id} - P)a_1$ is known, that Pa_1 satisfies (6.43) and that $(\mathrm{Id} - P)a_2$ is given by (6.22), we see that the equation is equivalent to

$$\begin{aligned} -(\partial_t + \mathbf{v}\nabla_x)(\mathrm{Id} - Q)A_0 P a_2 = {} & (\partial_T + \partial_\theta^{-1} S(\partial_x))(\mathrm{Id} - Q)A_0 P a_1 \\ & + (\mathrm{Id} - Q)\Phi'(a_0)Pa_1 + (\mathrm{Id} - Q)b_0 \end{aligned}$$

where b_0 is determined by a_0, together with an explicit determination of $(\mathrm{Id} - P)a_3$ in terms of a_0, a_1, a_2. In particular, $b_0(T, t, x, \theta) = \mathrm{b}_0(T, x - t\mathbf{v}, \theta)$ with b_0 odd in θ. The right-hand side f_1 satisfies $(\partial_t + \mathbf{v}\nabla_x) f_1 = 0$. Therefore, using Lemma 6.6, we see that a_2 has a sublinear growth in t if and only if

$$(\partial_t + \mathbf{v}\nabla_x)Pa_2 = 0, \tag{6.49}$$

and $Pa_1(T,t,x,\theta) = P\mathrm{a}_1(T, x - t\mathbf{v}, \theta)$ is therefore determined by the equation

$$\partial_T \mathrm{a}_1 + \partial_\theta^{-1} S(\partial_y)\mathrm{a}_1 + R_0\Phi'(\mathrm{a}_0)\mathrm{a}_1 + R_0\mathrm{b}_0 = 0. \tag{6.50}$$

The equations for the higher profiles are derived in a similar fashion. We see that the profiles a_n in (6.34) have the form $a_n(T,t,x,\theta) = \mathrm{a}_n(T, x - t\mathbf{v}, \theta)$, so that

$$u^\varepsilon(t,x) \sim \varepsilon \sum_{n\geq 0} \varepsilon^n \mathrm{a}_n(\varepsilon t, x - t\mathbf{v}, \varphi(t,x)/\varepsilon). \tag{6.51}$$

The $(I-P)\mathrm{a}_a$ are determined explicitly in terms of $(\mathrm{a}_0,\dots,\mathrm{a}_{n-1})$ and the slow dynamics for $P\mathrm{a}_n$ is determined by an equation of the form

$$\partial_T \mathrm{a}_n + \partial_\theta^{-1} S(\partial_y)\mathrm{a}_n + R_0\Phi'(\mathrm{a}_0)\mathrm{a}_n + R_0\mathrm{b}_{n-1} = 0 \tag{6.52}$$

with b_{n-1} explicitly given in terms of $(\mathrm{a}_0,\dots,\mathrm{a}_{n-1})$.

THEOREM 6.8 (*Existence of profiles at all orders*). *Suppose that for all $n \in \mathbb{N}$, initial data $h_n \in \mathcal{S}(\mathbb{R}^d \times \mathbb{T})$ are given, odd in θ and satisfying $Ph_n = h_n$. Then there exists $T_* \in]0,\infty]$ and a unique sequence $\{\mathrm{a}_n\}_{n\geq 0}$ of functions in $\mathcal{S}([0,T_*]\times\mathbb{R}^d\times\mathbb{T})$, odd in θ and satisfying the profile equations and the initial conditions $P\mathrm{a}_n(0,y,\theta) = Ph_n(y,\theta)$.*

Given such a sequence of profiles, Borel's theorem constructs

$$\mathrm{a}^\varepsilon(T,y,\theta) \sim \sum_{n\geq 0} \varepsilon^n \mathrm{a}_n(T,y,\theta)$$

in the sense that for all s and $m \in \mathbb{N}$, there holds for all $\varepsilon \in]0,1]$:

$$\Big\| \mathrm{a}^\varepsilon - \sum_{j\leq m} \varepsilon^j \mathrm{a}_j \Big\|_{H^s([0,T_*]\times\mathbb{R}^d\times\mathbb{T})} \leq C_{s,m}\varepsilon^{m+1}.$$

Then

$$u^\varepsilon_{\mathrm{app}}(t,x) := \varepsilon \mathrm{a}^\varepsilon(\varepsilon t, x - t\mathbf{v}, \varphi(t,x)/\varepsilon)$$

are approximate solutions of (6.7), in the sense that

$$L(\varepsilon\partial_{t,x})u^\varepsilon_{\mathrm{app}} + F(u^\varepsilon_{\mathrm{app}}) = \mathrm{r}^\varepsilon(\varepsilon t, x - t\mathbf{v}, \varphi(t,x)/\varepsilon),$$

with $\mathrm{r}^\varepsilon \sim 0$ in the sense above.

Using Theorem 4.21 ([59], see also [39,40]), one can produce exact solutions which have the asymptotic expansion (6.51). Consider initial data

$$h^\varepsilon(x) \sim u^\varepsilon_{\mathrm{app}}(0,x) \sim \varepsilon \sum_{n\geq 0} \varepsilon^n \mathrm{a}_n(0, x, \varphi(0,x)/\varepsilon).$$

THEOREM 6.9 (*Exact solutions*). *If ε is small enough, the Cauchy problem for (6.7) with initial data h^ε has a unique solution $u^\varepsilon \in \mathcal{S}([0, T_*/\varepsilon] \times \mathbb{R}^d)$ which satisfies for all s, m and ε:*

$$\left\| u^\varepsilon - u^\varepsilon_{\text{app}} \right\|_{H^s} \leq C_{s,m}\varepsilon^m.$$

6.2.2. Dispersive equations; the Schrödinger equation as a generic model for diffractive optics Consider now the case of a dispersive equation (6.7) with $E \neq 0$ in (6.6). Assume that the matrices A_j are real symmetric and that E is real skew symmetric, as usual in physical applications. Therefore the characteristic polynomial $p(\tau, \xi) = \det(\tau A_0 + \sum \xi_j A_j - \mathrm{i}E)$ is real, and thus equal to $\det(\tau A_0 + \sum \xi_j A_j + \mathrm{i}E)$. This implies that $p(-\tau, -\xi) = (-1)^N p(\tau, \xi)$ so that the characteristic variety $\mathcal{C} = \{p = 0\}$ is symmetric, i.e. $-\mathcal{C} = \mathcal{C}$.

We fix a planar phase φ with $\beta = d\varphi = (-\omega, k)$ satisfying Assumption 6.2. We assume that only the harmonics $n\beta$ with $n \in \{-1, 0, 1\}$ are characteristic. We further assume that the nonlinearity F is odd, with a nontrivial cubic term Φ. Again, we look for solutions u^ε of the form (6.34), with profiles $a_n(T, t, x, \theta)$ which are odd in θ, so that the characteristic phase 0 never shows up in the computations. All these assumptions are realistic in applications.

Note that the symmetry assumption of the characteristic variety implies that the eigenvalues $\lambda_\pm$ which describe the characteristic variety $\mathcal{C}$ near $\pm\beta$ (see (6.28)) satisfy the property that

$$\lambda_-(\xi) = -\lambda_+(-\xi). \tag{6.53}$$

In particular, the group velocities $\mathbf{v}_\pm = \nabla\lambda_\pm(\pm k)$ are equal. We denote their common value by $\mathbf{v}$.

The projector $\mathcal{P}$ on $\ker \mathcal{L}(\beta\partial_\theta)$ acting on odd functions reduces to

$$\mathcal{P}\left(\sum_{k\in 1+2\mathbb{Z}} \hat{a}_k \mathrm{e}^{\mathrm{i}k\theta}\right) = P_- \hat{a}_{-1}\mathrm{e}^{-\mathrm{i}\theta} + P_+ \hat{a}_1 \mathrm{e}^{\mathrm{i}\theta}$$

where $P_\pm$ are projectors on $\ker \mathcal{L}(\pm\mathrm{i}\beta)$. Since $\mathcal{L}(-\mathrm{i}\beta) = \overline{\mathcal{L}(\mathrm{i}\beta)}$, one can choose $P_- = \overline{P_+}$, and choose similarly projectors $Q_- = \overline{Q_+}$ on the images of $\mathcal{L}(\pm\mathrm{i}\beta)$ and partial inverses $R_- = \overline{R_+}$.

All that which has been done in the non-dispersive case can be repeated in this framework: one can construct asymptotic solutions

$$u^\varepsilon(t, x) \sim \varepsilon \sum_{n\geq 0} \varepsilon^n \mathrm{a}_n(\varepsilon t, x - t\mathbf{v}, \varphi(t, x)/\varepsilon), \tag{6.54}$$

with profiles $\mathbf{a}_n$ which are odd in θ. They are determined inductively. The Fourier coefficients $\widehat{\mathbf{a}}_{n,k}$ for $k \neq \pm 1$ and $(\mathrm{Id} - P_\pm)\widehat{\mathbf{a}}_{n,\pm 1}$ are given by explicit formulas and $P_\pm \widehat{\mathbf{a}}_{n,\pm 1}$ are given by solving a Schrödinger equation.

In particular, the main term contains only the harmonics ± 1

$$\mathbf{a}_0(T,t,x) = \widehat{\mathbf{a}_0}_{,-1}(T,t,x)\mathrm{e}^{-\mathrm{i}\theta} + \widehat{\mathbf{a}_0}_{,+1}(T,t,x)\mathrm{e}^{\mathrm{i}\theta} \tag{6.55}$$

satisfying the polarization conditions:

$$\widehat{\mathbf{a}_0}_{,\pm 1}(T,t,x) = P_\pm \widehat{\mathbf{a}_0}_{,\pm 1}(T,t,x) \tag{6.56}$$

and the propagation equations

$$\partial_T \widehat{\mathbf{a}_0}_{,\pm 1} \mp \mathrm{i}S(\partial_y)\widehat{\mathbf{a}_0}_{,\pm 1} + B_\pm \widehat{\Phi(\mathbf{a}_0)}_{\pm 1} = 0 \tag{6.57}$$

where $B_\pm$ denotes the partial inverse of $(\mathrm{Id}\ Q_\pm)A_0 P_\pm$ on the image of $P_\pm$. Denoting by $\widetilde{\Phi}(u,v,w)$ the symmetric trilinear mapping such that $\Phi(u) = \widetilde{\Phi}(u,u,u)$, the harmonic of $\Phi(\mathbf{a}_0)$ is

$$\widehat{\Phi(\mathbf{a}_0)}_1 = \widetilde{\Phi}\left(\widehat{\mathbf{a}_0}_{,-1}\widehat{\mathbf{a}_0}_{,1}, \widehat{\mathbf{a}_0}_{,1}\right), \tag{6.58}$$

and there is a similar formula for $\widehat{\Phi(\mathbf{a}_0)}_{-1}$.

This is the usual cubic Shrödinger equation for nonlinear optics that can be found in physics books (see e.g. [6,8,10,107,109]).

The BKW solutions provide approximate solutions of the equations, which can be converted into exact solutions, exactly as in Theorem 6.9. This mathematical analysis was first made in [39]. We refer to this paper for explicit examples coming from various models in nonlinear optics, as discussed in Section 2.

6.2.3. Rectification In the two analyses above, two facts were essential:

- Only regular phases were present. In applications, this excludes the harmonic 0. This was made possible by assuming that both the nonlinearity F and the profiles a were odd. When F is not odd, for instance when F is quadratic, the harmonic 0 is expected to be present. In general, and certainly for non-disoersive equations, 0 is not a regular point of the characteristic variety, implying that there will be no simple analogue of Lemma 6.6. The creation of a nontrivial mean value field by nonlinear interaction of waves is called *rectification*.
- All the characteristic harmonics $k\beta$ are regular and have the same group velocities $\mathbf{v}_k$. If not, the Fourier coefficients $\widehat{a_{n,k}}$ are expected to travel at different speeds $\mathbf{v}_k$, so that the right-hand sides in equations like (6.21) or (6.37) have no definite propagation speed. In this case too, Lemma 6.6 must be revisited.

We now give the general principles which yield the construction of the principal profile a_0, referring to [81,90] for precise statements and proofs.

For a general Eq. (6.7), with F satisfying (6.10), we look for solutions of the form

$$u^\varepsilon(t,x) = \varepsilon^p a^\varepsilon(\varepsilon t, t, x, \varphi/\varepsilon) \tag{6.59}$$

with $a^\varepsilon = a_0 + \varepsilon a_1 + \varepsilon^2 a_2$ and φ a planar phase with $\beta = d\varphi$ satisfying Assumption 6.2. The polarization condition (6.18) and the first Eq. (6.19) for a_0 are such that its Fourier coefficients $\widehat{a_0}_{,k}$ satisfy $P_k \widehat{a_0}_{,k} = \widehat{a_0}_{,k}$ and

$$H_0(\partial_{t,x})\widehat{a_0}_{,0} = 0, \tag{6.60}$$

$$(\partial_t + \mathbf{v}_k \cdot \nabla_x)\widehat{a_0}_{,k} = 0, \quad k \neq 0 \tag{6.61}$$

where $H_0 = (\mathrm{Id} - Q_0)L_1(\partial_{t,x})P_0$. Moreover, the components $(\mathrm{Id} - P_k)\widehat{a_1}_{,k}$ of the Fourier coefficients of a_1 are given by (6.20):

$$\widehat{a_1}_{,k} = -R_k L_1(\partial_{t,x}) P_k \widehat{a_0}_{,k}. \tag{6.62}$$

Next, the equations for $P_k \widehat{a_1}_{,k}$ deduced from (6.21) read

$$H_0(\partial_{t,x}) P_0 \widehat{a_1}_{,0} + \widehat{f}_{0,0} = 0, \tag{6.63}$$

$$(\partial_t + \mathbf{v}_k \cdot \nabla_x) P_k \widehat{a_1}_{,k} + \widehat{f}_{0,k} = 0, \quad k \neq 0, \tag{6.64}$$

where

$$\widehat{f}_{0,k} = (\mathrm{Id} - Q_k)(A_0 \partial_T) - L_1(\partial_{t,x}) R_k L_1(\partial_{t,x}) P_k \widehat{a_0}_{,k} + (\mathrm{Id} - Q_k)\widehat{\Phi(a_0)}_k.$$

We recall that the main problem is *to find conditions on the $\widehat{f}_{0,k}$ which imply that the equations for $P_k \widehat{a_1}_{,k}$ have solutions with sublinear growth in time*. This is what we briefly discuss below.

1. *Large time asymptotics for homogeneous hyperbolic equations*

The operator H_0 is symmetric hyperbolic (on the image of P_0) with constant coefficients. For instance, in the non-dispersive case, $E = 0$, $P = \mathrm{Id}$ and $H_0 = L_1$. This leads one to consider the large time asymptotics of solutions of

$$H(\partial_{t,x})a = f, \qquad a_{|t=0} = h \tag{6.65}$$

when H is a symmetric hyperbolic constant coefficient first order system. The characteristic variety $\mathcal{C}_H$ is a homogeneous algebraic manifold. *It may contain hyperplanes*, say $\mathcal{C}_\alpha = \{(\tau, \xi), \tau + \mathbf{c}_\alpha \cdot \xi = 0\}$. This *does* happen for instance for the Euler equation. The *flat* part $\mathcal{C}_{\mathrm{flat}}$ of $\mathcal{C}_H$ is the union of the $\mathcal{C}_\alpha$. We denote by $\pi_\alpha(\xi)$ the spectral projector on $\ker H(-\mathbf{c}_\alpha \cdot \xi, \xi)$ and by $\pi_\alpha(D_x)$ the corresponding Fourier multiplier. The following lemma is easily obtained by using the Fourier transform of the equation.

LEMMA 6.10. *Suppose that $h \in \mathcal{S}(\mathbb{R}^d)$ and a is the solution of the initial value problem $H(\partial_{t,x})a = 0$, $a_{|t=0} = h$. Then*

$$\lim_{t\to\infty} \left\| a(t,\cdot) - \sum (\pi_\alpha(D_x)h)(\cdot - t\mathbf{c}_\alpha) \right\|_{L^\infty(\mathbb{R}^d)} = 0.$$

2. *Large time asymptotics for* $\Phi(a_0)$

For harmonics $k \neq 1$, the transport Eq. (6.61) implies that $\widehat{a_0}_{,k}$ is a function of $x - t\mathbf{v}_k$. In the dispersive case, there is a finite number of frequencies k in play; in the non-dispersive case, all the speeds $\mathbf{v}_k$ are equal. Thus, in all cases, a finite number of speeds $\mathbf{v}_k$ occur, and the Fourier coefficients of a_0 are the sum of rigid waves moving at the speed c_α or $\mathbf{v}_k$ and of spreading waves which tends to 0 at infinity. The Fourier component $\widehat{\Phi(a_0)}_{,k}$ appears as a sum of products of such waves. Two principles can be observed:

– the product of rigid waves moving at different speeds tends to 0 at infinity;
– the product of a spreading wave and anything bounded, also tends to 0 at infinity.

This implies that for all T, the coefficients $\widehat{f}_{0,k}$ have the form

$$g(t,x) = \sum g_j(x - t\mathbf{w}_j) + g_0(t,x) \tag{6.66}$$

where the set of speeds is $\{\mathbf{w}_j\} = \{\mathbf{v}_k\} \cup \{\mathbf{c}_\alpha\}$, and $g_0(t,\cdot)$ tends to 0 in H^s as $t \to \infty$, uniformly in T.

3. *Large time asymptotics for inhomogeneous hyperbolic equations*

Using the equation on the Fourier side, one proves the following estimate:

LEMMA 6.11. *Consider g as in (6.66). The solution of $H(\partial_{t,x})a = g$, $a_{|t=0} = 0$ satisfies*

$$\lim_{t\to\infty} \frac{1}{t} \left\| a(t,\cdot) - t \sum_\alpha \sum_{\{j:\mathbf{w}_j = \mathbf{c}_\alpha\}} \pi_\alpha(D_x) g_j(\cdot - t\mathbf{c}_\alpha) \right\|_{L^\infty(\mathbb{R}^d)} = 0.$$

In particular, there is a solution with sublinear growth if and only if

$$\pi_\alpha(\partial_x) g_j = 0 \quad \text{when } \mathbf{w}_j = \mathrm{c}_\alpha. \tag{6.67}$$

Applied to $H = \partial_t + \mathbf{v}_k \cdot \nabla_x$, this lemma implies that the equation $(\partial_t + \mathbf{v}_k \cdot \nabla_x)a = g$ has a solution with sublinear growth if and only if

$$g_j = 0 \quad \text{when } \mathbf{w}_j = \mathrm{v}_k. \tag{6.68}$$

These principles can be applied to the equations (6.63) (6.64) to derive (necessary and) sufficient conditions for the existence of solutions with sublinear growth. These conditions give the desired equations for a_0. Their form depends on the relations between the speeds $\mathbf{v}_k$ and c_α. To give an example, we consider a particular, simple case and refer to [81,90] for more general statements.

Consider a non-dispersive system. Denote by $\mathbf{v}$ the common value of the propagation speeds $\mathbf{v}_k$ for $k \neq 0$. Assume that there is one hyperplane in the characteristic variety of L_1, with speed $\mathbf{c}$ and projector π. The principal profile a_0 is split into its mean value $\langle a_0 \rangle = \underline{a}_0$ and its oscillation a_0^*. The oscillation satisfies the polarization and propagation condition

$$a_0^* = P a_0^*, \qquad a_0^*(T,t,x,\theta) = \mathrm{a}_0^*(T, x - t\mathbf{v}, \theta). \tag{6.69}$$

Moreover, the mean value satisfies

$$L_1(\partial_{t,x})\underline{a}_0 = 0 \tag{6.70}$$

and

$$\lim_{t\to\infty} \sup_{T\in[0,T_*]} \|\underline{a}_0(T,t,\cdot) - \pi(D_x)\underline{a}_0(T,t,\cdot)\|_{L^\infty} = 0, \tag{6.71}$$

$$\pi(D_x)\underline{a}_0(T,t,x) = \pi(D_x)\underline{\mathrm{a}}_0(T,x - t\mathbf{c}). \tag{6.72}$$

The propagation equations link $\pi\underline{a}_0$ and a_0^* as follows:

$$A_0\partial_T\underline{\mathrm{a}}_0 + \pi(\partial_y)\left\langle\Phi(\underline{\mathrm{a}}_0 + \delta \mathrm{a}_0^*)\right\rangle = 0, \tag{6.73}$$

$$(\partial_T) + \partial_\theta^{-1}S(\partial_y)\mathrm{a}_0^* + R_0\left(\Phi(\delta\underline{\mathrm{a}}_0 + \mathrm{a}_0^*)\right)^* = 0, \tag{6.74}$$

with $\delta = 1$ when $\mathbf{c} = \mathbf{v}$ and $\delta = 0$ otherwise. There is no "Schrödinger" term in the equation for $\underline{\mathrm{a}}$ since the eigenvalue $\xi \cdot \mathbf{c}$ is flat.

The property (6.71) implies that $\underline{a}_0 \sim \pi(D_x)\underline{a}_0$ for large times $t = T/\varepsilon$.

This brief discussion explains how one can find the first profile a_0 and correctors a_1 and a_2 in order to cancel out the three first terms (6.13), (6.14), (6.15). We refer to [81,90] for more precise statements and proofs. We note that the sublinear growth of a_1 does not allow one to continue the expansion and get approximations at all order. In [81] examples are given showing that this is sharp.

6.3. *Other diffractive regimes*

We briefly mention several other problems, among many, which have been studied in the diffractive regime.

The case of *slowly varing phases* is studied in [44,45], after the formal analysis given in [67]. This concerns expansions of the form

$$u^\varepsilon(t,x) \sim \varepsilon^p a(t,x,\psi(t,x)/\varepsilon,\varphi(t,x)/\varepsilon^2)$$

or equivalent formulations that can be deduced by rescaling.

In the next section, we study *multi-phase* expansions in the regime of geometric optics. This can also be done in the diffractive regime.

The case of *transparent* nonlinearities can be also studied, in the spirit of what is done in the regime of geometric optics. See [73,30] for Maxwell–Bloch equations.

The Schrödinger equation is "generic" in diffractive optics. However, in many applications, there are different relevant multi-scale expansions which yield different dispersive equations. Among them, KdV in dimension $d = 1$, Davey–Stewartson equations (see e.g. [28,31]), K-P equations, (see e.g. [91]), Zakharov system (see e.g. [29]).

7. Wave interaction and multi-phase expansions

The aim of this section is to discuss a framework which describes wave interaction. We restrict ourselves here to the regime of geometric optics. The starting point is to consider multi-phase asymptotic expansion

$$u^\varepsilon(t,x) \sim \varepsilon^p \mathbf{u}^\varepsilon(t,x,\Phi(t,x)/\varepsilon) \tag{7.1}$$

where $\Phi = (\varphi_1, \dots, \varphi_m)$ and possibly $\mathbf{u}^\varepsilon \sim \sum \varepsilon^n \mathbf{u}_n$. Points to discuss are: what are the periodicity conditions in θ for the profiles $\mathbf{u}$; in which function spaces can we look for the profiles; in which framework can we prove convergence? We start with examples and remarks showing that the phase generation by nonlinear interaction can yield very complicated situations, and indicating that restrictive assumptions on the set of phases are necessary for the persistence and stability of representations like (7.1). After a short digression concerning the general question of representation of oscillations, the main outcomes of this section are:

– under an assumption of weak coherence in the phases, there are natural equations for the main profile which can be solved, yielding approximate solutions of the equation, that is, modulo errors which tends to zero as ε tends to 0 without any rate of convergence.
– the stability of such approximate solutions requires stronger assumptions (in general) called strong coherence.
– the strong coherence framework applies to physical situations and thus provides a rigorous justification of several asymptotic representations of solutions.

The wave interaction problem is also of crucial importance in the regime of diffractive optics. Another difficulty appears here: because the intensities are weaker, to be visible the interaction requires longer intervals of time. We will not tackle this problem in these notes but refer, for instance, to [44,45,81,90,7,30,31,98,124,125] among many other.

7.1. *Examples of phase generation*

7.1.1. Generation of harmonics The first occurrence of nonlinear interaction of oscillations is the creation of harmonics. Consider the elementary example

$$\partial_t u^\varepsilon + c\partial_x u^\varepsilon = (u^\varepsilon)^2, \quad u^\varepsilon_{|t=0} = a(x)\cos(x/\varepsilon).$$

For $0 \le t < \|a\|_{L^\infty}$, the solution is $u^\varepsilon(t,y) = \mathbf{u}(t,x,(x-ct)/\varepsilon)$ with

$$\mathbf{u}(t,x,\theta) = \frac{a(x)\cos\theta}{1 - ta(x)\cos\theta} = \sum_{n=1}^{\infty} t^{n-1} a(x)^n (\cos(\theta))^n.$$

Thus the solution is a periodic function of the phase $\theta = (x-ct)/\varepsilon$, where all the harmonics $n\theta$ are present, while the initial data has only the harmonics 1 and -1.

7.1.2. Resonance and phase matching; weak vs. strong coherence Given wave packets with phases φ_1 and φ_2, nonlinearities introduce terms with phases $\nu_1\varphi_1 + \nu_2\varphi_2$. These oscillations can be propagated: this is *resonance* or *phase matching*. To illustrate these phenomena, consider the system

$$\begin{cases} \partial_t u_1 + c_1\partial_x u_1 = 0, \\ \partial_t u_2 + c_2\partial_x u_2 = 0, \\ \partial_t u_3 = u_1 u_2, \end{cases} \tag{7.2}$$

with initial data

$$u_j(0,x) = a_j(x)\mathrm{e}^{\mathrm{i}\psi_j/\varepsilon}, \quad j \in \{1,2,3\}.$$

We assume that $0 \neq c_1 \neq c_2 \neq 0$. Then for $j \in \{1,2\}$,

$$u_j = a_j(x - c_j t)\mathrm{e}^{\mathrm{i}\varphi_j/\varepsilon}, \qquad \varphi_j(t,x) = \psi_j(x - c_j t)$$

and

$$u_3(t,x) = a_3(x)\mathrm{e}^{\mathrm{i}\psi_3(x)} + \int_0^t a_1(x - c_1 s)a_2(x - c_2 s)\mathrm{e}^{\mathrm{i}(\varphi_1(s,x)+\varphi_2(s,x))/\varepsilon}\,\mathrm{d}s. \tag{7.3}$$

The behavior of u_3^ε depends on the existence of critical points for the phase $\varphi_1 + \varphi_2$.

• *Resonance or phase matching* occurs when $\varphi_1 + \varphi_2$ is characteristic for the third field ∂_t, that is when

$$\partial_t(\varphi_1 + \varphi_2) = 0. \tag{7.4}$$

In this case,

$$u_3(t,x) = a_3(x)\mathrm{e}^{\mathrm{i}\psi_3(x)} + \mathrm{e}^{\mathrm{i}(\psi_1(x)+\psi_2(x))/\varepsilon}\int_0^t a_1(x - c_1 s)a_2(x - c_2 s)\,\mathrm{d}s,$$

showing that the oscillations of u_1 and u_2 interact constructively and bring a contribution of order 1 to u_3.

• *Weak coherence.* If

$$\partial_t(\varphi_1 + \varphi_2) \neq 0 \quad \text{almost everywhere} \tag{7.5}$$

then the integral in (7.3) tends to 0 as ε tends to 0. It is a *corrector* with respect to the principal term $a_3\mathrm{e}^{\mathrm{i}\psi_3/\varepsilon}$. The oscillations of u_1 and u_2 do not interact constructively to modify the principal term of u_3. Note that there is no general estimate of the size of the corrector:

it is $O(\sqrt{\varepsilon})$ if the phase has only non-degenerate critical points, it is $O(\varepsilon^{1/p})$ when there are degenerate critical points of finite order.

• *Strong coherence.* Suppose that

$$\partial_t(\varphi_1 + \varphi_2) \neq 0 \quad \text{everywhere.} \tag{7.6}$$

In this case, and in this case only, one can perform direct integration by parts of the integral and the corrector is $O(\varepsilon)$. Indeed, if the coefficients a are smooth, repeated integration by parts yields a complete asymptotic expansion in powers of ε.

CONCLUSION 7.1. *Complete asymptotic expansion in powers of ε can exist only if all the phases in play are either resonant or strongly coherent. Otherwise, the convergence to zero of correctors can be arbitrarily slow.*

REMARK 7.2. Resonance is a rare phenomenon. In the example, the phases φ_1 and φ_2 satisfy

$$(\partial_t + c_1\partial_x)\varphi_1 = 0, \qquad (\partial_t + c_2\partial_x)\varphi_2 = 0, \qquad \partial_t(\varphi_1 + \varphi_2) = 0.$$

Thus

$$0 = (\partial_t + c_2\partial_x)\partial_t(\varphi_1 + \varphi_2) = \partial_t(\partial_t + c_2\partial_x)\varphi_1 = c_1(c_1 - c_2)\psi_1''(y - c_1 t),$$

implying that $\psi_1'' = 0$. Thus ψ_1 must be a plane phase and $\varphi_1 = \alpha_1(x - c_1 t) + \beta_1$ and, similarly, $\varphi_2 = \alpha_2(x - c_2 t) + \beta_2$. The resonance occurs if and only if $\alpha_1 c_1 + \alpha_2 c_2 = 0$, showing that not only must the phases be planar but also the coefficients must be suitably chosen.

7.1.3. Strong coherence and small divisors Consider initial data for (7.2) which are periodic in y/ε. Thus, for $j = 1, 2$:

$$u_j(0, x) = \sum_n a_{j,n}(x)\mathrm{e}^{\mathrm{i}nx/\varepsilon}, \qquad u_j(t, x) = \left(\sum_n a_{j,n}(x - c_j t)\right)\mathrm{e}^{\mathrm{i}n\varphi_j/\varepsilon}$$

with $\varphi_j = x - c_j t$. Suppose that $u_3(0, x) = 0$. Then u_3 is the sum of integrals

$$I_{m,n} = \left(\int_0^t a_{1,m}(x - c_1 s)a_{2,n}(x - c_2 s)\mathrm{e}^{-\mathrm{i}(mc_1 + nc_2)s/\varepsilon}\,\mathrm{d}s\right)\mathrm{e}^{\mathrm{i}(m+n)x/\varepsilon}.$$

Resonances occur when $(m, n) \in \mathbb{Z}^2$ are such that

$$mc_1 + nc_2 = 0. \tag{7.7}$$

This shows that the presence of resonances also depends on *arithmetical conditions*. For instance, if $c_1/c_2 \notin \mathbb{Q}$, there will be no resonant interaction.

Since $\partial_t(m\varphi_1 + n\varphi_2)$ is a constant, the phases $m\varphi_1 + n\varphi_2$ are either resonant or strongly coherent, yielding integrals which are either $O(1)$ or $O(\varepsilon)$. More precisely, in the nonresonant case

$$I_{m,n} = \varepsilon \frac{O(1)}{mc_1 + nc_2}.$$

However, the convergence of the series $\sum I_{m,n}$ depends on the smallness of the denominators $mc_1 + nc_2$.

CONCLUSION 7.3. *Interaction phenomena also depend on the arithmetic properties of the spectrum of the oscillations. Moreover, the behavior of correctors may also involve problems of small divisors.*

7.1.4. Periodic initial data may lead to almost periodic solutions Consider the following dispersive system

$$\begin{cases} \partial_t u_1 + \partial_x u_1 + \dfrac{1}{\varepsilon} u_2 = 0, \\ \partial_t u_2 - \partial_x u_2 - \dfrac{1}{\varepsilon} u_1 = 0. \end{cases} \tag{7.8}$$

The phenomenon which we describe below also occurs for non-dispersive system, but in dimension $d \geq 2$. Consider initial data with a planar phase

$$u_j(0,x) = \sum a_{j,n} \mathrm{e}^{\mathrm{i}nx/\varepsilon}. \tag{7.9}$$

The solution reads

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \sum_n \alpha_n \begin{pmatrix} \mathrm{i} \\ n + \omega_n \end{pmatrix} \mathrm{e}^{\mathrm{i}(ny+\omega_n t)/\varepsilon} + \beta_n \begin{pmatrix} n + \omega_n \\ \mathrm{i} \end{pmatrix} \mathrm{e}^{\mathrm{i}(ny-\omega_n t)/\varepsilon}, \tag{7.10}$$

with $\omega_n = \sqrt{n^2+1}$. This solution is periodic in y/ε. However, the set $\{\omega_n\}_{n\in\mathbb{Z}}$ of time frequencies is not contained in a finitely generated $\mathbb{Z}$-module. The solution is not periodic in time, it is only almost periodic.

CONCLUSION 7.4. *Even for periodic initial data, the solution cannot always be written in the form* (7.1) $\mathbf{u}(x, \Phi(x)/\varepsilon)$ *with a finite number of phases* $\Phi_1, \ldots, \Phi_m$ *and periodic profiles* $\mathbf{u}(x,\theta)$. *The introduction of almost periodic functions can be compulsory.*

7.1.5. A single initial phase can generate a space of phases of infinite dimension over $\mathbb{R}$ Consider (7.8) with initial data

$$u_j(0,x) = \sum a_{j,n} \mathrm{e}^{\mathrm{i}nx^2/\varepsilon}. \tag{7.11}$$

We have simply replaced the planar phases nx by the curved phases $\psi_n(x) = nx^2/2$. The outgoing phases are the solutions of the eikonal equation

$$\partial_t \varphi_n^{\pm} = \pm\sqrt{1 + (\partial_y \varphi_n^{\pm})^2}, \qquad \varphi_n^{\pm}(0, y) = \psi_n(y).$$

All the phases $\varphi_n^{\pm}$ are defined and smooth on $\Omega = \{(t, x); x > |t|, |t| < 1\}$. One can check that the phases φ_n are independent, not only over $\mathbb{Q}$, but also over $\mathbb{R}$.

CONCLUSION 7.5. *A one dimensional vector space of initial phases can generate a space of phases of infinite dimension over* $\mathbb{R}$. *In this case, any description* (7.1) *must include an infinite number of phases, whatever the properties imposed on the spectrum of the profiles.*

7.1.6. Multidimensional effects We have already discussed the effects of focusing in Section 5.4: the linear amplification can be augmented by a nonlinear term and produce blow up. We briefly discuss below two other examples taken from [76].

A more dramatic multidimensional effect is *instantaneous blow up*, which may occur when the initial phase φ has a stationary point. This is illustrated in the following example. Consider in dimension $d = 3$

$$\begin{cases} (\partial_t^2 - \Delta_x)u^\varepsilon = 0, \; u^\varepsilon_{|t=0} = 0, & \partial_t u^\varepsilon_{|t=0} = h(|x|^2)\mathrm{e}^{\mathrm{i}|x|^2/\varepsilon}, \\ \partial_t v^\varepsilon = |\partial_t u^\varepsilon|^2 |v^\varepsilon|^2, & v^\varepsilon_{|t=0} = v_0(x). \end{cases} \tag{7.12}$$

Then

$$u^\varepsilon(t, 0) = th(t)\mathrm{e}^{\mathrm{i}t^2/\varepsilon} \qquad \partial_t u^\varepsilon = 2i\varepsilon^{-1}t^2 h(t)\mathrm{e}^{\mathrm{i}t^2/\varepsilon} + O(1).$$

Therefore, if $h(0) = 1$, $|\partial_t u^\varepsilon| \geq \varepsilon^{-1}t^2$ in an interval $[0, T_0]$ and if $v_0(0) = 1$,

$$v(t, 0) \geq \frac{1}{1 - t^5/5\varepsilon^2}.$$

This proves that the maximal time of existence of bounded solutions is $T(\varepsilon) \lesssim \varepsilon^{\frac{2}{5}}$, and there is no uniform domain of existence of solutions.

When combined with phase interaction, the focusing effects can come from linear combinations of the phases, before the initial phase focus. More precisely, the initial data can launch regular phases φ_j; nonlinear interaction creates new phases $\psi = \sum n_j \varphi_j$, and one among them may have a critical point at $t = 0$, producing instantaneous explosion. This phenomenon is called *hidden focusing* in [76] and is illustrated by the following example

$$\begin{cases} (\partial_t^2 - \Delta_x)u^\varepsilon = 0, & u^\varepsilon_{|t=0} = 0, & \partial_t u^\varepsilon_{|t=0} = \mathrm{e}^{\mathrm{i}x_1/\varepsilon}, \\ (\partial_t^2 - c\Delta_x)v^\varepsilon + a\partial_1 v^\varepsilon = 0, & v^\varepsilon_{|t=0} = \varepsilon \mathrm{e}^{\mathrm{i}x_1/\varepsilon}, & \partial_t u^\varepsilon_{|t=0} = b(0, x)\mathrm{e}^{\mathrm{i}x_1/\varepsilon}, \\ (\partial_t^2 - 3\Delta_x)w^\varepsilon = f^\varepsilon, & u^\varepsilon_{|t=0} = 0, & \partial_t u^\varepsilon_{|t=0} = \mathrm{e}^{\mathrm{i}x_1/\varepsilon}, \\ \partial_t z^\varepsilon = |\partial_{x_1} w^\varepsilon|^2 |z^\varepsilon|^2, & v^\varepsilon_{|t=0} = 0, & \end{cases}$$

with

$$f^\varepsilon = \frac{1}{(b(t,x))^2}\left(\partial_t \overline{u^\varepsilon} + \partial_{x_1}\overline{u^\varepsilon}\right)^3 \left(\partial_t u^\varepsilon - \partial_{x_1} u^\varepsilon\right)(\partial_t v^\varepsilon)^2.$$

The phases in play are $\varphi_1 = x - t$, $\varphi_2 = x + t$ and the solutions φ_3 and φ_4 of the eikonal equation for $\partial_t^2 - c\Delta_x$ with initial value x_1. If $c(0,0) = 2$, $\varphi_3 = x + 2t + 0(t^2 + |tx|)$ and $\varphi_4 = x - 2t + 0(t^2 + |tx|)$. In particular, the phase $\phi = -2\varphi_2 + \varphi_1 + 2\varphi_3 = O(t^2 + |tx|)$, which is revealed by the interaction f^ε, has a critical point at the origin, producing blow up in the fourth equation as in the previous example:

PROPOSITION 7.6 (*[76]*). *One can choose the functions $a(t,x)$, $b(t,x)$ and $c(t,x)$ such that $c(0,0) = 2$ and the maximal time of existence of bounded solutions is $T(\varepsilon) \lesssim \varepsilon^{\frac{1}{6}}$.*

7.2. *Description of oscillations*

In this section we deviate slightly from the central objective of nonlinear optics, and evoke the more general question of how one can model oscillations.

7.2.1. Group of phases, group of frequencies In the representation (7.1), Φ is thought of as a vector-valued function of (t,x), for instance $\Phi = (\varphi_1, \ldots, \varphi_m)$, and the phases are linear combinations $\varphi = \sum \alpha_j \varphi_j = \alpha \cdot \Phi$; the important property of the set of α is that it must be an additive group, to account for wave interaction. As explained in the examples above, this group is not necessarily finitely generated, for instance in the case of almost periodic oscillations. This leads to the following framework:

ASSUMPTION 7.7. Λ is an Abelian group (called the group of frequencies) and we have an injective homomorphism from Λ to $C^\infty(\overline{\Omega}; \mathbb{R})$ where Ω is an open set in $\mathbb{R}^{1+d}$. We denote it by $\alpha \mapsto \alpha \cdot \Phi$.

The set $\mathcal{F} = \Lambda \cdot \Phi$ is an additive subgroup of $C^\infty(\overline{\Omega}; \mathbb{R})$ (the group of phases) assumed to contain no nonzero constant function.

We also assume that for all $\varphi \in \mathcal{F} \setminus \{0\}$, $d\varphi \neq 0$ a.e. in $\overline{\Omega}$.

We denote by $\alpha \cdot \Phi(t,x)$ the value at (t,x) of the phase $\alpha \cdot \Phi$, and by $\alpha \cdot d\Phi(t,x)$ its differential. The notation $\alpha \mapsto \alpha \cdot \Phi$ for the homomorphism from $\Lambda \mapsto C^\infty(\overline{\Omega}; \mathbb{R})$ is chosen to mimic the case where $\Lambda \subset \mathbb{R}^m$ and $\Phi \in C^\infty(\overline{\Omega}; \mathbb{R}^m)$.

EXAMPLES 7.8.
(a) When $\mathcal{F}$ is finitely generated, one can choose a basis over $\mathbb{Z}$ and we are reduced to the case where $\Lambda = \mathbb{Z}^p$.
(b) Another interesting case occurs when the $\mathbb{R}$-vector space generated by $\mathcal{F}$, called $\widetilde{\mathcal{F}}$, is of finite dimension m over $\mathbb{R}$. Then one can choose a finite basis (ψ^ℓ) of $\widetilde{\mathcal{F}}$ and denote by $\cdot\Psi$ the mapping $\beta \cdot \Psi = \sum_{1\leq \ell \leq m} \beta_\ell \psi^k$ from $\mathbb{R}^m$ to $\widetilde{\mathcal{F}}$. Then, we have a group homomorphism $M : \Lambda \mapsto \mathbb{R}^m$ such that

$$\alpha \cdot \Phi = (M\alpha) \cdot \Psi. \tag{7.13}$$

7.2.2. The fast variables The question is to know what is the correct space Θ for the fast variables θ in the profiles $\mathbf{u}(t,x,\theta)$. In the periodic case, Θ is a torus $\mathbb{T}^m$, the set of frequencies is $\mathbb{Z}^m$, and the link is clear in Fourier expansions: the substitution $\theta = \Phi/\varepsilon$,

$$e^{i\alpha\theta} \mapsto e^{i\alpha\cdot\Phi/\varepsilon} \tag{7.14}$$

transforms the exponential into the desired oscillation. This approach extends to general Abelian group, using the corresponding Fourier analysis. We refer for instance to [126,63] for a detailed presentation of Fourier analysis on groups.

The group Λ is equipped with the discrete topology. Its *characters* are the homomorphisms from Λ to the unit circle $S^1 \subset \mathbb{C}$, and the set of characters is another Abelian group, denoted by Θ, which is compact when equipped with point-wise convergence. The duality theorem of Pontryagin–Van Kampen asserts that the group of characters of Θ is Λ. We denote this duality by

$$(\alpha,\theta) \in \Lambda \times \Theta \quad \mapsto \quad e_\alpha(\theta) \in S^1. \tag{7.15}$$

This extends to the general setting, the functions $e^{i\alpha\theta}$ for the duality $\alpha \in \mathbb{Z}^m$, $\theta \in \mathbb{T}^m$.

One advantage of working on a compact group Θ is that there is a nice L^2 Fourier theory. Equipped with the Haar probability $d\theta$, the $\{e_\alpha, \alpha \in \Lambda\}$ form an orthonormal basis of $L^2(\Theta)$ and any $\mathbf{u} \in L^2(\Theta)$ has the Fourier series decomposition

$$\mathbf{u}(\theta) = \sum_\Lambda \hat{\mathbf{u}}(\alpha) e_\alpha(\theta), \qquad \hat{\mathbf{u}}(\alpha) = \int_\Theta \mathbf{u}(\theta)\overline{e_\alpha(\theta)}\, d\theta.$$

The *spectrum* of $\mathbf{u}$ is the support of $\hat{\mathbf{u}}$ in Λ.

7.2.3. Profiles and oscillations Profiles are functions $\mathbf{u}(t,x,\theta)$ on $\Omega \times \Theta$. The simplest examples are trigonometric polynomials, that is finite linear combinations of the characters $e_\alpha(\theta)$ with coefficients $\hat{\mathbf{u}}_\alpha \in C^\infty(\overline{\Omega})$.

We now present the intrinsic definition of the substitution Φ/ε in place of θ in profiles. For all $(t,x) \in \overline{\Omega}$ and $\varepsilon > 0$ the mapping $\alpha \mapsto e^{i\alpha\cdot\Phi(t,x)/\varepsilon}$ is a character on Λ, and therefore defines $p^\varepsilon(t,x) \in \Theta$ such that

$$e^{i\alpha\cdot\Phi(t,x)/\varepsilon} = e_\alpha(p^\varepsilon(t,x)). \tag{7.16}$$

For $\mathbf{u} \in C^0(\overline{\Omega} \times \Theta)$ one can define the family of oscillating functions:

$$\mathbf{p}^\varepsilon(\mathbf{u})(t,x) = \mathbf{u}(t,x,p^\varepsilon(t,x)). \tag{7.17}$$

This coincides with the expected definition for trigonometric polynomials $\mathbf{u} = \sum \hat{\mathbf{u}}_\alpha e_\alpha$, since then

$$\mathbf{p}^\varepsilon(\mathbf{u})(t,x) = \sum \hat{\mathbf{u}}_\alpha(t,x)\, e^{i\,\alpha\cdot\Phi(t,x)/\varepsilon}.$$

7.2.4. Profiles in L^p and associated oscillating families The definition of the substitution can be extended to profiles $\mathbf{u} \in L^p$, $p < +\infty$, not for a fixed ε, but in an asymptotic way. The result is a class of bounded families in L^p, modulo families which converge strongly to 0. The key observation is the following:

PROPOSITION 7.9. *For* $\mathbf{u} \in C_0^0(\overline{\Omega} \times \Theta)$ *and* $1 \leq p < +\infty$*, there holds*

$$\|\mathbf{u}\|_{L^p(\Omega\times\Theta)} = \lim_{\varepsilon\to 0} \|\mathbf{p}^\varepsilon(\mathbf{u})\|_{L^p(\Omega)}.$$

PROOF. $\mathbf{v} = |\mathbf{u}\,|^p \in C_0^0(\overline{\Omega} \times \Theta)$ and it is sufficient to prove that

$$\int_{\Omega\times\Theta} \mathbf{v}(t, x, \theta)\, \mathrm{d}t\mathrm{d}x\mathrm{d}\theta = \lim_{\varepsilon\to 0} \int_\Omega \mathbf{p}^\varepsilon(\mathbf{v})(t, x)\, \mathrm{d}t\mathrm{d}x.$$

By density one can assume that $\mathbf{v}$ is a trigonometric polynomial. In this case

$$\int_\Omega \mathbf{p}^\varepsilon(\mathbf{u})(t, x)\, \mathrm{d}t\mathrm{d}x = \int_\Omega \sum \hat{\mathbf{u}}(t, x, \alpha)\, \mathrm{e}^{\mathrm{i}\,\alpha\cdot\Phi(t,x)/\varepsilon}\, \mathrm{d}t\mathrm{d}x$$
$$\underset{\varepsilon\to 0}{\longrightarrow} \int_\Omega \hat{\mathbf{u}}(t, x, 0)\, \mathrm{d}t\mathrm{d}x = \int_{\Omega\times\Theta} \mathbf{u}(t, x, \theta)\, \mathrm{d}t\mathrm{d}x\mathrm{d}\theta$$

where the convergence follows from Lebesgue's lemma and the assumption that $d(\alpha\cdot\Phi) \neq 0$, $a.e.$ if $\alpha \neq 0$. □

DEFINITION 7.10. Let $\mathcal{L}_\varepsilon^p(\Omega)$ denote the space of families (u^ε) which are bounded in $L^p(\Omega)$. Two families (u^ε) and (v^ε) are equivalent if $(u^\varepsilon - v^\varepsilon)$ tends to 0 in $L^p(\Omega)$. This is denoted by $(u^\varepsilon) \sim (v^\varepsilon)$.

Consider the following semi-norm on $\mathcal{L}_\varepsilon^p(\Omega)$:

$$\|(u^\varepsilon)\|_p = \limsup_{\varepsilon\to 0} \|u^\varepsilon\|_{L^p(\Omega)}. \tag{7.18}$$

The set $\mathcal{N}_\varepsilon^p$ of families (u^ε) such that $\|(u^\varepsilon)\|_p = 0$ is the subspace of families $(u^\varepsilon) \sim 0$ which converge strongly to zero in $L^p(\Omega)$. The quotient space $\mathcal{L}_\varepsilon^p/\mathcal{N}_\varepsilon^p$ is a space of bounded families in L^p modulo families which converge strongly to 0. With the norm (7.18) it is a Banach space.

For $\mathbf{u} \in C_0^0(\overline{\Omega} \times \Theta)$ the family $(u^\varepsilon) = (\mathbf{p}^\varepsilon(\mathbf{u}))$ is bounded in $L^p(\Omega)$. Seen in the quotient space L_ε^p, Proposition 7.9 implies that $\mathbf{p}^\varepsilon$ defined on $C_0^0(\overline{\Omega} \times \Theta)$ is an isometry for the norms of $L^p(\Omega \times \Theta)$ and $L_\varepsilon^p(\Omega)$. Thus by density, $\mathbf{p}^\varepsilon$ extends as an isometry $\widetilde{\mathbf{p}}^\varepsilon$ from $L^p(\Omega \times \Theta)$ into L_ε^p.

DEFINITION 7.11 (*$\mathcal{F}$-oscillating families in L^p*). $L_{\varepsilon,\mathcal{F}}^p$ denotes the image of $L^p(\Omega \times \Theta)$ by $\widetilde{\mathbf{p}}^\varepsilon$. A family (u^ε) in $\mathcal{L}_\varepsilon^p$ is called $\mathcal{F}$-oscillating, if its image in the quotient L_ε^p belongs to $L_{\varepsilon,\mathcal{F}}^p$. The space of such families is denoted bu $\mathcal{L}_{\varepsilon,\mathcal{F}}^p \subset \mathcal{L}_\varepsilon^p$.

Because $\widetilde{\mathbf{p}}^\varepsilon$ is an isometry, for $(u^\varepsilon) \in \mathcal{L}^p_{\varepsilon,\mathcal{F}}$, there is a *unique* $\mathbf{u} \in L^p(\Omega \times \Theta)$ such that $(u^\varepsilon) \in \tilde{\mathbf{p}}^\varepsilon(\mathbf{u})$. This $\mathbf{u}$ is called *the profile of the oscillating family* (u^ε) and we can think of it as

$$u^\varepsilon(t,x) \sim \mathbf{u}(t,x,p^\varepsilon(t,x))$$

where the $\sim$ is taken in the sense that the difference converges to 0 in L^p, the substitution is exact if $\mathbf{u}$ is continuous, and taken in the extended sense if $\mathbf{u} \in L^p$.

7.2.5. Multi-scale weak convergence, weak profiles

PROPOSITION 7.12. *If Λ is countable, then for all $(u^\varepsilon) \in \mathcal{L}^p_\varepsilon(\Omega)$, $1 < p < +\infty$, there is a sequence $\varepsilon_n \to 0$ and also $\mathbf{u} \in L^p(\Omega\times\Theta)$, such that for all trigonometric polynomials $\mathbf{a}$:*

$$\int_\Omega u^{\varepsilon_n}(t,x)\mathbf{p}^{\varepsilon_n}(\mathbf{a})(t,x)\mathrm{d}t\mathrm{d}x \xrightarrow[\varepsilon_n\to 0]{} \int_{\Omega\times\Theta} \mathbf{u}(t,x,\theta)\,\mathbf{a}(t,x,\theta)\,\mathrm{d}t\mathrm{d}x\,\mathrm{d}\theta. \quad (7.19)$$

This $\mathbf{u}$ is called the *weak profile with respect to the group of phases $\mathcal{F}$* of the family (u^{ε_n}). In other contexts it is also called the *two-scale weak limit*.

If $(u^\varepsilon) \in \mathcal{L}^p_{\varepsilon,\mathcal{F}}$ is $\mathcal{F}$ oscillating with profile $\mathbf{u}$, then $\mathbf{u}$ is also the weak profile of the family and no extraction of subsequence is necessary.

We do not investigate these notions further. They have been used in the context of geometric optics in [79,80], see also Section 7.6.4 below. They are also useful in other situations such as homogeneisation.

7.3. *Formal multi-phases expansions*

The goal of this section is to show that the formal analysis of Section 5.2 can be carried out in the more general setting of multi-phase expansions. We consider the weakly nonlinear framework, that is equations of the form

$$A_0(a,\varepsilon u)\partial_t u + \sum_{j=1}^d A_j(a,\varepsilon u)\partial_{x_j} u + \frac{1}{\varepsilon}E(a)u = F(u) \quad (7.20)$$

where the coefficient $a = a(t,x)$ is smooth and given. The matrices A_j and E are assumed to be smooth functions of their arguments. Moreover, the A_j are self-adjoint with A_0 positive definite and E is skew-adjoint. The source term F is a smooth function of its arguments.

7.3.1. Phases and profiles. Weak coherence Following the general principles, we consider a group of phases $\mathcal{F}$ and also introduce the vector space $\widetilde{\mathcal{F}}$ they span. The characteristic phases play a prominent role. We use the notations

$$\begin{cases} A_j(a) = A_j(a, 0), \\ L(a, \partial) = A_0(a)\partial_t + \sum A_j(a)\partial_{x_j} + E(a) = L_1(a, \partial) + E(a). \end{cases} \tag{7.21}$$

ASSUMPTION 7.13 (*Weak coherence*). We are given an additive subgroup $\mathcal{F}$ of $C^\infty(\overline{\Omega}; \mathbb{R})$ such that the vector space $\widetilde{\mathcal{F}}$ generated by $\mathcal{F}$ is finite dimensional over $\mathbb{R}$. We assume that

(i) for all $\varphi \in \widetilde{\mathcal{F}}$, either $\det L(a(t, x), id\varphi(t, x))$ vanishes everywhere on $\overline{\Omega}$ and the dimension $\ker L(a(t, x), id\varphi(t, x))$ is independent of (t, x), or the determinant $\det L(a(t, x), id\varphi(t, x))$ is different from zero almost everywhere.

(ii) for all $\varphi \in \widetilde{\mathcal{F}} \setminus \{0\}$, $d\varphi(t, x) \neq 0$ for almost all $(t, x) \in \overline{\Omega}$.

This condition is called *weak coherence*. The condition (i) means that either φ is a characteristic phase of constant multiplicity in the sense of Definitions 5.3, or it is characteristic almost nowhere. For the model (7.2), and $\mathcal{F}$ generated by the phases φ_1 and φ_2, weak coherence occurs when $\alpha\partial_t\varphi_1 + \alpha_2\partial_t\varphi_2 \neq 0$ almost everywhere when $(\alpha_1, \alpha_2) \neq (0, 0)$. We also refer to Section 7.6.1 for generic examples in dimension $d = 1$.

In dimension $d > 1$, an example (which is indeed strongly coherent as explained below) is the case where $a = \underline{a}$ is constant and $\mathcal{F}$ is a group of planar phases $\{\beta_0 t + \sum \beta_j x_j, \beta \in \Lambda \subset \mathbb{R}^{1+d}\}$.

Taking a basis $(\psi_1, \dots, \psi_m)$ for $\widetilde{\mathcal{F}}$, then any phase $\varphi \in \widetilde{\mathcal{F}}$ is represented as $\varphi = \beta \cdot \Psi = \sum \beta_j \psi_j$ with $\beta \in \mathbb{R}^m$. The subgroup $\mathcal{F}$ corresponds to frequencies β in a subgroup $\Lambda \subset \mathbb{R}^m$. There are three possibilities:

• The group $\mathcal{F}$ (or Λ) is finitely generated of rank $p = m$. Then one can choose a basis of $\mathcal{F}$ over $\mathbb{Z}$ which is also a basis of $\widetilde{\mathcal{F}}$ over $\mathbb{R}$. Then the oscillations are naturally represented as

$$u^\varepsilon(t, x) = \mathbf{u}(t, x, \Psi(t, x)/\varepsilon), \tag{7.22}$$

with $\mathbf{u}(t, x, \theta)$ 2π periodic in $(\theta_1, \dots, \theta_m)$. The space for the variables θ is $\mathbb{T}^m = (\mathbb{R}/2\pi\mathbb{Z})^m$.

• The group $\mathcal{F}$ is finitely generated of rank $p > m$. Then one can choose a basis $(\varphi_1, \dots, \varphi_p)$ of $\mathcal{F}$ such that $\mathcal{F} = \{\alpha \cdot \Phi = \sum \alpha_j \varphi_j; \alpha \in \mathbb{Z}^p\}$. Writing the φ_j in the base $(\psi_1, \dots, \psi_m)$ yields a $m \times p$ matrix M as in (7.13), such that

$$\Lambda = M\mathbb{Z}^p, \qquad \Phi(t, x) = {}^tM\Psi(t, x). \tag{7.23}$$

The oscillations can be represented in two forms:

$$u^\varepsilon(t, x) = \mathbf{u}(t, x, \Psi(t, x)/\varepsilon), \tag{7.24}$$

$$u^\varepsilon(t, x) = \mathbf{v}(t, x, \Phi(t, x)/\varepsilon), \tag{7.25}$$

with $\mathbf{u}(t,x,Y)$ *quasi-periodic* in Y with its spectrum contained in Λ, or with $\mathbf{v}(t,x,\theta)$ *periodic in* $(\theta_1,\dots,\theta_p)$. In accordance with (7.23), the link between $\mathbf{u}$ and $\mathbf{v}$ is that

$$\mathbf{u}(t,x,Y)=\mathbf{v}(t,x,{}^tMY). \tag{7.26}$$

Identifying periodic functions with functions on the torus, the variables θ can be seen as living in $\Theta=\mathbb{T}^p$.

• The group $\mathcal{F}$ is not finitely generated. Then the representation (7.22) involve profiles $\mathbf{u}(t,x,Y)$ which are *almost periodic in* $Y\in\mathbb{R}^m$. Recall that the space of almost periodic functions of Y, with values in a Banach space B, is the closure in $L^\infty(\mathbb{R}^m;B)$ of the set of finite sums $\sum e^{i\beta Y}b_\beta$, with $\beta\in\mathbb{R}^m$ and $b_\beta\in B$, called *trigonometric polynomials*.

The analogue of (7.25) involves a *Bohr compactification* of $\mathbb{R}^m$ (see e.g. [63,126]). The dual group of the discrete group, Λ, is a compact group Θ (see Section 7.2.2). There is a natural mapping $\pi:\mathbb{R}^m\mapsto\Theta$ defined as follows. A point $Y\in\mathbb{R}^m$ is mapped to the character $\alpha\mapsto e^{i\alpha\cdot Y}$. In the quasi-periodic case, $\pi=\varpi\circ{}^tM$, where ϖ is the natural map from $\mathbb{R}^p$ to $\mathbb{T}^p$. The link between almost periodic profiles with spectra contained in Λ and functions on the group Θ is that

$$\mathbf{u}(t,x,Y)=\mathbf{v}(t,x,\pi Y). \tag{7.27}$$

Similarly, the mapping p^ε in (7.16) is $p^\varepsilon=\pi(\Psi/\varepsilon)$. In this setting, the oscillations can be written

$$u^\varepsilon(t,x)=\mathbf{v}(t,x,\pi(\Phi(t,x)/\varepsilon)), \tag{7.28}$$

with $\mathbf{v}(t,x,\theta)$ defined on $\Omega\times\Theta$.

7.3.2. Phases and profiles for the Cauchy problem We make the elementary but important remark that *giving initial oscillating data for the solution u^ε is not the same as giving initial data for the profile* $\mathbf{u}$, since taking $t=0$ in (7.22) yields:

$$u^\varepsilon(0,x)=\mathbf{u}(0,x,\Psi(0,x)/\varepsilon). \tag{7.29}$$

Given initial oscillations

$$u^\varepsilon(0,x)=h^\varepsilon(x)\sim\mathbf{h}(x,\Psi_0(x)/\varepsilon) \tag{7.30}$$

we have to make the links between Ψ_0 and Ψ and between $\mathbf{u}$ and $\mathbf{h}$ which *do not depend on the same set of fast variables*. The situation is that we have a group $\mathcal{F}$ for the phases in Ω and a group $\mathcal{F}_0$ of initial phases. Clearly, the first relation is that $\mathcal{F}_0=\{\varphi_{|t=0};\varphi\in\mathcal{F}\}$. But this is not sufficient: we also want to show that for all $\varphi_0\in\mathcal{F}_0$ the solutions of the eikonal equation with initial value φ_0 belong to $\mathcal{F}$. This is a very strong assumption indeed. It is satisfied for instance in the framework of planar phases and also in other interesting circumstances. The conditions can be summarized in the following assumption:

ASSUMPTION 7.14 (*Weak coherence for the Cauchy problem*). Let $\mathcal{F} = \Lambda \cdot \Psi \subset \widetilde{\mathcal{F}} \subset C^\infty(\overline{\Omega}; \mathbb{R})$ satisfy Assumption 7.13. We denote by $\mathcal{C}_\Lambda$ the set of frequencies $\alpha \in \Lambda$ such that $\det L(a, id\alpha \cdot \Psi) = 0$. When $\alpha \notin \mathcal{C}_\Lambda$ let $P_\alpha = 0$ and when $\alpha \in \mathcal{C}_\Lambda$ let $P_\alpha(t, x)$ denote the spectral projector of

$$A_0^{-1}(a) \left(\sum_{j=1}^{d} \mathrm{i}\partial_{x_j}(\alpha \cdot \Psi) A_j(a) + E(a) \right)$$

associated with the eigenvalue $-\mathrm{i}\partial_t(\alpha \cdot \Psi)$.

With $\overline{\omega} = \overline{\Omega} \cap \{t = 0\}$ let $\mathcal{F}_0 \subset \widetilde{\mathcal{F}}_0 \subset C^\infty(\overline{\omega})$ be the group and vector space of the restrictions to $t = 0$ of the phases φ in $\mathcal{F}$ and $\widetilde{\mathcal{F}}$, respectively. We assume that for all $\varphi_0 \in \widetilde{\mathcal{F}}_0$, $d\varphi_0 \neq 0$ almost every where on $\overline{\omega}$.

$\widetilde{\mathcal{F}}_0$ is finite dimensional, of dimension $m_0 \leq m$; choosing a basis, defines $\Psi_0 \in C^\infty(\overline{\omega}; \mathbb{R}^{m_0})$ and a group $\Lambda_0 \subset \mathbb{R}^{m_0}$ such that $\mathcal{F}_0 = \Lambda_0 \cdot \Psi_0$.

The mapping $\varphi \mapsto \varphi_{|t=0}$ induces a surjective group homomorphism from Λ to Λ_0, called ρ, such that

$$\alpha \cdot \Psi_{|t=0} = (\rho(\alpha)) \cdot \Psi_0. \tag{7.31}$$

We assume that

$$\forall \alpha_0 \in \Lambda_0, \quad \forall x \in \overline{\omega}: \qquad \sum_{\alpha \in \rho^{-1}(\alpha_0)} P_\alpha(0, x) = \mathrm{Id}\,. \tag{7.32}$$

Note that the sum in (7.32) has at most N terms different from zero, since different α's in $\rho^{-1}(\alpha_0)$ correspond to different eigenvalues of $A_0^{-1}\left(\sum \mathrm{i}\partial_{x_j}(\alpha_0 \cdot \Psi_0) A_j + E\right)$. That the sum is equal to the identity means that all the eigenvalues are present in the sum and thus correspond to a characteristic phase $\alpha \cdot \Psi \in \mathcal{F}$ with initial value $\alpha_0 \cdot \Psi_0$.

LEMMA 7.15. *For a trigonometric polynomial* $\mathbf{h}(x, Y_0) = \sum \widehat{\mathbf{h}}_{\alpha_0}(x) \mathrm{e}^{\mathrm{i}\alpha_0 \cdot Y_0}$, *define*

$$\mathbf{u}(0, x, Y) = \sum \widehat{P}_\alpha(0, x) \widehat{\mathbf{h}}_{\rho(\alpha)}(x) \mathrm{e}^{\mathrm{i}\alpha \cdot Y}. \tag{7.33}$$

Then

$$\mathbf{u}(0, x, \Psi(0, x)/\varepsilon) = \mathbf{h}(x, \Psi_0(x)/\varepsilon). \tag{7.34}$$

We note that (7.33) also defines a trigonometric polynomial, since there are finitely many α such that $P_\alpha \neq 0$ and $\rho(\alpha)$ belongs to the spectrum of $\mathbf{h}$. The identity (7.34) is an immediate consequence of (7.31) and (7.32).

This lemma, when extended to more general classes of profiles, will provide the initial data for $\mathbf{u}$, when the profile $\mathbf{h}$ of the initial data is known.

7.3.3. Formal BKW expansions Given a basis $(\psi_1, \ldots, \psi_m)$ of $\widetilde{\mathcal{F}}$, we look for solutions

$$u^\varepsilon \sim \sum_{n \geq 0} \varepsilon^n \mathbf{u}_n(t, x, \Psi(t, x)/\varepsilon) \tag{7.35}$$

with profiles $\mathbf{u}_n(t,x,Y)$ which are periodic/quasi-periodic/almost periodic in $Y \in \mathbb{R}^m$. They have Fourier expansions

$$\mathbf{u}_n(t,x,Y) = \sum_{\alpha\in\Lambda} \hat{\mathbf{u}}_{n,\alpha}(t,x)\, \mathrm{e}^{\mathrm{i}\alpha Y}$$

and for the moment we leave aside the question of convergence of the Fourier series.

As in Section 5.2, one obtains a sequence of equations

$$\mathcal{L}(a,\partial_Y)\mathbf{u}_0 = 0 \tag{7.36}$$

and for $n \geq 0$

$$\mathcal{L}(a,\partial_Y)\mathbf{u}_{n+1} + \mathcal{L}_1(a,\mathbf{u}_0,\partial_{t,x,Y})\mathbf{u}_n = F_n(\mathbf{u}_0,\ldots,\mathbf{u}_n) = \mathbf{F}_n \tag{7.37}$$

where

$$\mathcal{L}(a,\partial_Y) = \sum_{j=1}^{m} L_1(a,d\psi_j)\partial_{Y_j} + E(a) \tag{7.38}$$

$$\mathcal{L}_1(a,v,\partial_{t,x,Y}) = L_1(a,\partial_{t,x}) + \sum_{j=1}^{m} v\cdot\nabla_v \widetilde{L}_1(a,0,d\psi_j)\partial_{Y_j}. \tag{7.39}$$

Here $\widetilde{L}_1(a,v,\tau,\xi)$ denotes the complete symbol $\tau A_0(a,v) + \sum \xi_j A_j(a,v)$. Moreover, $\mathbf{F}_0 = F(\mathbf{u}_0)$ and for $n > 0$, $\mathbf{F}_n = F'(\mathbf{u}_0)\mathbf{u}_n +$ terms which depend only on $(\mathbf{u}_0,\ldots,\mathbf{u}_{n-1})$.

The equations are analyzed in (formal) Fourier series. In particular,

$$\mathcal{L}(a,\partial_Y)\mathrm{e}^{\mathrm{i}\alpha Y} = \mathrm{e}^{\mathrm{i}\alpha Y} L(a, id(\alpha\cdot\Psi)), \qquad \alpha\in\Lambda.$$

By Assumption 7.26:

(1) either the phase $\alpha\cdot\Psi$ is characteristic of constant multiplicity and we can introduce smooth projectors $P_\alpha(t,x)$ and $Q_\alpha(t,x)$ on the kernel and the image of $L(a(t,x), id(\alpha\cdot\Psi)(t,x))$, respectively, and the partial inverse R_α, such that $R_\alpha(\mathrm{Id}-Q_\alpha)=0$, $P_\alpha R_\alpha = 0$, $R_\alpha\mathcal{L}(a,d(\alpha\cdot\Psi))=0$.

(2) or the phase $\alpha\cdot\Psi$ is almost nowhere characteristic, and we *define*

$$P_\alpha = 0, \qquad Q_\alpha = \mathrm{Id}\,, \qquad R_\alpha = (L(a(t,x), id(\alpha\cdot\Psi)(t,x)))^{-1}\,. \tag{7.40}$$

They define operators $\mathcal{P}$, $\mathcal{Q}$ and $\mathcal{R}$, acting on Fourier series (at least on formal Fourier series). The Eq. (7.36) reads

$$\mathbf{u}_0 = \mathcal{P}\mathbf{u}_0 \Leftrightarrow \forall\alpha,\ \hat{\mathbf{u}}_{0,\alpha}\in\ker L(a, id(\alpha\cdot\Psi)). \tag{7.41}$$

The Eq. (7.37) is projected by $\mathcal{Q}$ and Id $-\mathcal{Q}$; the first part gives

$$(\mathrm{Id} - \mathcal{P})\mathbf{u}_{n+1} = \mathcal{R}(\mathbf{F}_n - \mathcal{L}_1(a, \partial_{t,x,Y})\mathbf{u}_n) \tag{7.42}$$

and the second gives the equation

$$(\mathrm{Id} - \mathcal{Q})(\mathcal{L}_1(a, \mathbf{u}_0, \partial_{t,x,Y})\mathbf{u}_n - \mathbf{F}_n) = 0. \tag{7.43}$$

In particular, the equation for the principal profile is

$$(\mathrm{Id} - \mathcal{Q})\mathcal{L}_1(a, \mathbf{u}_0, \partial_{t,x,Y})\mathcal{P}\mathbf{u}_0 = (\mathrm{Id} - \mathcal{Q})F(\mathbf{u}_0). \tag{7.44}$$

Knowing $\mathbf{u}_0$, (7.42) with $n = 0$ determines $(\mathrm{Id} - \mathcal{P})\mathbf{u}_1$; injecting in (7.43) for $n = 1$ gives an equation for $\mathcal{P}\mathbf{u}_1$, and so on, by induction.

This is the general scheme at the formal level. To make it rigorous, we have to define the operators $\mathcal{P}$, $\mathcal{Q}$ and $\mathcal{R}$ in function spaces and solve the profile equations.

7.3.4. Determination of the main profiles It is important that Eq. (7.44) depends only on the definition of the projectors $\mathcal{P}$ and $\mathcal{Q}$, and not on the partial inverse $\mathcal{R}$. To fix the idea, we assume from now on that the projectors $P_\alpha(t, x)$ are the projectors on $\ker \mathcal{L}(a(t, x), i\alpha)$ which are *orthogonal for the scalar product induced by* $A_0(a(t, x))$, and that $Q_\alpha = A_0(\mathrm{Id} - P)A_0^{-1} = (\mathrm{Id} - P_\alpha^*)$.

By the symmetry assumption for L, the projectors P_α and Q_α are uniformly bounded in α. Moreover, smoothess in x for fixed α follows from the constant multiplicity assumption. However, when studying the convergence properties of the Fourier series defining $\mathcal{P}$ and $\mathcal{Q}$, we need a little bit more:

ASSUMPTION 7.16 (*Uniform coherence*). The projectors P_α and Q_α are bounded, as well as their derivatives, uniformly with respect to $(t, x) \in \overline{\Omega}$ and $\alpha \in \Lambda$.

EXAMPLE 7.17. This condition is trivially satisfied if $a(t, x) = \underline{a}$ is constant, since then the projectors are bounded and independent of (t, x). As in Proposition 4.18, it also holds if the symbol $\mathcal{L}(a(t, x), \eta)$ is symmetric hyperbolic in a direction $\underline{\eta}$ with constant multiplicities in (t, x, η) for $\eta \neq 0$.

Several frameworks can be considered:

(H1) The group $\mathcal{F}$ is finitely generated of rank $p \geq m$.

In this case, instead of the quasi-periodic representation (7.22) one uses the periodic one (7.25). With obvious modifications, the equation reads

$$(\mathrm{Id} - \mathcal{Q})\mathcal{L}_1(a, \mathbf{u}_0, \partial_{t,x,\theta})\mathcal{P}\mathbf{u}_0 = (\mathrm{Id} - \mathcal{Q})F(\mathbf{u}_0). \tag{7.45}$$

The profiles are sought in Sobolev spaces $H^s(\Omega \times \mathbb{T}^p)$. One key argument for the convergence of Fourier series is Plancherel's theorem: for $\mathbf{u}(x, \theta) = \sum \widehat{\mathbf{u}}_\alpha(x)e^{i\alpha\cdot\theta}$ the following expression holds

$$\|\mathbf{u}\|^2_{H^s(\omega\times\mathbb{T}^p)} \approx \sum_\alpha \|\widehat{\mathbf{u}}_\alpha\|^2_{H^s(\omega)} + |\alpha|^{2s}\,\|\widehat{\mathbf{u}}_\alpha\|^2_{L^2(\omega)}\,. \tag{7.46}$$

The profile equation inherits symmetry from the original equation and can be solved by an iterative scheme, in the spirit of the general theory of symmetric hyperbolic systems (see Theorem 3.2).

(H2) The equation is semi-linear and the nonlinearity F is real analytic.

In this case, one can use Wiener's algebra of almost periodic functions

$$\mathbb{A}^s = \left\{ \mathbf{u}(x, Y) = \sum \hat{\mathbf{u}}_\alpha(x)\mathrm{e}^{\mathrm{i}\alpha Y};\ \sum_\alpha \left\|\hat{\mathbf{u}}_\alpha\right\|_{H^s} < \infty \right\}.$$

The profile equation on Fourier series reads

$$(\mathrm{Id} - Q_\alpha)L_1(a, \partial_{t,x})P_\alpha\hat{\mathbf{u}}_{0,\alpha} = (\mathrm{Id} - Q_\alpha)\hat{\mathbf{F}}_{0,\alpha}, \quad \alpha \in \Lambda. \tag{7.47}$$

For each α, $(\mathrm{Id} - Q_\alpha)L_1(a, \partial_{t,x})P_\alpha$ is a hyperbolic system (Lemma 5.7). Assuming $\mathbf{F}_0$ given, or given by an iterative process, the coefficients $\hat{\mathbf{u}}_{0,\alpha}$ can be determined from their initial values. Because the projectors are uniformly bounded the equations (7.47) can be uniformly solved. Next, we note that because F is real analytic, it maps the space $\mathbb{E}^s$ into itself if $s > \frac{d}{2}$.

To simplify the exposition we use the following notations and terminology:

NOTATIONS 7.18. Ω *is a truncated cone* $\{(t, x) : 0 \le t \le T\ \lambda_* t + |x| \le R\}$ *with* $T \le T_0$ *and* λ_* *will be chosen large enough so that* Ω *will be contained in the domain of determinacy of* $\omega = \{x : |x| \le R\}$ *whenever this is necessary (see the discussion in Section 3.6).*

Recall also from this section, u *defined on* Ω *is said to be continuous in time with values in* L^2 *if its extension by 0 outside* Ω *belongs to* $C^0([0, T_0]; L^2(\mathbb{R}^d))$*; for* $s \in \mathbb{N}$*, we say that* u *is continuous with values in* H^s *if the derivatives* $\partial_x^\alpha u$ *for* $|\alpha| \le s$ *are continuous in time with values in* L^2*. We denote these spaces by* $C^0H^s(\Omega)$.

There are analogous definitions and notations for functions defined on $\Omega \times \mathbb{T}^p$ *or* $\Omega \times \mathbb{R}^m$.

THEOREM 7.19 (*Existence of the principal profile under the uniform weak coherence assumption*). *Suppose that Assumptions 7.13, 7.16 and (H1) [resp. (H2)] are satisfied. For all initial data* $\mathcal{P}\mathbf{u}_{0|t=0}$ *given in* $H^s(\omega \times \mathbb{T}^p)$ *with* $s > \frac{d+p}{2} + 1$ *[resp.* $\mathbb{E}^s(\omega)$ *with* $s > \frac{d}{2}$*], there exists* $T > 0$ *such that the profile equations (7.41) (7.44) have a unique solution* $\mathbf{u}_0 \in C^0H^s((\Omega \cap \{t \le T\}) \times \mathbb{T}^p)$ *[resp.* $C^0\mathbb{A}^s(\Omega \cap \{t \le T\})$*]*

For details, we refer the reader to [74,75].

7.3.5. Main profiles for the Cauchy problem The previous theorem solves the profile equations knowing the value of $\mathbf{u}_0$ at time $t = 0$. When considering the Cauchy problem with oscillating data (7.30), one has to lift the initial profile $\mathbf{h}$ to an initial condition for $\mathbf{u}_0$.

We consider the framework of Assumption 7.14 with a group of phases $\mathcal{F} = \Lambda \cdot \Psi$ and $\mathcal{F}_0 = \Lambda_0 \cdot \Psi_0$. We choose a left inverse ℓ from $\widetilde{\mathcal{F}}_0$ to $\widetilde{\mathcal{F}}$ of the restriction map $\rho : \varphi \mapsto \varphi_{|t=0}$, so that

$$\widetilde{\mathcal{F}} = \ell\widetilde{\mathcal{F}}_0 \oplus \widetilde{\mathcal{F}}_1, \qquad \widetilde{\mathcal{F}}_1 = \ker \rho. \tag{7.48}$$

We can choose accordingly a basis in $\widetilde{\mathcal{F}}_0$ and a basis in $\widetilde{\mathcal{F}}_0$, defining, accordingly, functions $\Psi_0 \in C^\infty(\omega; \mathbb{R}^{m_0})$ and $\Psi_1 \in C^\infty(\Omega; \mathbb{R}^{m_1})$. The former is lifted in $\ell\Psi_0 \in C^\infty(\Omega; \mathbb{R}^{m_0})$, so that the functions in $\widetilde{\mathcal{F}}$ can be represented as $\alpha_0 \cdot \ell\Psi_0 + \alpha_1 \cdot \Psi_1$, with $\alpha_0 \in \mathbb{R}^{m_0}$ and $\alpha_1 \in \mathbb{R}^{m_1}$. The group $\mathcal{F} \subset \widetilde{\mathcal{F}}$ can be written in this decomposition showing that one can assume that

$$\Lambda \subset \Lambda_0 \times \Lambda_1, \qquad \Lambda_1 \subset \mathbb{R}^{m_1} \tag{7.49}$$

and that for $\alpha = (\alpha_0, \alpha_1) \in \Lambda$

$$\alpha \cdot \Psi = \alpha_0 \cdot \ell\Psi_0 + \alpha_1 \cdot \Psi_1. \tag{7.50}$$

The Lemma 7.15 gives the natural lifting $\mathbf{h} \mapsto \mathbf{u}_{0|t=0}$ for trigonometric polynomials. To extend it to spaces of profiles, we need some assumptions to ensure the convergence of Fourier series.

ASSUMPTION 7.20. With the notations of Assumption 7.14 we suppose that Λ_0 is finitely generated.

Choosing a $\mathbb{Z}$-basis in $\mathcal{F}_0$, we determine a matrix M_0 such that $\Lambda_0 = M_0\mathbb{Z}^{p_0}$. Let $\Phi_0 = {}^tM\Psi_0$. We represent the initial oscillations with profiles $\mathbf{h}(x, \theta_0)$, which are periodic in θ_0:

$$\mathbf{h}(x, \theta_0) = \sum_{\alpha_0 \in \mathbb{Z}^{p_0}} \widehat{\mathbf{h}}_{\alpha_0}(x)e^{i\alpha_0 \cdot \theta_0} \quad \in H^s(\omega \times \mathbb{T}^{p_0}). \tag{7.51}$$

In accordance with (7.50) we define for $\alpha_0 \in \mathbb{Z}^{p_0}$ and $\alpha_1 \in \Lambda_1$

$$(\alpha_0, \alpha_1) \cdot \Phi = \alpha_0 \cdot \ell\Phi_0 + \alpha_1 \cdot \Psi_1, \qquad (M\alpha_0) \cdot \ell\Psi_0 + \alpha_1 \cdot \Psi_1 \tag{7.52}$$

and accordingly we look for profiles $\mathbf{u}$ which are functions of $(t, x) \in \Omega$, $\theta_0 \in \mathbb{T}^{p_0}$ and $Y_1 \in \mathbb{R}^{m_1}$.

For fixed t, the profiles $\mathbf{u}$ we will consider will live in spaces

$$\mathbb{E}^s := C^0_{pp}(\mathbb{R}^{m_1}; H^s(\omega \times \mathbb{T}^{p_0})) \tag{7.53}$$

of almost periodic functions of Y_1 valued in $H^s(\omega \times \mathbb{T}^{p_0})$. This space is the closure in $L^\infty(\mathbb{R}^{m_1} : H^s(\omega \times \mathbb{T}^{p_0}))$ of the space of trigonometric polynomials, that is finite sums

$$\mathbf{v}(x, \theta_0, Y_1) = \sum_{\alpha_1} \widehat{\mathbf{v}}_{\alpha_1}(x, \theta_0)e^{i\alpha_1 \cdot Y_1} \tag{7.54}$$

with coefficients $\widehat{\mathbf{v}}_{\alpha_1} \in H^s(\omega\times\mathbb{T}^{p_0})$. Each coefficient can be in its turn expanded in Fourier series in θ_0.

Following the notations of Assumptions 7.14 and 7.20, we denote by $\tilde{\mathcal{C}}_\Lambda$ the set of $(\alpha_0,\alpha_1) \in \mathbb{Z}^{p_0} \times \Lambda_1$ such that $(M_0\alpha_0,\alpha_1)$ belongs to Λ and the corresponding phase $(\alpha_0,\alpha_1)\cdot\Phi$ is characteristic. With little risk of confusion, we denote by $P_{(\alpha_0,\alpha_1)}$ and $Q_{(\alpha_0,\alpha_1)}$ the associated projectors. For $(\alpha_0,\alpha_1) \notin \tilde{\mathcal{C}}_\Lambda$, we set $P_{(\alpha_0,\alpha_1)} = 0$ and $Q_{(\alpha_0,\alpha_1)} = \mathrm{Id}$.

The projectors $\mathcal{P}$ and $\mathcal{Q}$ are obviously defined on trigonometric polynomials,

$$\sum \widehat{\mathbf{v}}_{\alpha_0,\alpha_1}(x)\mathrm{e}^{\mathrm{i}(\alpha_0\cdot\theta_0+\alpha_1\cdot Y_1)}.$$

Lemma 7.21. *The projector $\mathcal{P}$ extends as a bounded projector in $\mathbb{E}^0$.*

Proof. For a trigonometric polynomial with coefficients $\mathbf{v}_{\alpha_0,\alpha_1}$, let $\mathbf{w} = \mathcal{P}\mathbf{v}$ and define

$$V_{\alpha_0}(x,Y_1) = \sum_{\alpha_1} \widehat{\mathbf{v}}_{\alpha_0,\alpha_1}(x)\mathrm{e}^{\mathrm{i}\alpha_1\cdot Y_1},$$
$$W_{\alpha_0}(x,Y_1) = \sum_{\alpha_1} \widehat{\mathbf{w}}_{\alpha_0,\alpha_1}(x)\mathrm{e}^{\mathrm{i}\alpha_1\cdot Y_1} = \sum_{\alpha_1} P_{(\alpha_0,\alpha_1)}\widehat{\mathbf{v}}_{\alpha_0,\alpha_1}(x)\mathrm{e}^{\mathrm{i}\alpha_1\cdot Y_1}.$$

Because the number of α_1 such that $(\alpha_0,\alpha_1) \in \tilde{\mathcal{C}}_\Lambda$ is at most N and the projectors are bounded, the following holds

$$\begin{aligned}\left|W_{\alpha_0}(x,Y_1)\right|^2 &\le NC\sum_{\alpha_1}\left|\widehat{\mathbf{v}}_{\alpha_0,\alpha_1}(x)\right|^2\\ &= NC\lim_{R\to\infty}\frac{1}{(2R)^{m_1}}\int_{[-R,R]^{m_1}}\left|V_{\alpha_0}(x,Y)\right|^2\mathrm{d}Y.\end{aligned}$$

Thus, by Fatou's lemma,

$$\begin{aligned}\|\mathbf{w}(\cdot,\cdot,Y_1)\|^2_{L^2(\omega\times\mathbb{T}^{p_0})} &\le NC\lim_{R\to\infty}\frac{1}{(2R)^{m_1}}\\ &\quad\times\int_{[-R,R]^{m_1}}\|\mathbf{w}(\cdot,\cdot,Y)\|^2_{L^2(\omega\times\mathbb{T}^{p_0})}\,\mathrm{d}Y\\ &\le NC\sup_{Y\in\mathbb{R}^{m_1}}\|\mathbf{w}(\cdot,\cdot,Y)\|^2_{L^2(\omega\times\mathbb{T}^{p_0})}.\end{aligned}$$

This proves that $\mathcal{P}$ is bounded on trigonometric polynomials for the norm of $\mathbb{E}^0$. The lemma follows by density. □

Using Assumption 7.16, one can repeat the proof for derivatives and $\mathcal{P}$ acts from $\mathbb{E}^s$ to $\mathbb{E}^s$. There is a similar proof for the action of $\mathrm{Id} - \mathcal{Q}$ and thus for $\mathcal{Q}$. Finally, for $\mathbf{h} \in H^s(\omega\times\mathbb{T}^{p_0})$, one can construct

$$\boldsymbol{\ell}\mathbf{h}(x,\theta_0,Y_1) = \sum_{\alpha_0,\alpha_1} P_{(\alpha_0,\alpha_1)}\widehat{\mathbf{h}}_{\alpha_0}(x)\mathrm{e}^{\mathrm{i}(\theta_0\cdot\theta_0+\alpha_1\cdot Y_1)} \in \mathbb{E}^s \tag{7.55}$$

using again the fact that for each α_0 there are at most N non-vanishing terms in the sum. Therefore:

PROPOSITION 7.22. *Under Assumptions 7.14, 7.16 and 7.20, the projectors $\mathcal{P}$ and $\mathcal{Q}$ are well defined in the spaces $\mathbb{E}^s$ for all $s \geq 0$. In addition, the lifting operator (7.55) is bounded from $H^\infty(\omega \times \mathbb{T}^{p_0})$ into $\mathbb{E}^s$ and $(\mathrm{Id} - \mathcal{P})\boldsymbol{\ell} = 0$.*

Using the Notations 7.18, we can repeat

THEOREM 7.23 (*Existence of the principal profile for the Cauchy problem under the uniform weak coherence assumption*). *Suppose that Assumptions 7.14, 7.16 and 7.20 are satisfied. For all data $\mathbf{h} \in H^s(\omega \times \mathbb{T}^{p_0})$ with $s > \frac{d+p}{2} + 1$, there exists $T > 0$ such that the profile equations (7.41) (7.44) have a unique solution $\mathbf{u}_0 \in C^0\mathbb{E}^s\,(\Omega \cap \{t \leq T\})$ satisfying*

$$\mathbf{u}_0(0, x, \theta_0, Y_1) = \boldsymbol{\ell}\mathbf{h}. \tag{7.56}$$

This is proved in [76] when $\Lambda_1 = \mathbb{R}$, but the proof extends immediately to the general case considered here. It is based on the symmetry property inherited by the profile equation, and the fact that the projectors act in the proper spaces.

7.3.6. Approximate solutions at order $o(1)$ in L^2 When $\mathbf{u}_0$ is known, to solve the equation with an error $O(\varepsilon)$, it would be sufficient to determine $\mathbf{u}_1$ such that

$$\mathcal{L}(a, \partial_Y)\mathbf{u}_1 = \mathcal{Q}\left(\mathbf{F}_0 - \mathcal{L}_1(a, \partial_{t,x,Y})\mathbf{u}_0\right), \tag{7.57}$$

that is, formally,

$$(\mathrm{Id} - \mathcal{P})\mathbf{u}_1 = \mathcal{R}\left(\mathbf{F}_0 - \mathcal{L}_1(a, \partial_{t,x,Y})\mathbf{u}_0\right). \tag{7.58}$$

Obviously, the principal difficulty is to show that $\mathcal{R}$ is a bounded operator on suitable spaces of profiles. The weak coherence Assumption 7.13 permits phases $\alpha \cdot \Psi$ such that the determinant $\det L(a, id\alpha \cdot \Psi)$ vanishes on a set of measure 0. In this case the inverse $R_\alpha = (L(a, id\alpha \cdot \Psi))^{-1}$ is defined almost everywhere but is certainly not bounded. However, this is sufficient to obtain approximate solutions in spaces of low regularity. To give an idea of a possible result consider the Cauchy problem for (7.20) in the framework of Theorem 7.23.

THEOREM 7.24. *Under the assumptions of Theorem 7.23 suppose that $\mathbf{h} \in H^s(\omega \times \mathbb{T}^{p_0})$ is given and $\mathbf{u}_0 \in C^0\mathbb{E}^s(\Omega)$ is a solution of the main profile equation with initial data (7.56). Then, there is a family of functions $u^\varepsilon \in C^0L^2(\Omega)$ such that*

$$u^\varepsilon(0, x) - \mathbf{h}(x, \Phi_0(x)/\varepsilon) \to 0 \quad \text{in } L^2(\omega),$$
$$u^\varepsilon(t, x) - \mathbf{u}_0(t, x, \Phi(t, x)/\varepsilon) \to 0 \quad \text{in } C^0L^2(\Omega)$$

and

$$A_0(a, \varepsilon u^\varepsilon)\partial_t u^\varepsilon + \sum_{j=1}^d \varepsilon A_j(a, \varepsilon u^\varepsilon)\partial_{x_j} u^\varepsilon + \frac{1}{\varepsilon}E(a)u^\varepsilon = F(u^\varepsilon) + f^\varepsilon \tag{7.59}$$

with $f^\varepsilon \to 0$ *in* $L^2(\Omega)$.

SCHEME OF THE PROOF. *Step 1.* The profile $\mathbf{f}_0 := \mathbf{F}_0 - \mathcal{L}_1(a, \partial_{t,x,Y})\mathbf{u}_0$ is known and can be approximated by trigonometric polynomials: for all $\delta > 0$, there is a finite sum $\mathbf{f}'_0 = \sum e^{i\alpha Y}\hat{\mathbf{f}}_\alpha$ such that $\|\mathbf{f}_0 - \mathbf{f}'\| \le \delta$, in the space of profiles.
Step 2. One can modify the Fourier coefficients $\hat{\mathbf{f}}_\alpha$, and assume that they are smooth and *vanish* near the negligible set where $\det L(a, d\alpha \cdot \Psi)$ vanishes. One can then ensure that

$$\left\|(\mathbf{f}_0 - \mathbf{f}'_0)(\,\cdot, \Phi(\cdot)/\varepsilon)\right\|_{L^2(\Omega)} \le 2\delta. \tag{7.60}$$

Step 3. The conditions on $\hat{\mathbf{f}}_\alpha$ allow one to define $R_\alpha \hat{\mathbf{f}}_\alpha$ and therefore an approximate first corrector $\mathbf{u}'_1$ which is a smooth trigonometric polynomial.

Consider $u^\varepsilon(t,x) = (\mathbf{u}_0 + \varepsilon\mathbf{u}'_1)(t, x, \Phi(t,x)/\varepsilon)$. It satisfies

$$\left|u^\varepsilon(t,x) - \mathbf{u}_0(t, x, \Phi(t,x)/\varepsilon)\right| \le \varepsilon K(\delta)$$

and satisfies the equation up to an error $e^\varepsilon(t,x)$ which is $(\mathbf{f}_0 - \mathbf{f}')(t, x, \Psi(t,x)/\varepsilon)$ plus terms factored by ε and involving $\mathbf{u}_1$ and $\mathbf{u}_0$. Thus

$$\left\|e^\varepsilon\right\|_{L^2} \le 2\delta + \varepsilon K(\delta).$$

Choosing $\delta = \delta(\varepsilon) \to 0$ sufficiently slowly, so that $\varepsilon K(\delta) \to 0$ proves the theorem. □

REMARK 7.25. Because we truncate the Fourier coefficients in Step 2, estimates of the error f^ε in L^2 (or in L^p but with $p < \infty$) is the best that one can expect with this method.

7.3.7. Strong coherence The difficulty caused by the phases such that $\det L(a, i\mathrm{d}\varphi)$ vanishes on a small set, is not only a technical difficulty met in the definition of the corrector $\mathbf{u}_1$, but the examples of Section 7.1.6 show that this problem can have severe consequences. To avoid them, the following condition is natural:

ASSUMPTION 7.26 (*Strong coherence*). We are given an additive subgroup $\mathcal{F}$ of $C^\infty(\Omega; \mathbb{R})$ such that the vector space $\widetilde{\mathcal{F}}$ generated by $\mathcal{F}$ is finite dimensional over $\mathbb{R}$. We assume that

(i) for all $\varphi \in \widetilde{\mathcal{F}}$ the dimension of the kernel of $\det L(a(t,x), id\varphi(t,x))$ is independent of $(t,x) \in \overline{\Omega}$,
(ii) for all $\varphi \in \widetilde{\mathcal{F}} \setminus \{0\}$, $d\varphi(t,x) \ne 0$ for all $(t,x) \in \overline{\Omega}$.

EXAMPLE 7.27. An important example is the case where $a = \underline{a}$ is constant and $\mathcal{F}$ is a group of planar phases $\{\beta_0 t + \sum \beta_j x_j, \beta \in \Lambda \subset \mathbb{R}^{1+d}\}$. We refer to [74,76] for other examples and a detailed discussion of this assumption. We just mention here the case of phases generated by $\{|x| \pm t, t\}$ in the case of the wave equation, or Maxwell equations, or more generally in the case of spherically invariant systems.

In particular, (i) implies that each phase $\varphi \in \widetilde{\mathcal{F}}$ is either characteristic everywhere or nowhere characteristic. This assumption implies that all the partial inverses R_α are well defined everywhere on $\overline{\Omega}$. Therefore Step 2 in the proof of Theorem 7.24 can be eliminated.

REMARK 7.28. When the strong coherence assumption is satisfied, the approximation results in Theorem 7.24 can be improved, so that the errors tend to zero in L^∞.

7.3.8. Higher order profiles, approximate solutions at all order Under Assumption 7.26, for each $\alpha \in \Lambda$, the partial inverse $R_\alpha(t,x)$ is well defined everywhere on $\overline{\Omega}$ and smooth. However, the partial inverses are not necessarily uniformly bounded with respect to α since the determinant $\det L(a, id(\alpha \cdot \Psi))$ can be very small. The summation of the Fourier series for $(\mathrm{Id} - \mathcal{P})\mathbf{u}_1$ can be delicate and ultimately impossible in certain cases. This is *a small divisors problem* which seems difficult to solve in the general framework of almost periodic profiles.

ASSUMPTION 7.29 (*Small divisors conditions*). Suppose that Assumption 7.26 is satisfied and that $\mathcal{F}$ is finitely generated. Taking a basis $(\varphi_1, \ldots, \varphi_p)$, and representing $\Lambda = M\mathbb{Z}^p$ as in (7.23), we assume that for all $\beta \in \mathbb{N}^{1+d}$ there are constants $c > 0$ and ν such that for all $\alpha \in \mathbb{Z}^d$ and for all $(t,x) \in \overline{\Omega}$:

$$\left|\partial_{t,x}^{\beta} R_{M\alpha}(t,x)\right| \le C(1+|\alpha|)^{\nu}. \tag{7.61}$$

EXAMPLE 7.30. When $a(t,x) = \underline{a}$ is constant and the phases are planar, the matrices $R_{M\alpha}$ are independent of (t,x). The condition (7.61) splits into two conditions: there are c and ν such that, when $M\alpha \cdot \Psi$ is not characteristic

$$\left|\det L(\underline{a}, id(M\alpha \cdot \Psi))\right| \ge c(1+|\alpha|)^{-\nu}, \tag{7.62}$$

and when $M\alpha \cdot \Psi$ is characteristic, the nonvanishing eigenvalues $\lambda(\alpha)$ of $L(\underline{a}, id(M\alpha \cdot \Psi))$ satisfy

$$|\lambda(\alpha)| \ge c(1+|\alpha|)^{-\nu}. \tag{7.63}$$

Under Assumption 7.29, for $\mathbf{F} \in H^\infty(\Omega \times \mathbb{T}^p)$ the Fourier series defining $\mathcal{R}\mathbf{F}$ converges and its sum belongs to $\mathbf{F} \in H^\infty(\Omega \times \mathbb{T}^p)$. This immediately implies the following:

PROPOSITION 7.31. *Suppose that Assumption 7.29 is satisfied. If the $\mathbf{u}_0 \in H^\infty(\Omega \times \mathbb{T}^p)$ is given, there is $\mathbf{u}_1 \in H^\infty(\Omega \times \mathbb{T}^p)$ satisfying (7.57).*

All the ingredients being now prepared, one can construct BKW solutions with arbitrary small residuals.

THEOREM 7.32 (*Complete asymptotic solution, in the uniformly strongly coherent case, with the small divisors condition*). *Suppose that $\Omega \subset [0, T_0] \times \mathbb{R}^d$ is contained in the domain of determinacy of the initial domain ω. Suppose that Assumptions 7.16, 7.26 and 7.29 are satisfied. Given initial data for $\mathcal{P}\mathbf{u}_{n|t=0}$ in $H^\infty(\omega \times \mathbb{T}^p)$ there exists $T > 0$ and a sequence of profiles $\mathbf{u}_n \in H^\infty((\Omega \cap \{t \le T\}) \times \mathbb{T}^p)$ which satisfy (7.36) (7.37).*

We refer to [74] for details. In this paper it is also proved that the small divisor condition is generically satisfied. For instance, for a constant coefficient system, the condition is satisfied for almost all choices of p characteristic planar phases.

Whence one has a BKW solution at all orders, so one can construct approximate solutions

$$u^{\varepsilon}_{\mathrm{app},n}(t,x) = \sum_{k=0}^{n} \varepsilon^n \mathbf{u}_k(t,x,\Phi(t,x)/\varepsilon) \tag{7.64}$$

which are solutions of the equation with error terms of order $O(\varepsilon^n)$. Using Borel's Theorem, one can construct approximate solutions at infinite orders.

7.4. *Exact oscillating solutions*

In this section we briefly describe several methods which aim to construct exact solutions of (7.20) such that

$$u^{\varepsilon}(t,x) - \mathbf{u}_0(t,x,\Psi(t,x))/\varepsilon \to 0 \quad \text{as } \varepsilon \to 0 \tag{7.65}$$

where $\mathbf{u}_0$ is a principal profile.

7.4.1. Oscillating solutions with complete expansions The easiest case is when one has a BKW solution at all orders, that is under Assumption 7.29. In this case, one can construct approximate solutions $u^{\varepsilon}_{\mathrm{app},n}(t,x)$ as in (7.64). Then one can write the equation for $u^{\varepsilon} - u^{\varepsilon}_{\mathrm{app},n}$, and if n is large enough, one can use Theorem 4.21 and obtain a result analogous to Theorem 5.33. We refer to [74] for details.

7.4.2. Prepared data, continuation of solutions We consider oscillations belonging to a *finitely generated* group of phases $\mathcal{F}$ which satisfies Assumption 7.26. We look for *exact* solutions of a non-dispersive equation (7.20) (with $E = 0$) of the form:

$$u^{\varepsilon}(t,x) = \mathbf{u}^{\varepsilon}(t,x,\Phi(t,x)/\varepsilon) \tag{7.66}$$

with $\mathbf{u}^{\varepsilon}(t,x,\theta)$ periodic in θ, and Φ related to Ψ as in (7.23). For u^{ε} to solve (7.20) it is *sufficient* that $\mathbf{u}^{\varepsilon}$ solves

$$L_1(a,\varepsilon\mathbf{u}^{\varepsilon},\partial_{t,x})\mathbf{u}^{\varepsilon} + \frac{1}{\varepsilon}\mathcal{L}(a,\varepsilon\mathbf{u}^{\varepsilon},\partial_\theta)\mathbf{u}^{\varepsilon} = F(\mathbf{u}^{\varepsilon}) \tag{7.67}$$

with

$$L_1(a,v,\partial_{t,x}) = A_0(a,v)\partial_t + \sum_{j=1}^{d} A_j(a,v)\partial_{x_j}, \tag{7.68}$$

$$\mathcal{L}(a,v,\partial_\theta) = \sum_{j=1}^{p} L_1(a,v,d\varphi_j)\partial_{\theta_j}. \tag{7.69}$$

This is precisely the type of equation (4.35) of (4.45) discussed in Section 4.1.

REMARK 7.33. It is remarkable that the *condition* (4.39) *for* (7.69), which is that the rank of $L(a(t,x), d\alpha \cdot \Phi)$ is constant, *is exactly condition* (i) *of* Assumption 7.26. This indicates that the conditions of Section 4.1.3, which came from commutation requirements, are not only technical but deeply related to the focusing effects.

In order to apply Propositions 4.18 and 4.20, we are led to supplement Assumption 7.26 with the following

ASSUMPTION 7.34. Assume that there is $\underline{\psi} \in \widetilde{\mathcal{F}}$, such that the system $L(a, \partial)$ is hyperbolic in the direction $d\underline{\psi}$.

The next result follows directly from Proposition 4.20:

LEMMA 7.35. *Under Assumptions* 7.26, 7.16 and 7.34, *the singular term* $\mathcal{L}(a(t,x), 0, \partial_\theta)$ *satisfies the Assumption* 4.12.

Therefore Theorem 4.14 applies to the Cauchy problem for (7.67) with initial data

$$\mathbf{u}^\varepsilon{}_{|t=0}(x,\theta) = \mathbf{g}^\varepsilon(x,\theta) \tag{7.70}$$

(see Remark 4.15). In this context, we have to assume that the *data are prepared* in the sense of Section 4.1.2, meaning that the times derivatives $\partial_t^k \mathbf{u}^\varepsilon{}_{|t=0}$ which can be computed from the equation satisfy

$$\sup_{\varepsilon \in]0,\varepsilon_0]} \left\| \partial_t^k \mathbf{u}^\varepsilon{}_{|t=0} \right\|_{H^{s-k}(\omega \times \mathbb{T}^p)} < \infty. \tag{7.71}$$

THEOREM 7.36. *Suppose that* $\Omega \subset [0, T_0] \times \mathbb{R}^d$ *is contained in the domain of determinacy of the initial domain* ω. *Under Assumptions* 7.26 and 7.34, *for* $s > 1 + (d+p)/2$ *and initial data satisfying* (7.71), *there exists* $T > 0$ *such that for* $\varepsilon \in]0, \varepsilon_0]$ *the Cauchy problem* (7.67) (7.70) *has a solution* $\mathbf{u}^\varepsilon$ *in* $CH^s((\Omega \cap \{t \le T\}) \times \mathbb{T}^p)$, *and the family* $\{\mathbf{u}^\varepsilon\}$ *is bounded in this space.*

Moreover, by Theorem 7.19, the Cauchy problem for the principal profile is well-posed.

THEOREM 7.37. *Under the assumptions of Theorem* 7.36 *suppose that in addition,* $\mathbf{g}^\varepsilon \to \mathbf{g}_0$ *in* $H^s(\omega \times \mathbb{T}^p)$. *Then,* $\mathbf{u}^\varepsilon \to \mathbf{u}_0$ *in* $CH^{s'}((\Omega \cap \{t \le T\}) \times \mathbb{T}^p)$ *for all* $s' < s$, *where* $\mathbf{u}^0$ *satisfies the main profile equation with initial data* $\mathbf{g}_0$.

SKETCH OF PROOF. By compactness, one can extract a subsequence such that $\mathbf{u}^\varepsilon \to \mathbf{u}_0$ in $CH^{s'}((\Omega \cap \{t \le T\}) \times \mathbb{T}^p)$ for all $s' < s$, implying that $\mathbf{u}_0(0,x,\theta) = \mathbf{g}_0(x,\theta)$. Passing to the limit in the equation (7.67) multiplied by ε, implies that $\mathcal{L}(a, \partial_\theta)\mathbf{u}_0 = 0$ and thus that $\mathbf{u}_0$ satisfies the polarization condition $\mathcal{P}\mathbf{u}_0 = \mathbf{u}_0$. Next, multiplying the equation by $\mathcal{P}$ kills the singular term, and passing to the limit implies that $\mathbf{u}_0$ satisfies the profile equation. By uniqueness of the limit, the full family converges. □

CONCLUSION 7.38. *The first theorem means that we have solved the Cauchy problem for (7.20) with initial data*

$$u^{\varepsilon}{}_{|t=0}(x) = \mathbf{g}^{\varepsilon}(x, \Phi(0, x)/\varepsilon) \tag{7.72}$$

and found the solution under the form (7.66). The second theorem implies the convergence in (7.65). To apply this method, the difficult part is to check the preparation conditions (7.71). This method is better adapted to continuation problems as in Theorem 4.5 when the oscillations are created, not by the initial data, but by source terms. We leave to the reader to duplicate Theorem 4.5 in the present context.

The assumptions of the theorem are satisfied when a is constant and the phases are planar; it also applies when the space $\widetilde{\mathcal{F}}$ is strongly coherent and the system is hyperbolic with constant multiplicities in the direction $d\underline{\psi}$ of Assumption 7.34 (see Proposition 4.20).

7.4.3. Construction of solutions. General oscillating data Our goal here is to solve the Cauchy problem with more general oscillating data, in particular without assuming the preparation conditions (7.71). The idea is similar to the one developed in the previous section, but different. We consider initial data of the form

$$h^{\varepsilon}(x) = \mathbf{h}^{\varepsilon}(x, \Phi_0(x)/\varepsilon) \tag{7.73}$$

where $\mathbf{h}^{\varepsilon}$ is a bounded family in $H^s(\omega \times \mathbb{T}^{p_0})$, and we look for solutions of the form

$$u^{\varepsilon}(t, x) = \mathbf{u}^{\varepsilon}_{\flat}(t, x, \Phi_{\flat}(t, x)/\varepsilon) \tag{7.74}$$

where $\Phi_{\flat}(0, x) = \Phi_0(x)$. The $\mathbf{u}^{\varepsilon}_{\flat}$ are sought in $C^0 H^s(\Omega \times \mathbb{T}^{p_0})$.

In this section, we consider a non-dispersive system (7.20) with $E = 0$. The strategy is to use Proposition 4.20 and Theorem 4.14, thanks to the following assumption:

ASSUMPTION 7.39. We are given a finitely generated subgroup $\mathcal{F}_{\flat} \subset C^{\infty}(\overline{\Omega})$ generating a finite dimensional linear space $\widetilde{\mathcal{F}}_{\flat}$, such that

(i) for all $\psi \in \widetilde{\mathcal{F}}_{\flat} \setminus \{0\}$ and all $(t, x) \in \overline{\Omega}$, $d_x\psi(t, x) \neq 0$,

(ii) for all $\psi \in \widetilde{\mathcal{F}}_{\flat} \setminus \{0\}$ $A_0^{-1}(a(t, x))L(a(t, x), d\psi)$ has eigenvalues independent of (t, x); moreover, the eigenprojectors belong to a bounded set in $C^{\infty}(\overline{\Omega})$ independent of $\psi \in \widetilde{\mathcal{F}}$.

Choose a $\mathbb{R}$-basis $\{\psi_1, \dots, \psi_{m_{\flat}}\}$ of $\widetilde{\mathcal{F}}_{\flat}$ and a $\mathbb{Z}$-basis $\{\varphi_1, \dots, \varphi_{p_{\flat}}\}$ of $\mathcal{F}_{\flat}$. Let $M_{\flat}$ be the $m_{\flat} \times p_{\flat}$ matrix such that $\Phi_{\flat} = {}^tM\Psi_{\flat}$, where $\Phi_{\flat} = {}^t(\varphi_1, \dots, \varphi_{p_{\flat}})$ and $\Psi_{\flat} = {}^t(\psi_1, \dots, \psi_{m_{\flat}})$. We look for solutions of (7.20) of the form (7.74). The equation for $\mathbf{u}^{\varepsilon}_{\flat}$ reads

$$L_1(a, \varepsilon\mathbf{u}^{\varepsilon}_{\flat}, \partial_{t,x})\mathbf{u}^{\varepsilon}_{\flat} + \frac{1}{\varepsilon}\mathcal{L}_{\flat}(a, \varepsilon\mathbf{u}^{\varepsilon}_{\flat})\mathbf{u}^{\varepsilon}_{\flat} = F(\mathbf{u}^{\varepsilon}_{\flat}) \tag{7.75}$$

with

$$\mathcal{L}_\flat(a, v, \partial_\theta) = \sum_{j=1}^{p_\flat} L_1(a, v, d\varphi_j)\partial_{\theta_j}. \tag{7.76}$$

Under Assumption 7.39 and Proposition 4.20, we can apply Theorem 4.14 and Remark 4.15, which imply the following result:

THEOREM 7.40. *Suppose that $\Omega \subset [0, T_0] \times \mathbb{R}^d$ is contained in the domain of determinacy of the initial domain ω. Under Assumption 7.39 for $s > 1 + (d + p_\flat)/2$ with initial data $\mathbf{h}^\varepsilon$ bounded in $H^s(\omega \times \mathbb{T}^{p_\flat})$, there are $T > 0$ and $\varepsilon_0 > 0$ such that for $\varepsilon \in]0, \varepsilon_0]$ the Cauchy problem for (7.75) with initial data,*

$$\mathbf{u}^\varepsilon_{\flat\,|t=0}(x, \theta_\flat) = \mathbf{h}^\varepsilon(x, \theta_\flat), \tag{7.77}$$

has a unique solution in $C^0H^s((\Omega \cap \{t \le T\}) \times \mathbb{T}^{p_\flat})$.

REMARK 7.41. It is important to note that the solutions $u^\varepsilon_\flat$ have bounded derivatives in x, *but not in t*. In particular, *$u^\varepsilon_\flat$ has rapid oscillations in time*.

This theorem allows one to solve the Cauchy problem for (7.20) with initial data (7.73) and initial phases $\Phi_0(x) = \Phi_\flat(0, x)$. To continue the analysis, we must make the connection with the framework elaborated in the previous sections for the Cauchy problem and the description of oscillations. The key is in the next result.

PROPOSITION 7.42. *Assume that for $\xi \neq 0$, $A_0^{-1} \sum \xi_j A_j$ has at least two different eigenvalues. Suppose that $\mathcal{F}_\flat \subset \widetilde{\mathcal{F}}_\flat \subset C^\infty(\overline{\Omega})$ satisfy Assumption 7.39. Then, the space $\widetilde{\mathcal{F}}$ generated by $\widetilde{\mathcal{F}}_\flat$ and $\psi_0 = t$ is strongly coherent, and that it satisfies Assumption 7.26.*

PROOF. Let $\psi = \tau t + \psi_\flat \in \widetilde{\mathcal{F}}$. Then the kernel of $L(a(t, x), d\psi)$ is the kernel of $\tau \mathrm{Id} + A_0^{-1}(a)L(a, d\psi_\flat)$ and thus has constant dimension.

If $d\psi$ vanishes at one point, then $L(a(t, x), d\psi) = 0$ at this point, and thus everywhere. Since $A_0^{-1} \sum \xi_j A_j$ has at least two eigenvalues when $\xi \neq 0$, this implies that $d_x\psi = d_x\psi_\flat$ vanishes everywhere, and hence $\psi_\flat = 0$ and $\psi = 0$. □

Conversely, this clearly shows how we can use Theorem 7.40 in the context of the framework of Assumptions 7.14 and 7.20 for the Cauchy problem. For the convenience of the reader we sum up the assumptions.

ASSUMPTION 7.43. Suppose that:

(i) $\mathcal{F}$ is a subgroup of a strongly coherent space $\widetilde{\mathcal{F}}$ satisfying Assumption 7.26.
(ii) The mapping $\rho : \varphi \mapsto \varphi_{|t=0}$ has a kernel in $\widetilde{\mathcal{F}}$ of dimension 1, generated by $\underline{\psi}$, such that $\partial_t\underline{\psi}$ never vanishes on $\overline{\Omega}$ and $d\underline{\psi}$ is a direction of hyperbolicity of $L(a, \partial_{t,x})$. Moreover, the group $\mathcal{F}_0 = \rho\mathcal{F}$ is finitely generated.
(iii) The uniform coherence Assumption 7.16 is satisfied.

With these notations, let $\widetilde{\mathcal{F}}_\flat$ be such that $\widetilde{\mathcal{F}} = \mathbb{R}\underline{\psi} \oplus \widetilde{\mathcal{F}}_\flat$. Let ℓ be a right inverse of ρ from $\widetilde{\mathcal{F}}_0 = \rho\widetilde{\mathcal{F}}$ to $\widetilde{\mathcal{F}}_\flat$,and let $\mathcal{F}_\flat = \ell\mathcal{F}_0$.

PROPOSITION 7.44. *If in addition, possibly after changing the time function, $\underline{\psi}(t, x) = t$, then $\mathcal{F}_\flat \subset \widetilde{\mathcal{F}}_\flat$ satisfies Assumption 7.39.*

PROOF. If $\underline{\psi} = t$, saying that the dimension of the kernel of $L(a(t, x), d(\tau t + \psi_\flat))$ is constant for all τ, is equivalent to saying that the real eigenvalues of $A_0^{-1} L(a(t, x), d\psi_\flat)$ have constant multiplicity. By the hyperbolicity of $d\underline{\psi}$ all the eigenvalues are real. □

REMARK 7.45. In practice, when solving the Cauchy problem with oscillating initial data, one starts with a finitely generated group of phases $\mathcal{F}_0 \subset C^\infty(\overline{\omega})$. One constructs all the solutions of the eikonal equation with initial data in $\mathcal{F}_0$, and the group $\mathcal{F}$ that they generate. The big constraint is to check that $\mathcal{F}$ satisfies Assumption 7.26. The examples given in Section 7.1 show on one hand that this assumption is very strong, but on the other hand that in the multi-dimensional case direct and hidden focusing effects require such strong conditions.

Under Assumption 7.43 one can solve both the Cauchy problem for data (7.73) by Theorem 7.40 and the principal profile equation by Theorem 7.23. Note that the principal profile here is a function $\mathbf{u}_0(t, x, Y, \theta_\flat)$ with $Y \in \mathbb{R}$. Y is the placeholder for $\underline{\psi}/\varepsilon = t/\varepsilon$.

THEOREM 7.46 (*Rigorous justification of the geometric optics approximation*). *Suppose that $\mathbf{h}^\varepsilon$ is a bounded family in $H^s(\omega \times \mathbb{T}^{p_\flat})$ with $s > 1 + (d + p_\flat)/2$, and that $\mathbf{h}^\varepsilon \to \mathbf{h}^0$ in this space. Let $\mathbf{u}_0$ be the solution of the main profile equation given by Theorem 7.23 corresponding to $\mathbf{h}^0$. Then the Cauchy problem for (7.20) with initial data (7.73) has a solution u^ε such that*

$$\lim_{\varepsilon\to 0} \sup_{t,x} \left|u^\varepsilon(t, x) - \mathbf{u}_0(t, x, t/\varepsilon, \Phi_\flat(t, x)/\varepsilon)\right| = 0. \tag{7.78}$$

SKETCH OF PROOF (*See also [76,117,59,60] and Section 7.5*). As in Section 7.3.6, one can construct $\mathbf{u}^\varepsilon_{\mathrm{app}}(t, x, \theta_\flat)$ such that

$$\mathbf{u}^\varepsilon_{\mathrm{app}}(t, x, \theta_\flat) - \mathbf{u}_0(t, x, t/\varepsilon, \theta_\flat) \to 0$$

in L^∞ and $\mathbf{u}^\varepsilon_{\mathrm{app}}(t, x, \theta_flat)$ is an approximate solution of (7.75) in the sense that is satisfies the equation up to an error term which tends to 0 in $C^0 H^{s-1}(\Omega \times \mathbb{T}^{p_\flat})$. Then the stability property of the equation implies that $\mathbf{u}^\varepsilon_\flat - \mathbf{u}^\varepsilon_{\mathrm{app}} \to 0$ in $C^0 H^{s-1}(\Omega \times \mathbb{T}^{p_\flat})$. □

7.4.4. The case of constant coefficients and planar phases We illustrate the results of the previous sections in the important case where a is constant and the phases are planar. We consider here a quasi-linear system

$$A_0(u)\partial_t u + \sum_{j=1}^{d} A_{x_j}(u)\partial_j u = F(u). \tag{7.79}$$

In the regime of weakly nonlinear optics, the solutions are $u^\varepsilon = u_0 + O(\varepsilon)$ where u_0 is a solution with no rapid oscillations. Here we choose u_0 to be a constant, which we can take to be 0, assuming thus that $F(0) = 0$. *We choose planar initial phases and quasi-periodic initial profiles*. We therefore consider initial data of the form

$$u^\varepsilon{}_{|t=0}(x) = \varepsilon \mathbf{h}^\varepsilon(x, {}^t M_\flat x/\varepsilon) \tag{7.80}$$

where the $\mathbf{h}^\varepsilon(x, \theta_\flat)$ are periodic in $\theta_\flat$, bounded in $H^s(\mathbb{R}^d \times \mathbb{T}^{p_\flat})$, and $M_\flat$ is a $d \times p_\flat$ matrix. The initial phases are $(M_\flat \alpha) \cdot x$ for $\alpha \in \mathbb{Z}^{p_\flat}$.

Looking for solutions of the form

$$u^\varepsilon(t, x) = \varepsilon \mathbf{u}^\varepsilon_\flat(t, x, {}^t M_\flat x/\varepsilon), \tag{7.81}$$

yields the equation

$$A_0(\varepsilon \mathbf{u}^\varepsilon_\flat)\partial_t \mathbf{u}^\varepsilon_\flat + \sum_{j=1}^{d} A_j(\varepsilon \mathbf{u}^\varepsilon_\flat)\partial_{x_j} \mathbf{u}^\varepsilon_\flat + \frac{1}{\varepsilon}\sum_{j=1}^{p_\flat} B_j(\varepsilon \mathbf{u}^\varepsilon_\flat)\partial_{\theta_j} \mathbf{u}^\varepsilon_\flat = G(\varepsilon \mathbf{u}^\varepsilon_\flat)\mathbf{u}^\varepsilon_\flat \tag{7.82}$$

with symmetric matrices B_j which we do not compute explicitly here. We are in position to apply Theorem 4.2:

THEOREM 7.47. *With assumptions as above, let* $\{\mathbf{h}^\varepsilon; \varepsilon \in]0, \varepsilon_0]\}$ *be a bounded family in* $H^s(\mathbb{R}^d \times \mathbb{T}^{p_\flat})$ *with* $s > (d + p_\flat + 1)/2$. *There exists* $T > 0$ *such that for* $\varepsilon \in]0, \varepsilon_0]$ *the Cauchy problem for* (7.82) *with initial data* $\mathbf{h}^\varepsilon_\flat$ *has a unique solution* $\mathbf{u}^\varepsilon_\flat \in C^0([0, T]; H^s(\mathbb{R}^d \times \mathbb{T}^{p_\flat}))$.

The profile equation concerns profiles $\mathbf{u}_0(t, x, Y, \theta_\flat)$ periodic in $\theta_\flat$ and almost periodic in $Y \in \mathbb{R}$. The profiles satisfy the polarization condition (7.41):

$$\mathcal{P}\mathbf{u}_0 = \mathbf{u}_0 \tag{7.83}$$

and the propagation equation (7.43):

$$(\mathrm{Id} - \mathcal{Q})\big(L_1(\partial_t, \partial_x)\mathbf{u}_0 + \Gamma(\mathbf{u}_0, \partial_Y, \partial_{\theta_\flat})\mathbf{u}_0 - G(0)\mathbf{u}_0\big) = 0 \tag{7.84}$$

where

$$\Gamma(\mathbf{v}, \partial_Y, \partial_{\theta_\flat})\mathbf{w} = \mathbf{v} \cdot \nabla_u A_0(0)\partial_Y \mathbf{w} + \sum_{j=1}^{p_\flat} \mathbf{u}\nabla_u B_j(0)\partial_{\theta_j}\mathbf{w} \tag{7.85}$$

and the projectors $\mathcal{P}$ and $\mathcal{Q}$ are defined in Proposition 7.22. In particular, $\mathcal{P}$ is the projector on the kernel of $\mathcal{L}(\partial_Y, \partial_{\theta_\flat}) = A_0(0)\partial_Y + \sum B_j(0)\partial_{\theta_j}$. The initial condition for $\mathbf{u}_0$ is $\boldsymbol{\ell}\mathbf{h}$ as in (7.56).

THEOREM 7.48. *With assumptions as in Theorem 7.47, suppose that* $\mathbf{h}^\varepsilon \to \mathbf{h}$ *in* $H^s(\mathbb{R}^d \times \mathbb{T}^{p_\flat})$. *There exists* $T' > 0$ *such that*

(i) *the profile equation with initial data* $\boldsymbol{\ell}\mathbf{h}$ *has a solution* $\mathbf{u}_0 \in C^0([0, T']; \mathbb{E}^s)$ *with* $\mathbb{E}^s = C^0_{pp}(\mathbb{R}; H^s(\mathbb{R}^d \times \mathbb{T}^{p_\flat}))$;
(ii) *and*

$$\lim_{\varepsilon\to 0} \sup_{0\le t\le \min\{T,T'\}} \left\| \mathbf{u}^\varepsilon_\flat(t, \cdot, \cdot) - \mathbf{u}_0(t, \cdot, t/\varepsilon, \cdot)\right\|_{H^{s-1}(\mathbb{R}^d\times\mathbb{T}^{p_\flat})} = 0. \tag{7.86}$$

7.5. *Asymptotics of exact solutions*

In this section we present the principles of different methods which can be used to prove that exact solutions have an asymptotic expansion.

7.5.1. The method of simultaneous approximations *The principle of the method.* On one hand, there is an equation for exact solutions

$$\partial_t u^\varepsilon = M^\varepsilon(u^\varepsilon, \partial)u^\varepsilon \tag{7.87}$$

and on the other hand there is another equation for the profiles

$$\partial_t \mathbf{u} = \mathcal{M}(\mathbf{u}, \partial)\mathbf{u}. \tag{7.88}$$

The equations are supplemented by initial conditions. The profile $\mathbf{u}$ may depend on more variables than the functions u^ε. Moreover there is a law $\mathbf{u} \mapsto U^\varepsilon(\mathbf{u})$ which assigns a family of functions to a profile. Both equations are solved by iteration:

$$\partial_t u^\varepsilon_{\nu+1} = M^\varepsilon(u^\varepsilon_\nu, \partial)u^\varepsilon_{\nu+1}, \tag{7.89}$$

$$\partial_t \mathbf{u}_{\nu+1} = \mathcal{M}(\mathbf{u}_\nu, \partial)\mathbf{u}_{\nu+1}. \tag{7.90}$$

The principle of the method relies on the following elementary lemma:

LEMMA 7.49. *Suppose that*

(1) *the sequence* u^ε_ν *converges to* u^ε, *uniformly in* ε *in a space* X;
(2) *the sequence* $\mathbf{u}_\nu$ *converges to* $\mathbf{u}$ *in a space* $\mathcal{X}$;
(3) *the mapping* $\mathbf{u} \mapsto \{U^\varepsilon(\mathbf{u}), \varepsilon \in]0, \varepsilon_0]\}$ *maps continuously* $\mathcal{X}$ *to the space of bounded families in* X *equipped with the topology of uniform convergence;*
(4) *for the initial term* $\nu = 0$ *of the induction,* $u^\varepsilon_0 - U^\varepsilon(\mathbf{u}_0) \to 0$ *in* X;
(5) *for each* $\nu \ge 0$, *assuming that* $u^\varepsilon_\nu - U^\varepsilon(\mathbf{u}_\nu) \to 0$ *in* X, *then* $u^\varepsilon_{\nu+1} - U^\varepsilon(\mathbf{u}_{\nu+1}) \to 0$ *in* X.

Then $u^\varepsilon - U^\varepsilon(\mathbf{u}) \to 0$ *in* X.

PROOF. By (4) and (5), the property (6), $u^\varepsilon_\nu - U^\varepsilon(\mathbf{u}_\nu) \to 0$ in X, is true for all ν. By (1) for $\delta > 0$, there is ν_0 such that for all $\nu \geq \nu_0$ and all $\varepsilon \in]0, \varepsilon_0]$:

$$\|u^\varepsilon - u^\varepsilon_\nu\|_X \leq \delta.$$

Moreover, by (2) and (3), for ν large enough, the following holds

$$\|U^\varepsilon(\mathbf{u}) - U^\varepsilon(\mathbf{u}_\nu)\|_X \leq \delta.$$

Thus, choosing ν such that both properties are satisfied and using (6), we see that

$$\limsup_{\varepsilon\to 0} \|u^\varepsilon - U^\varepsilon(\mathbf{u})\|_X \leq 2\delta$$

and the lemma follows. □

Of course, this lemma is just a general principle which can (and often must) be adapted to particular circumstances of the cases under examination.

REMARK 7.50. In this strategy, the main step is the fifth, which amounts to prove that for *a linear equation* whose coefficients have given asymptotic expansions, the solution has an asymptotic expansion.

Application to the oscillating Cauchy problem. This method has been applied in $d = 1$ with spaces X of C^1 functions and spaces of profiles $\mathcal{X}$ of class C^1 in all the variables and almost periodic in the fast variables (see [82] and Section 7.6.1 below).

It has been applied also in the multidimensional case $d > 1$, for coherent almost periodic oscillations in the Wiener Algebra and semi-linear equations (see Theorem 7.19 for the context and [75] for precise results.)

We now briefly review how it is applied to prove the second part of Theorem 7.46 as in [76]. For simplicity, we consider the case of Eq. (7.79) and planar phases as in Theorem 7.48. The equations for exact solutions are given in (7.82) and for the profiles in (7.83) (7.84). The iterative scheme leads to linear problems

$$A_0(\varepsilon \mathbf{v}^\varepsilon_\flat)\partial_t \mathbf{u}^\varepsilon_\flat + \sum_{j=1}^{d} A_j(\varepsilon \mathbf{v}^\varepsilon_\flat)\partial_{x_j}\mathbf{u}^\varepsilon_\flat + \frac{1}{\varepsilon}\sum_{j=1}^{p_\flat} B_j(\varepsilon \mathbf{v}^\varepsilon_\flat)\partial_{\theta_j}\mathbf{u}^\varepsilon_\flat = G(\varepsilon \mathbf{v}^\varepsilon_\flat)\mathbf{v}^\varepsilon_\flat \tag{7.91}$$

and

$$\begin{cases} \mathcal{P}\mathbf{u} = \mathbf{u} \\ (\mathrm{Id} - \mathcal{Q})\big(L_1(\partial_t, \partial_x)\mathbf{u} + \Gamma(\mathbf{v}, \partial_Y, \partial_{\theta_\flat})\mathbf{u} - G(0)\mathbf{v}\big) = 0. \end{cases} \tag{7.92}$$

The key point is to prove that if the $\mathbf{v}^\varepsilon$ are bounded in $C^0[0,T]; H^s(\mathbb{R}^d \times \mathbb{T}^{p_\flat})$, if $\mathbf{u} \in C^0([0,T], \mathbb{E}^s(\mathbb{R} \times \mathbb{R}^d \times \mathbb{T}^{p_\flat}))$, and if $\mathbf{v}^\varepsilon_\flat(t, x, \theta_\flat) - \mathbf{v}(t, x, t/\varepsilon, \theta_\flat)$ tends to 0 in $C^0([0,T]; H^{s-1}(\mathbb{R}^d \times \mathbb{T}^{p_\flat}))$, then $\mathbf{u}^\varepsilon_\flat(t, x, \theta_\flat) - \mathbf{u}(t, x, t/\varepsilon, \theta_\flat)$ tends to 0 in $C^0([0,T]; H^{s-1}(\mathbb{R}^d \times \mathbb{T}^{p_\flat}))$. This means that the problem of the asymptotic description of solutions is completely transformed into linear problems. Using approximations, one can restrict the analysis to the case where $\mathbf{v}$ is a trigonometric polynomial, and in this case we can use BKW solutions.

7.5.2. The filtering method This method has been introduced in [86,87], with motivation coming from fluid Mechanics, in particular with applications to the analysis of the low Mach limit of Euler equations. Next it was extended to more general contexts such as (7.82) in [88,115–118], and then adapted to various other problems by many authors.

The principle of the method. The method applies to evolution equations where two time scales (at least) are present:

$$\partial_t u^\varepsilon + \frac{1}{\varepsilon} G u^\varepsilon + \Phi(u^\varepsilon) = 0, \qquad u^\varepsilon{}_{|t=0} = h^\varepsilon \tag{7.93}$$

where G generates a C^0 group e^{-tG}, which is uniformly bounded in a Hilbert space H_1, and also uniformly bounded in $H \subset H_1$. Moreover, Φ is a bounded family of continuous functionals from H to H_1 which map bounded sets in H to bounded sets in H_1.

It is assumed that there is a preliminary construction which provides us with a family of solutions u^ε bounded in $C^0([0,T]; H)$. The problem is to describe the asymptotic behavior of u^ε. Assume that the embedding $H \subset H_1$ is compact and $h^\varepsilon \to h$ in H. The problem then is to describe the fast oscillations in time of u^ε. The idea is very classical in the study of evolution systems, and can be related to themes like adiabatic limits or the construction of wave operators in scattering theory: set

$$v^\varepsilon(t) = \mathrm{e}^{\frac{t}{\varepsilon}G} u^\varepsilon(t) \Leftrightarrow u^\varepsilon(t) = \mathrm{e}^{-\frac{t}{\varepsilon}G} v^\varepsilon(t). \tag{7.94}$$

Thus v^ε is bounded in $C^0([0,T]; H)$ and

$$\partial_t v^\varepsilon(t) = -\mathrm{e}^{\frac{t}{\varepsilon}G} \Phi_\varepsilon(u^\varepsilon(t)) \tag{7.95}$$

is bounded in $C^0([0,T]; H_1)$. Therefore, v^ε is compact in $C^0([0,T]; H_1)$, and up to the extraction of a subsequence

$$v^\varepsilon \to v \quad \text{in } C^0([0,T]; H_1), \qquad v \in L^\infty([0,T]; H). \tag{7.96}$$

The asymptotic behavior of u^ε is given by

$$u^\varepsilon(t) - \mathrm{e}^{-\frac{t}{\varepsilon}G} v(t) \to 0 \quad \text{in } C^0([0,T]; H_1). \tag{7.97}$$

Two scales of time are present in the main term $\mathrm{e}^{-\frac{t}{\varepsilon}G} v(t)$, yielding the profile

$$\mathbf{u}(t, Y) := \mathrm{e}^{-YG} v(t) \tag{7.98}$$

so that

$$u^\varepsilon(t) = \mathbf{u}(t, t/\varepsilon). \tag{7.99}$$

By construction, $\mathbf{u}$ satisfies the "polarization" condition

$$(\partial_Y + G)\mathbf{u}(t, Y) = 0. \tag{7.100}$$

Note that in (7.98), v appears as a parametrization of the profiles. The slow evolution of the profile is determined from the slow evolution of v, which is obtained by *passing to the weak limit* in Eq. (7.95). Because we already know the approximation (7.97), the slow evolution reads

$$\partial_t v + \underline{\Phi}(v(t)) = 0 \tag{7.101}$$

where $\underline{\Phi}$ is the nonlinear operator defined by

$$\underline{\Phi}(v) = w - \lim_{\varepsilon\to 0} \mathrm{e}^{\frac{t}{\varepsilon}G}\Phi(\mathrm{e}^{-\frac{t}{\varepsilon}G}v). \tag{7.102}$$

Application to the oscillatory Cauchy problem. We sketch the computation in the case of Eq. (7.82) and planar phases as in Theorem 7.48. The equation is of the form (7.93) with

$$G = \sum_{j=1}^{p_\flat} A_0^{-1}B_j(0)\partial_{\theta_j}$$

and a nonsingular term Φ_ε given by:

$$\begin{aligned} A_0(0)\Phi_\varepsilon(\mathbf{u}) = (A_0(\varepsilon\mathbf{u}) - A_0(0))\,\partial_t\mathbf{u} + \sum A_j(\varepsilon\mathbf{u})\partial_{x_j}\mathbf{u} \\ + \sum \frac{1}{\varepsilon}\big(B_j(\varepsilon\mathbf{u}) - B_j(0)\big)\,\partial_{\theta_j}\mathbf{u} - G(\varepsilon\mathbf{u})\mathbf{u}. \end{aligned}$$

The fast evolution is made explicit in Fourier series in $\theta_\flat$:

$$\mathrm{e}^{-YG}\left(\sum_\alpha \hat{v}_\alpha \mathrm{e}^{\mathrm{i}\alpha\theta_\flat}\right) = \sum_{\tau,\alpha} \mathrm{e}^{\mathrm{i}(\tau Y + \alpha\theta_\flat)} P_{\tau,\alpha}\hat{v}_\alpha \tag{7.103}$$

where the $P_{\tau,\alpha}$ are the spectral projectors on $\ker(\tau\,\mathrm{Id} + \sum \alpha_j A_0^{-1}B_j)$. In particular, Proposition 7.22 implies that e^{-tG} is bounded in the spaces $H^s(\mathbb{R}^d \times \mathbb{T}^{p_\flat})$.

In accordance with Theorem 7.47, consider a family of solutions $\mathbf{u}_\flat^\varepsilon$ of (6.60) bounded in $C^0([0, T]; H^s(\mathbb{R}^d \times \mathbb{T}^{p_\flat}))$. Then $\mathbf{v}^\varepsilon = \mathrm{e}^{\frac{t}{\varepsilon}G}\mathbf{u}_\flat^\varepsilon(t)$ is bounded in $C^0([0, T]; H^s)$ and $\partial_t\mathbf{v}^\varepsilon$

is bounded in $C^0([0,T], H^{s-1})$, so that extracting a subsequence,

$$\mathbf{v}^\varepsilon \to \mathbf{v} \quad \text{in } C^0([0,T], H^{s'}_{loc}),$$

for all $s' < s$. This implies that

$$\left\| u^\varepsilon_\flat(t,\,\cdot\,) - \mathbf{u}_0(t, t/\varepsilon,\,\cdot\,)\right\|_{H^{s'}_{loc}} \to 0 \tag{7.104}$$

where the profile $\mathbf{u}_0$ is defined by

$$\mathbf{u}_0(t, Y,\,\cdot\,) = \mathrm{e}^{-YG}\mathbf{v}(t,\,\cdot\,). \tag{7.105}$$

The explicit form (7.103) of the evolution implies that $\mathbf{u}_0$ belongs to a class $C^0([0,T];\mathbb{E}^{s'})$ (see Proposition 7.22) and satisfies the polarization condition $\mathcal{P}\mathbf{u}_0 = \mathbf{u}_0$.

The evolution of $\mathbf{v}$ is obtained by passing to the limit in Eq. (7.95) for $\partial_t\mathbf{v}^\varepsilon$. Using (7.104), the problem is reduced to the following computation:

$$\partial_t\mathbf{v}(t,\,\cdot\,) = -w - \lim_{\varepsilon\to 0} \mathrm{e}^{\frac{t}{\varepsilon}G}\Phi_0\left(\mathbf{u}_0(t, t/\varepsilon,\,\cdot\,)\right) \tag{7.106}$$

where

$$\Phi_0(\mathbf{u}) = \sum A_j(0)\partial_{x_j}\mathbf{u} + \Gamma(\mathbf{u}, \partial_{Y,\theta})\mathbf{u} - G(0)\mathbf{u}$$

and Γ is given by (7.85).

LEMMA 7.51. *If $\mathbf{f}$ is a profile in $C^0([0,T];\mathbb{E}^s)$, then $\mathrm{e}^{\frac{t}{\varepsilon}G}\mathbf{f}(t, t/\varepsilon,\,\cdot\,)$ converges in the sense of distributions to $\mathcal{P}\mathbf{f}(t, 0,\,\cdot\,)$.*

PROOF. It is sufficient to prove the convergence when $\mathbf{f}$ is a trigonometric polynomial, that is a finite sum

$$\mathbf{f}(t, Y, x, \theta_\flat) = \sum_{\tau,\alpha} \hat{\mathbf{f}}_{\tau,\alpha}(t, x)\mathrm{e}^{\mathrm{i}\tau Y + \alpha\theta_\flat}.$$

In this case

$$e^{\frac{t}{\varepsilon}G}\mathbf{f}(t, t/\varepsilon, x, \theta_\flat) = \sum_{\lambda,\tau\alpha} \mathrm{e}^{\mathrm{i}(\tau-\lambda)t/\varepsilon} P_{\lambda,\alpha}\hat{\mathbf{f}}_{\tau,\alpha}(t, x)\mathrm{e}^{\mathrm{i}\alpha\theta_\flat}$$

converges weakly to

$$\sum_{\tau\alpha} P_{\tau,\alpha}\hat{\mathbf{f}}_{\tau,\alpha}(t, x)\mathrm{e}^{\mathrm{i}\alpha\theta_\flat} = \mathcal{P}\mathbf{f}(t, 0, x, \theta_\flat). \qquad \square$$

COROLLARY 7.52. *The main profile satisfies the propagation equation*

$$\partial_t \mathbf{u}_0 + \mathcal{P}\Phi_0(\mathbf{u}_0) = 0,$$

that is the profile equation (7.84).

PROOF. Let $\mathbf{f} = \Phi_0(\mathbf{u}_0)$. Then $\partial_t \mathbf{v}(t, \cdot) = -\mathcal{P}\mathbf{f}(t, 0, \cdot)$ and

$$\partial_t \mathbf{u}_0(t, Y, \cdot) = \mathrm{e}^{-YG}\partial_t \mathbf{v}(t, \cdot) = -\mathrm{e}^{-YG}\mathcal{P}\mathbf{f}(t, 0, \cdot) = -\mathcal{P}\mathbf{f}(t, Y, \cdot)$$

where the last identity follows immediately from (7.103). □

REMARK 7.53. By the uniqueness of the profile equation, it follows that the limit is independent of the extracted subsequence so that the full sequence $\mathbf{v}^\varepsilon$ converges, and the approximation (7.104) holds for the complete family as $\varepsilon \to 0$.

7.6. *Further examples*

7.6.1. One dimensional resonant expansions In the one dimensional case $d = 1$, strong coherence assumptions such as Assumption 7.26 are not necessary. Weak coherence is sufficient for the construction of the main profiles (as in Theorem 7.23). The strong coherence conditions were used to construct solutions in Sobolev spaces. In 1-D, one can use different methods such as integration along characteristics and construct solutions in L^∞ (or $W^{1,\infty}$ is the quasi-linear case) avoiding the difficulty of commutations. Moreover, there is a special way to describe the interaction operators in 1-D which is interesting in itself.

For simplicity, we consider strictly hyperbolic non-dispersive, semilinear systems, refereeing to [68,70,82,85] for the quasilinear case. After a diagonalization and a change of dependent variables, the equations are

$$X_k u_k := \partial_t u_k + \lambda_k(t, x)\partial_x u_k = f_k(t, x, u_1, \dots, u_N), \tag{7.107}$$

for $k \in \{1, \dots, N\}$. The λ_k are real and $\lambda_1 < \lambda_2 < \lambda_3$. The f_k are smooth functions of the variables $(t, x, u) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^N$.

The initial oscillations are associated wth phases in a group $\mathcal{F}_0 \subset C^\infty(\overline{\omega})$, or in the vector space $\widetilde{\mathcal{F}}_0$ they span, where ω is a bounded interval in $\mathbb{R}$. We assume that $\widetilde{\mathcal{F}}_0$ is finite dimensional and that for all $\varphi \in \widetilde{\mathcal{F}}_0$, the derivative $\partial_x \varphi$ is different from zero almost everywhere. Thus the initial data will be of the form

$$u_k^\varepsilon(0, x) = h_k^\varepsilon(x) \sim \mathbf{h}_k(x, \Psi_0(x)/\varepsilon) \tag{7.108}$$

where we have chosen a basis $\{\psi_{0,1}, \dots, \psi_{0,m}\}$ and $\Psi_0 \in C^\infty(\overline{\omega}; \mathbb{R}^m)$ is the function with components $\psi_{0,j}$.

For all k we consider the solutions of the kth eikonal equation

$$X_k \varphi = 0, \qquad \varphi_{|t=0} \in \mathcal{F}_0. \tag{7.109}$$

It is one of the main features of the 1D case, that the eikonal equation factors are products of linear equations. In particular, the set of solutions of (7.109) is a group $\mathcal{F}^k$, isomorphic to $\mathcal{F}_0$, and all the phases are defined and smooth on $\overline{\Omega}$ the domain of determinacy of $\overline{\omega}$. Similarly, the vector space $\widetilde{\mathcal{F}}^k$ spanned by $\mathcal{F}^k$ is the set of solutions of $X_k\varphi = 0$ with initial data in the vector space $\widetilde{\mathcal{F}}_0$. The initial phases $\psi_{0,j}$ are propagated by X_k, defining bases $\{\psi_{k,1}, \dots, \psi_{k,m}\}$ of $\widetilde{\mathcal{F}}_k$ and functions $\Psi_k \in C^\infty(\overline{\Omega}; \mathbb{R}^m)$ satisfying

$$X_k \Psi^k = 0, \qquad \Psi^k{}_{|t=0} = \Psi_0.$$

The complete group of phases which are obtained by nonlinear interaction is $\mathcal{F} = \mathcal{F}^1 + \cdots + \mathcal{F}^N$, generating the vector space $\widetilde{\mathcal{F}} = \widetilde{\mathcal{F}}^1 + \cdots + \widetilde{\mathcal{F}}^N$. Note that, by construction,

$$\bigl\{\varphi \in \widetilde{\mathcal{F}};\ X_k\varphi = 0\bigr\} = \widetilde{\mathcal{F}}^k. \tag{7.110}$$

We assume that $\widetilde{\mathcal{F}}$ satisfies the weak coherence assumption, which here means that

$$\forall \varphi \in \widetilde{\mathcal{F}} \setminus \widetilde{\mathcal{F}}_k : \quad X_k\varphi \neq 0, \quad \text{a.e. on } \Omega. \tag{7.111}$$

To follow the general discussion in the previous section, we should introduce a basis of $\widetilde{\mathcal{F}}$, a function Ψ and accordingly represent the solution as in (7.22). This can be done, but it does not give the simplest and most natural description. Indeed, the diagonal form of the system and (7.110) imply that the projector $\mathcal{P}$ has a simple form: the oscillation with phase $\varphi \neq 0$ in $\mathcal{F}$ is characteristic if it belongs to one of the $\mathcal{F}^k$, and then the projector P is simply the projector on the kth axis (we admit here, or assume, that the index k is necessarily unique). Thus, the polarization condition (7.41) implies that the kth component of the main term depends only on the phases in $\widetilde{\mathcal{F}}^k$ so that its main oscillation is of the form

$$u_k^\varepsilon(t, x) \sim \mathbf{u}_k(t, x, \Psi_k(t, x)/\varepsilon) \tag{7.112}$$

with $\mathbf{u}_k(t, x, Y)$ periodic / quasi-periodic / almost-periodic in $Y \in \mathbb{R}^m$.

One should consider different copies $Y^k \in \mathbb{R}^m$ of the variable Y, one for each component. However, *when replaced by* Ψ^k/ε, *the variables* Y^k *are not necessarily independent*. This is precisely the phenomenon of *resonance* or *phase matching*. As a result, the description of the projector $\mathcal{P}$ is not immediate. This is the price of using the description (7.112).

DEFINITION 7.54. A resonance is an N-tuple $(\varphi_1, \dots, \varphi_N)$ of functions such that $X_k\,\varphi_k = 0$ for all k and $\sum d\varphi_k = 0$. The resonance is trivial when all the $d\varphi_k$ are identically zero.

Note that the definition does not depend on the vector fields themselves but on the foliations they define. The data of N foliations by curves in $\mathbb{R}^2$ is called an N-web in differential geometry, and resonances correspond to Abelian relations on webs (see [111, 5]). The existence of resonances is a rare phenomenon: *for an N-web, the dimension of the space of resonances (modulo constants functions) is equal to, at most,* $(N-1)(N-2)/2$. Moreover, for "generic" vector fields X_k, there are no nontrivial resonances. Therefore, in order to have resonances, first the vector fields X_k must be suitably chosen, and second, the phases φ_k must also to be chosen carefully. The case $N = 3$ is well illustrated by the example of Section 7.1.2. However, resonance is a very important phenomenon. Constant coefficient vector fields $X_k = \partial_t - \lambda_k \partial_x$, are resonant with maximal dimension. We refer to [73] for a more general discussion of resonances and examples.

NOTATIONS 7.55. 1. *We denote by* $\vec{\Psi}$ *the function* $(\Psi^1, \dots, \Psi^N) \in C^\infty(\overline{\Omega}; (\mathbb{R}^m)^N)$.

2. *The resonances (within the set of phases under consideration) are described by the vector space* $R \subset (\mathbb{R}^m)^M$ *of the* $\vec{\alpha} = (\alpha^1, \dots, \alpha^N)$, *such that* $\vec{\alpha} \cdot \vec{\Psi} := \sum \alpha^k \cdot \Psi^k = 0$.
3. *The combinations of phases which are characteristic for* X_k *are described by the vector space* $R_k \subset (\mathbb{R}^m)^M$ *of the* $\vec{\alpha}$,*such that* $\vec{\alpha} \cdot \vec{\Psi} \in \widetilde{\mathcal{F}}_k$. R_k *is the sum of* R *and* $A_k := \{\vec{\alpha}\ \forall l \neq k, \alpha^l = 0\}$. *This sum is direct since* $\alpha^k \cdot \Psi^k = 0$ *only if* $\alpha^k = 0$. *We denote by* ρ_k *the mapping*

$$\rho_k : R_k \mapsto \mathbb{R}^m \quad \text{such that } \vec{\alpha} \cdot \vec{\Psi} = \rho_k(\vec{\alpha}) \cdot \Psi^k. \tag{7.113}$$

4. *We denote by* $\vec{Y} = (Y^1, \dots, Y^N)$ *the variable in* $(\mathbb{R}^m)^N$, *dual to the* $\vec{\alpha}$, *and* $\vec{\alpha} \cdot \vec{Y} = \sum \alpha^k \cdot Y^k$. *The natural vector space for the placeholder of* $\vec{\Psi}$ *in the profiles is* $V = R^\perp = \{\vec{Y}\ \forall \vec{\alpha} \in R, \vec{\alpha} \cdot \vec{Y} = 0\}$.
5. *We denote by* $V_k \subset V$ *the orthogonal of* R_k.
6. *The projection* $\pi_k : \vec{Y} \mapsto Y^k$ *is surjective from* V *to* $\mathbb{R}^m$, *since if* $\alpha^k \cdot Y^k = 0$ *for all* $\vec{Y} \in V$, *this implies that* $\alpha^k \cdot \Psi^k = 0$ *and thus* $\alpha^k = 0$. *Its kernel* $\{\vec{Y} \in V \ :\ Y^k = 0\}$ *is* V_k. *Therefore there is a natural isomorphism*

$$V/V_k \leftrightarrow \mathbb{R}^m. \tag{7.114}$$

Equivalently, one can choose a lifting linear operator $\pi_k^{(-1)}$ *from* $\mathbb{R}^m$ *to* V *such that* $\pi_k \pi_k^{(-1)} = \mathrm{Id}$. *In particular, with (7.113), there holds*

$$\forall \vec{\alpha} \in R_k, \forall Y^k \in \mathbb{R}^m, \qquad \vec{\alpha} \cdot \pi_k^{(-1)} Y^k = \rho_k(\vec{\alpha}) \cdot Y^k. \tag{7.115}$$

With these notations, we can now describe the projectors. We consider profiles *which are almost periodic*: given a finite dimensional space E, denote by $C^0_{pp}(\overline{\Omega} \times E)$ the space of functions $\mathbf{u}(t, x, Y)$ which are almost periodic in $Y \in E$, that is the closure in L^∞ of finite sums $\sum \hat{\mathbf{u}}_\alpha(t, x) \mathrm{e}^{\mathrm{i}\alpha \cdot Y}$ Given such profiles $\mathbf{u}_k(t, x, Y^k)$, which are almost periodic in Y^k, the nonlinear coupling $f_k(t, x, \mathbf{u}_1, \dots, \mathbf{u}_N)$ appears as a function

$$\mathbf{f}(t, x, \vec{Y}), \quad (t, x) \in \Omega, \vec{Y} \in V, \tag{7.116}$$

which is almost periodic in $\vec{Y}$. Expanding $\mathbf{f}_k$ in Fourier series, the projector $\mathbf{E}_k$ must keep [resp. eliminate] the exponentials $\mathrm{e}^{\mathrm{i}\vec{\alpha} \cdot \vec{\Psi}}$ when $\vec{\alpha} \cdot \vec{\Psi} \in \widetilde{\mathcal{F}}^k$, [resp. $\notin \widetilde{\mathcal{F}}^k$]. Translated to the Y variables, this means that

$$\begin{cases} \mathbf{E}_k(\mathrm{e}^{\mathrm{i}\vec{\alpha} \cdot \vec{Y}}) = 0 & \text{when } \vec{\alpha} \notin R_k, \\ \mathbf{E}_k(\mathrm{e}^{\mathrm{i}\vec{\alpha} \cdot \vec{Y}}) = \mathrm{e}^{\mathrm{i}\rho_k(\vec{\alpha}) \cdot Y^k} & \text{when } \vec{\alpha} \in R_k. \end{cases} \tag{7.117}$$

This operator is linked to the *averaging operator with respect to* V_k acting on almost periodic functions:

$$\mathbf{M}_k\mathbf{f}(\vec{Y}) = \lim_{T\to\infty} \frac{1}{\operatorname{vol}(TB_k)} \int_{TB_k} \mathbf{f}(\vec{Y} + \vec{Y}')\, \mathrm{d}\vec{Y}' \tag{7.118}$$

where B_k is any open sphere (or cube) in V_k and $\mathrm{d}\vec{Y}'$ a Lebesgue measure on V_k, which also measures the volume in the denominator. This average is invariant by translations in V_k, so that $\mathbf{M}_k\mathbf{f}$ can be seen as a function on V/V_k. Using (7.114), it can be seen as a function of $Y^k \in \mathbb{R}^m$. Equivalently, using the lifting $\pi_k^{(-1)}$, we set

$$\mathbf{E}_k\mathbf{f}(Y^k) = \mathbf{M}_k\mathbf{f}(\pi_k^{(-1)}Y^k) \tag{7.119}$$

which is independent of the particular choice of $\pi_k^{(-1)}$.

PROPOSITION 7.56. *Using the notations above, in the representation* (7.112) *of the oscillations, the profile equations read:*

$$\begin{aligned}&(\partial_t + \lambda_k(t,x)\partial_x)\mathbf{u}_k(t,x,Y^k)\\ &\quad = \mathbf{E}_k f_k(t,x,\mathbf{u}_1(t,x,Y^1),\dots,\mathbf{u}_N(t,x,Y^N))\end{aligned} \tag{7.120}$$

for $k = 1,\dots,N$, *with initial data*

$$\mathbf{u}_k(0,x,Y) = \mathbf{h}_k(x,Y). \tag{7.121}$$

EXAMPLE 7.57. Consider the case $N = 3$ with one initial phase $\psi_0(x)$. Let ψ_k be the solution of $X_k\psi_k = 0$ with initial data equal to ψ_0. The resonant set is $R = \{\alpha \in \mathbb{R}^3 : \sum_k \alpha_k\psi_k = 0\}$. There are two cases:

– either $R = \{0\}$; then averaging operators are defined by

$$(\mathbf{E}_1 F)(t,x,\theta_1) := \lim_{T\to+\infty} \frac{1}{T^2} \int_0^T \int_0^T F(\theta_1,\theta_2,\theta_3)\, \mathrm{d}\theta_2\, \mathrm{d}\theta_3,$$

with similar definitions for $\mathbf{E}_2$ and $\mathbf{E}_3$.

– or there is a nontrivial resonance $\alpha \in R \setminus \{0\}$, and

$$(\mathbf{E}_1 F)(t,x,\theta_1) := \lim_{T\to+\infty} \frac{1}{T} \int_0^T F\left(\theta_1, -\frac{\alpha_1\theta_1 + \alpha_3\sigma}{\alpha_2}, \sigma\right) \mathrm{d}\sigma,$$

with similar formulas for $k = 2$ and $k = 3$. Since the vector fields are pairwise linearly independent, all three components α_k of $\alpha \in R \setminus \{0\}$ are not equal to 0. Thus the right-hand side makes sense.

THEOREM 7.58 (*[82]*). *Suppose that* Ω *is contained in the domain of determinacy of the interval* $\omega \subset \mathbb{R}$. *Suppose that* $\mathbf{h}_k \in C^0_{pp}(\omega \times \mathbb{R}^m)$, *and that* $\{u^\varepsilon_{0,k}\}_{\varepsilon\in]0,1]}$ *is a bounded* *family in* $L^\infty(\omega)$ *such that*

$$u^\varepsilon_{0,k}(.) - \mathbf{h}_k(., \Psi_0(.)/\varepsilon) \to 0 \quad in\ L^1(\omega), \quad as\ \varepsilon \to 0. \tag{7.122}$$

Then there exists $T > 0$ *such that:*

(i) *for all* $\varepsilon \in]0,1]$, *the Cauchy problem for* (7.107) *with Cauchy data* $u^\varepsilon_{0,k}$ *has a unique solution* $u^\varepsilon = (u^\varepsilon_1, \dots, u^\varepsilon_N)$ *in* $C^0(\overline{\Omega}_T)$, *where* $\Omega_T := \Omega \cap \{0 < t < T\}$,

(ii) *the profile equations* (7.120) (7.121) *have a unique solution* $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_N)$ *with* $\mathbf{u}_k \in C^0_{pp}(\Omega_T \times \mathbb{R}^m)$,

(iii)

$$u^\varepsilon_k(.) - \mathbf{u}_k(., \Psi_k(.)/\varepsilon) \to 0 \quad in\ L^1(\Omega_T), \quad as\ \varepsilon \to 0. \tag{7.123}$$

REMARKS 7.59. In (7.122) (7.123) the L^1 norm can be replaced by L^p norms for all finite p. There is a similar result in L^∞, if the condition (7.111) is slightly reinforced, requiring that $X_k\varphi \neq 0$ a.e. on all integral curves of X_k when $\varphi \in \widetilde{\mathcal{F}} \setminus \widetilde{\mathcal{F}}_k$.

2. A formal derivation of the profile equations is given in [68,105,70]. Partial rigorous justifications were previously given by [122,72,85].

3. The results quoted above concern continuous solutions of semilinear first order systems, or C^1 solutions in the quasi-linear case. For weak solutions of quasi-linear systems of conservation laws, which may present shocks, the validity of weakly nonlinear geometric optics is now proved in a general setting in [20] generalizing a result of [118] (see also [38]) The interaction of (strong) shock waves or contact discontinuities and small amplitude oscillations is described in [32–34].

7.6.2. Generic phase interaction for dispersive equations We give here a more explicit form of the profile equation (7.44), in cases that are important in applications to nonlinear optics.

We consider a constant coefficient dispersive system

$$L(\varepsilon\partial_{t,x})u = \varepsilon A_0 \partial_t u + \sum_{j=1}^{d} \varepsilon A_j \partial_{x_j} u + Eu = F(u), \tag{7.124}$$

with $F(u) = O(|u|^3)$. We assume that the cubic part of F does not vanish and we denote it by $F_3(u,u,u)$. In this context, the weakly nonlinear regime concerns waves of amplitude $O(\varepsilon^{\frac{1}{2}})$.

We consider four characteristic planar phases φ_j which satisfy the resonance (or phase matching) condition:

$$\varphi_1 + \varphi_2 + \varphi_3 + \varphi_4 = 0. \tag{7.125}$$

We denote by $\mathcal{F}$ the group that they generate. It has finite rank ≤ 3. We make the following assumption (where the dispersive character is essential):

ASSUMPTION 7.60. In $\mathcal{F}\setminus\{0\}$, the only characteristic phases are $\pm\varphi_j$ for $j \in \{1,2,3,4\}$. Moreover, we assume that the $d\varphi_j$ are regular points of the characteristic variety.

In our general framework, we would choose a basis of $\mathcal{F}$, for instance $(\varphi_1, \varphi_2, \varphi_3)$, and use profiles $\mathbf{u}(t, x, \theta_1, \theta_2, \theta_3)$ periodic in θ_1, θ_2 and θ_3. However, the assumption above and the polarization condition (7.41) imply that $\mathbf{u}_0$ has only 9 non-vanishing coefficients, and is better represented as

$$\begin{aligned}\mathbf{u}_0(t,x,\theta) = \underline{\mathbf{u}}_{0,0}(t,x) + \mathbf{u}_{0,1}(t,x,\theta_1) + \mathbf{u}_{0,2}(t,x,\theta_2)\\ + \mathbf{u}_{0,3}(t,x,\theta_3) + \mathbf{u}_{0,4}(t,x,-\theta_1-\theta_2-\theta_3)\end{aligned}$$

with $\mathbf{u}_{0,j}$ periodic in *one* variable with the only harmonics $+1$ and -1:

$$\mathbf{u}_{0,j}(t,x,\theta) = a_j(t,x)\mathrm{e}^{\mathrm{i}\theta} + \tilde{a}_j(t,x)\mathrm{e}^{-\theta}.$$

We are interested in real solutions, meaning that $\tilde{a}_j$ is the complex conjugate of a_j, with mean value equal to zero, meaning that $\underline{\mathbf{u}}_{0,0} = 0$. This corresponds to waves of the form

$$u^\varepsilon(t,x) = \varepsilon^{\frac{1}{2}} \sum_{j=1}^{4} a_j(t,x)\mathrm{e}^{\mathrm{i}\varphi_j/\varepsilon} + \varepsilon^{\frac{1}{2}} \sum_{j=1}^{4} \overline{a_j}(t,x)\mathrm{e}^{-\mathrm{i}\varphi_j/\varepsilon}. \tag{7.126}$$

The a_j must satisfy the polarization condition

$$a_j \in \ker L(id\varphi_j). \tag{7.127}$$

Introducing the group velocity v_j associated with the regular point $d\varphi_j$ of the characteristic variety, the dynamics of a_j is governed by the transport field $\partial_t + \mathrm{v}_j \cdot \nabla_x$.

For a cubic interaction $(\mathrm{Id} - \mathcal{Q})F_3(\mathbf{u}_0, \mathbf{u}_0, \mathbf{u}_0)$, the αth Fourier coefficient is

$$\sum_{\beta_1+\beta_2+\beta_3=\alpha} (\mathrm{Id} - Q_\alpha)F_3(\hat{\mathbf{u}}_{0,\beta_1}, \hat{\mathbf{u}}_{0,\beta_2}, \hat{\mathbf{u}}_{0,\beta_3}). \tag{7.128}$$

Applied to the present case, this shows that the equations have the form

$$(\partial_t + \mathrm{v}_1 \cdot \nabla_x)a_1 = r_1 F_3(\overline{a}_2, \overline{a}_3, \overline{a}_4), \tag{7.129}$$

$$(\partial_t + \mathrm{v}_2 \cdot \nabla_x)a_2 = r_1 F_3(\overline{a}_3, \overline{a}_4, \overline{a}_1), \tag{7.130}$$

$$(\partial_t + \mathrm{v}_3 \cdot \nabla_x)a_3 = r_1 F_3(\overline{a}_4, \overline{a}_1, \overline{a}_2), \tag{7.131}$$

$$(\partial_t + \mathrm{v}_3 \cdot \nabla_x)a_4 = r_1 F_3(\overline{a}_1, \overline{a}_2, \overline{a}_3) \tag{7.132}$$

where the r_j are projectors on $\ker L(i\mathrm{d}\varphi_j)$.

Note that the condition $\underline{\mathbf{u}}_{0,0} = 0$ is consistent with Eq. (7.44) for cubic nonlinearities.

This is the *generic form of four wave interaction*. There are analogues for three wave interaction. These settings cover fundamental phenomena such as *Raman scattering*, *Brillouin scattering* or *Rayleigh scattering*, see e.g. [6,10,107,109].

7.6.3. A model for Raman interaction We consider here a simplified model, based on a one dimensional version of the Maxwell–Bloch equation (see Section 2). The electric field E is assumed to have a constant direction, orthogonal to the direction of propagation. B is perpendicular both to E and the axis of propagation. The polarization P is parallel to E. This leads to the following set of equations

$$\begin{cases} \partial_t b + \partial_x e = 0, \\ \partial_t e + \partial_x b = -\partial_t \operatorname{tr}(\Gamma\rho), \\ \mathrm{i}\varepsilon\partial_t \rho = [\Omega, \rho] - e[\Gamma, \rho] \end{cases}$$

where e and b take their values in $\mathbb{R}$ and Γ is a Hermitian symmetric matrix, with entries in $\mathbb{C}$. In this case, note that $\operatorname{tr}(\Gamma[\Gamma, \rho]) = 0$ and

$$\varepsilon\partial_t \operatorname{tr}(\Gamma\rho) = \frac{1}{i} \operatorname{tr}(\Gamma[\Omega, \rho]).$$

A classical model to describe Raman scattering in one space dimension (see e.g. [10,107, 109]) uses a three level model for the electrons, with Ω having three simple eigenvalues, $\omega_1 < \omega_2 < \omega_3$. Moreover, the states 1 and 2 have the same parity, while the state 3 has the opposite parity. This implies that the interaction coefficient $\gamma_{1,2}$ vanishes. Finally, we assume that

$$\Omega = \begin{pmatrix} \omega_1 & 0 & 0 \\ 0 & \omega_2 & 0 \\ 0 & 0 & \omega_3 \end{pmatrix}, \qquad \Gamma = \begin{pmatrix} 0 & 0 & \gamma_{1,3} \\ 0 & 0 & \gamma_{2,3} \\ \gamma_{3,1} & \gamma_{3,2} & 0 \end{pmatrix}.$$

As noticed in Section 5.3.2, the scaling of the waves requires some care for Maxwell-Bloch equations (see [73]). The density matrix ρ is assumed to be a perturbation of the ground state

$$\underline{\rho} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The scaling for amplitudes is

$$\begin{aligned} & e = \varepsilon\tilde{e}, \quad b = \varepsilon\tilde{b}, \\ & \rho_{1,1} = 1 + \varepsilon^2 \tilde{\rho}_{1,1}, \qquad \rho_{1,k} = \varepsilon\tilde{\rho}_{1,k} \quad \text{for } k = 2, 3, \\ & \rho_{j,k} = \varepsilon^2 \tilde{\rho}_{j,k} \quad \text{for } 2 \le j, k \le 3. \end{aligned} \tag{7.133}$$

The different scaling for the components of ρ is consistent with the fact that ρ must remain a projector, as implied by the Liouville equation for ρ.

Substituting in the equation, and neglecting $O(\varepsilon^2)$ terms which do not affect the principal term of the expansions, yields the following model system for $u = (\tilde{b}, \tilde{e}, \tilde{\rho}_{1,2}, \tilde{\rho}_{1,3}, \tilde{\rho}_{2,1}, \tilde{\rho}_{3,1})$

$$\begin{cases} \varepsilon\partial_t\tilde{b} + \varepsilon\partial_x\tilde{e} = 0 \\ \varepsilon\partial_t\tilde{e} + \varepsilon\partial_x\tilde{b} - \mathrm{i}\omega_{3,1}(\gamma_{1,3}\tilde{\rho}_{3,1} - \gamma_{3,1}\tilde{\rho}_{1,3}) \\ \qquad + \mathrm{i}\varepsilon\omega_{3,2}(\gamma_{2,3}\tilde{\rho}_{3,1}\tilde{\rho}_{1,2} - \gamma_{3,2}\tilde{\rho}_{2,1}\tilde{\rho}_{1,3}) = 0, \\ \varepsilon\partial_t\tilde{\rho}_{1,3} - \mathrm{i}\omega_{3,1}\tilde{\rho}_{1,3} + \mathrm{i}\tilde{e}\gamma_{1,3} + \mathrm{i}\varepsilon\tilde{e}\tilde{\rho}_{1,2}\gamma_{2,3} = 0, \\ \varepsilon\partial_t\tilde{\rho}_{3,1} + \mathrm{i}\omega_{3,1}\tilde{\rho}_{3,1} - \mathrm{i}\tilde{e}\gamma_{3,1} - \mathrm{i}\varepsilon\tilde{e}\gamma_{3,2}\tilde{\rho}_{2,1} = 0, \\ \varepsilon\partial_t\tilde{\rho}_{1,2} - \mathrm{i}\omega_{2,1}\tilde{\rho}_{1,2} + \mathrm{i}\varepsilon\tilde{e}\tilde{\rho}_{1,3}\gamma_{3,2} = 0, \\ \varepsilon\partial_t\tilde{\rho}_{2,1} + \mathrm{i}\omega_{2,1}\tilde{\rho}_{2,1} - \mathrm{i}\varepsilon\tilde{e}\gamma_{2,3}\tilde{\rho}_{3,1} = 0. \end{cases}$$

This is of the form

$$L(\varepsilon\partial_{t,x})u + \varepsilon f(u) = 0 \tag{7.134}$$

where f is quadratic. As a consequence of the equality $\gamma_{1,2} = 0$, the linear operator $L(\varepsilon\partial_{t,x})$ splits into two independent systems, where we drop the tildes:

$$L_1(\varepsilon\partial_{t,x}) \begin{pmatrix} b \\ e \\ \rho_{1,3} \\ \rho_{3,1} \end{pmatrix} := \begin{cases} \varepsilon\partial_t b + \varepsilon\partial_x e, , \\ \varepsilon\partial_t e + \varepsilon\partial_x b - \mathrm{i}\omega_{3,1}(\gamma_{1,3}\tilde{\rho}_{3,1} - \gamma_{3,1}\tilde{\rho}_{1,3}), \\ \varepsilon\partial_t\rho_{1,3} - \mathrm{i}\omega_{3,1}\rho_{1,3} + \mathrm{i}e\gamma_{1,3}, \\ \varepsilon\partial_t\rho_{3,1} + \mathrm{i}\omega_{3,1}\rho_{3,1} - \mathrm{i}e\gamma_{3,1}, \end{cases}$$

and

$$L_2(\varepsilon\partial_{t,x}) \begin{pmatrix} \rho_{1,2} \\ \rho_{2,1} \end{pmatrix} := \begin{cases} \varepsilon\partial_t\rho_{1,2} - \mathrm{i}\omega_{2,1}\rho_{1,2}, \\ \varepsilon\partial_t\rho_{2,1} + \mathrm{i}\omega_{2,1}\rho_{2,1}. \end{cases}$$

The characteristic varieties of L_1 and L_2 are respectively

$$\begin{cases} \mathcal{C}_{L_1} = \{(\tau, \xi) \in \mathbb{R}^2; \xi^2 = \tau^2(1 + \chi(\tau))\}, \qquad \chi(\tau) := \dfrac{2\omega_{3,1}|\gamma_{1,3}|^2}{(\omega_{3,1})^2 - \tau^2}, \\ \mathcal{C}_{L_1} = \{(\tau, \xi) \in \mathbb{R}^2; \tau = \pm\omega_{2,1}\}. \end{cases}$$

Raman interaction occurs when a laser beam of wave number $\beta_L = (\omega_L, \kappa_L) \in \mathcal{C}_{L_1}$ interacts with an electronic excitation $\beta_E = (\omega_{2,1}, \kappa_E) \in \mathcal{C}_{L_2}$ to produce a scattered wave $\beta_S = (\omega_S, \kappa_S) \in \mathcal{C}_{L_1}$ via the resonance relation

$$\beta_L = \beta_E + \beta_S.$$

One further assumes that $\beta_L \notin \mathcal{C}_{L_2}$, $\beta_E \notin \mathcal{C}_{L_1}$ and $\beta_S \notin \mathcal{C}_{L_2}$.

We represent the oscillations, using the two independent phases

$$\varphi_L(t,x) = \omega_L t + \kappa_L x, \qquad \varphi_S(t,x) = \omega_S t + \kappa_S x.$$

The third phase φ_E is a linear combination of these two phases:

$$\varphi_E = \varphi_L - \varphi_S. \tag{7.135}$$

Accordingly, we use profiles $\mathbf{u}$ depending on two fast variables θ_L and θ_S, so that

$$u^\varepsilon(t,x) \sim \mathbf{u}(t,y,\varphi_L/\varepsilon,\varphi_S/\varepsilon). \tag{7.136}$$

Because f is quadratic, the conditions on $(\beta_L,\beta_S,\beta_E)$ show that the profile equations have solutions with spectra satisfying

$$\operatorname{spec}(\mathbf{u}) \subset \{\pm(1,0), \pm(0,1), \pm(1,-1)\}.$$

We now restrict our attention to these solutions. It is convenient to call $u_{\pm L}$, $u_{\pm S}$ and $u_{\pm E}$ the Fourier coefficients $\hat{\mathbf{u}}_{\pm 1,0}$, $\hat{\mathbf{u}}_{0,\pm 1}$ and $\hat{\mathbf{u}}_{\pm 1,\mp 1}$, respectively. Thus, for real solutions

$$\begin{aligned} u^\varepsilon(t,x) = {} & u_L(x)\mathrm{e}^{\mathrm{i}\varphi_L(t,x)/\varepsilon} + u_S(x)\mathrm{e}^{\mathrm{i}\varphi_S(t,x)/\varepsilon} + u_E(x)\mathrm{e}^{\mathrm{i}\varphi_E(t,x)/\varepsilon} \\ & u_{-L}(x)\mathrm{e}^{-\mathrm{i}\varphi_L(t,x)/\varepsilon} + u_{-S}(x)\mathrm{e}^{-\mathrm{i}\varphi_S(t,x)/\varepsilon} + u_{-E}(x)\mathrm{e}^{-\mathrm{i}\varphi_E(t,x)/\varepsilon}. \end{aligned}$$

In addition, we use the notations b_L, e_L, $\rho_{j,k,L}$ for the components of u_L etc.

LEMMA 7.61. *The polarization conditions for the principal profile read*

$$\begin{aligned} & b_{\pm E} = e_{\pm E} = \rho_{1,3,\pm E} = \rho_{3,1,\pm E} = 0, \\ & \rho_{1,2,\pm L} = \rho_{1,2,\pm S} = \rho_{2,1,\pm L} = \rho_{2,1,\pm S} = 0, \\ & \rho_{1,2,-E} = \rho_{2,1,E} = 0, \end{aligned} \tag{7.137}$$

and

$$\begin{aligned} & b_L = -\frac{1}{\omega_L}\kappa_L e_L, \qquad \rho_{1,3,L} = \frac{\gamma_{1,3}}{\omega_{3,1} - \omega_L} e_L, \\ & \qquad \rho_{3,1,L} = \frac{\gamma_{3,1}}{\omega_{3,1} + \omega_L} e_L, \end{aligned} \tag{7.138}$$

with similar formula when L is replaced by $-L$ and $\pm S$. Furthermore, for real fields and Hermitian density matrices, one has the relations

$$e_{-L} = \overline{e_L}, \qquad e_{-S} = \overline{e_S}, \qquad \rho_{2,1,-E} = \overline{\rho_{1,2,E}}. \tag{7.139}$$

Therefore, the profile equations for $\mathbf{u}_0$ involve only (e_L, e_S) and $\sigma_E := \rho_{1,2,E}$:

THEOREM 7.62. *For real solutions, the evolution of the principal term is governed by the system*

$$\begin{cases}(\partial_t + v_L\partial_y)e_L + \mathrm{i}c_1 e_S\sigma_E = 0,\\ (\partial_t + v_S\partial_y)e_S + \mathrm{i}c_2 e_S\overline{\sigma_E} = 0,\\ \partial_t\sigma_E + \mathrm{i}c_3 e_L\overline{e_S} = 0\end{cases} \tag{7.140}$$

where v_L and v_S are the group velocities associated with the frequencies β_L and β_S, respectively. The complete fields are recovered using the polarization conditions stated in the preceding lemma.

This is the familiar form of the equations relative to three wave mixing. We refer, for example, to [10,107] for an explicit calculation of the constants c_k (see also [9]). *The Raman instability* (spontaneous or stimulated) is related to the amplification properties of this system.

7.6.4. Maximal dissipative equations For maximal dissipative systems, energy solutions exist and are stable. In this context the weakly nonlinear approximations can be justified in the energy norm. The new point is that profiles need not be continuous, they belong to L^p spaces, in which case the substitution $\theta = \Psi/\varepsilon$ has to be taken in the sense of Definition 7.11.

In $\mathbb{R}^{1+d}$, consider the Cauchy problem:

$$\partial_t u + \sum_{j=1}^{d} A_j\partial_j u + f(u) = 0, \quad u_{|t=0} = h. \tag{7.141}$$

ASSUMPTION 7.63. The matrices A_j are symmetric with constant coefficients and the distinct eigenvalues $\lambda_k(\xi)$ of $A(\xi) = \sum \xi_j A_j$ have constant multiplicity for $\xi \neq 0$. We denote by Π_k the orthogonal spectral projector on $\ker A(\xi) - \lambda_k(\xi)\mathrm{Id}$.

ASSUMPTION 7.64 (*Maximal monotonicity*). f is a C^1 function from $\mathbb{C}^N$ to $\mathbb{C}^N$, with $f(0) = 0$ and there are $p \in [1, +\infty[$ and constants $0 < c < C < +\infty$ such that for all $(t, x) \in \mathbb{R}^{1+d}$ and all u and v in $\mathbb{C}^N$,

$$\begin{aligned}&|f(u)| \leq C|u|^p, \qquad |\partial_{u,\overline{u}} f(u)| \leq C|u|^{p-1},\\ &\mathrm{Re}\,((f(u) - f(v)) \cdot (\overline{u} - \overline{v})) \geq c|u - v|^{p+1}.\end{aligned}$$

Solutions of the Cauchy problem (7.141) are constructed in [121] (see also [99] or [61]): we consider smooth domains $\omega \subset \mathbb{R}^d$ and $\Omega \subset [0 + \infty[\times\mathbb{R}^d$contained in the domain of determination of ω. The key ingredients are the energy estimates which follow from the monotonicity assumptions:

$$\|u(t)\|^2_{L^2} + c\int_0^t \left\|u(t')\right\|^{p+1}_{L^{p+1}} \mathrm{d}t' \leq \|u(0)\|^2_{L^2} \tag{7.142}$$

and

$$\|u(t) - v(t)\|_{L^2}^2 + c \int_0^t \left\| u(t') - v(t') \right\|_{L^{p+1}}^{p+1} \mathrm{d}t' \leq \|u(0) - v(0)\|_{L^2}^2 . \qquad (7.143)$$

PROPOSITION 7.65. *For all $h \in L^2(\omega)$, there is a unique solution of (7.141) $u \in C^0 L^2 \cap L^{p+1}(\Omega)$.*

We assume that the initial data have periodic oscillations

$$h^\varepsilon(x) \sim \mathbf{h}(x, \varphi_0(x)/\varepsilon) \qquad (7.144)$$

with $\mathbf{h}(x, \theta)$ periodic in $\theta \in \mathbb{T}$ and $\varphi_0 \in C^\infty(\overline{\omega})$ satisfying $d\varphi_0(x) \neq 0$ for all $x \in \overline{\omega}$. For $1 \leq k \leq K$, we denote by φ_k the solution of the eikonal equation

$$\partial_t \varphi_k + \lambda_k(\partial_x \varphi_k) = 0, \qquad \varphi_{k\,|t=0} = \varphi_0.$$

We assume that the space generated by the phases is weakly coherent and that no resonances occur:

ASSUMPTION 7.66. The φ_k are defined and smooth on $\overline{\Omega}$ and $d_x \varphi_k(x) \neq 0$ at every point. Moreover,

For all $\alpha \in \mathbb{Z}^K$, $\det L(\mathrm{d}\varphi(t, x)) \neq 0$ a.e. on Ω, except when α belongs to one of the coordinate axes.

Here $L(\partial_t, \partial_x) = \partial_t + A(\partial_x)$. In this setting, the profiles are functions $\mathbf{u}(t, x, \theta)$ which are periodic in $\theta \in \mathbb{R}^K$ and periodic in θ. The projector $\mathcal{P}$ is

$$\mathcal{P} \sum (\hat{\mathbf{u}}_\alpha(t, x) \mathrm{e}^{\mathrm{i}\alpha\theta}) = \sum P_\alpha (\hat{\mathbf{u}}_\alpha(t, x) \mathrm{e}^{\mathrm{i}\alpha\theta})$$

and P_α is the orthogonal projector on $\ker L(\alpha \cdot \Phi)$. Introducing the averaging operator $\underline{\mathbf{E}}$ with respect to all the variables θ, and $\mathbf{E}_k$ with respect to the all the variables except θ_k, there holds

$$\mathcal{P}\mathbf{u} = \underline{\mathbf{E}}\mathbf{u} + \sum_{k=1}^{K} \Pi_k(\mathrm{d}\varphi_k) \mathbf{E}_k \mathbf{u}^* \qquad (7.145)$$

with $\mathbf{E}\mathbf{u} = \hat{\mathbf{u}}_0$, the average of $\mathbf{u}$, and $\mathbf{u}^* = \mathbf{u} - \hat{\mathbf{u}}_0$, its oscillation.

The polarization conditions $\mathcal{P}\mathbf{u} = \mathbf{u}$ read

$$\mathbf{u}(t, x, \theta) = \underline{\mathbf{u}}(t, x) + \sum_{k=1}^{K} \mathbf{u}_k^*(t, x, \theta_k), \qquad \mathbf{u}_k^* = \widetilde{\Pi}_k \mathbf{u}_k^* \qquad (7.146)$$

where $\widetilde{\Pi}_k = \Pi_k(\mathrm{d}\varphi_k)$ and $\mathbf{u}_k^*$ is a periodic solution with vanishing mean value. The profile equations,

$$\partial_t \mathbf{u} - \sum_{j=1}^{d} \mathcal{P} A_j \partial_j \mathbf{u} + \mathcal{P} f(\mathbf{u}) = 0,$$

decouple into a system

$$L(\partial_t, \partial_x)\underline{\mathbf{u}} + \mathbf{E} f(\mathbf{u}) = 0, \tag{7.147}$$

$$X_k \mathbf{u}_k^* + \widetilde{\Pi}_k \left(\mathbf{E}_k f(\mathbf{u})\right)^* = 0 \tag{7.148}$$

where $X_k = \widetilde{\Pi}_k L(\partial_t, \partial_x)\widetilde{\Pi}_k$ is the propagator associated with the kth eigenvalue. The initial conditions read

$$\underline{\mathbf{u}}_{|t=0}(x) = \underline{\mathbf{h}}(x), \tag{7.149}$$

$$\mathbf{u}^*_{k\,|t=0}(x, \theta_k) = \widetilde{\Pi}_k(0, x)\mathbf{h}^*(x, \theta_k) = 0 \tag{7.150}$$

where $\underline{\mathbf{h}}$ is its mean value and $\mathbf{h}^* = \mathbf{h} - \underline{\mathbf{h}}$ its oscillation.

The Eq. (7.147) inherits the dissipative properties from the original equation:

LEMMA 7.67. *If* $\mathbf{f}$ *is a profile and* $\mathbf{v}$ *satisfies the polarization condition* (7.146), *then*

$$\int_{\Omega} \underline{\mathbf{E}\mathbf{f}}\,\underline{\mathbf{v}}\,\mathrm{d}t\mathrm{d}x + \sum_{k=1}^{K} \int_{\Omega\times\mathbb{T}} \mathbf{E}_k \mathbf{f}^* \mathbf{v}_k^*\,\mathrm{d}t\mathrm{d}x\mathrm{d}\theta_k = \int_{\Omega\times\mathbb{T}^K} \mathbf{f}\mathbf{v}\,\mathrm{d}t\mathrm{d}x\mathrm{d}\theta.$$

PROOF. Because $\widetilde{\Pi}_k$ is self-adjoint and $\widetilde{\Pi}_k \mathbf{v}_k^* = \mathbf{v}_k^*$,

$$\int_{\Omega\times\mathbb{T}} \widetilde{\Pi}_k \mathbf{E}_k \mathbf{f}^* \mathbf{v}_k^*\,\mathrm{d}t\mathrm{d}x\mathrm{d}\theta_k = \int_{\Omega\times\mathbb{T}^K} \mathbf{f}(t, x, \theta)\mathbf{v}_k^*(t, x, \theta_k)\,\mathrm{d}t\mathrm{d}x\mathrm{d}\theta.$$

Adding up implies the lemma. □

Similarly,

$$\int L(\partial_t, \partial_x)\underline{\mathbf{u}}\,\underline{\mathbf{u}}\mathrm{d}x + \sum_{k=1}^{K} \int X_k \mathbf{u}_k^* \mathbf{u}_k^* \mathrm{d}t\mathrm{d}x\mathrm{d}\theta_k = \int L(\partial_t, \partial_x)\mathbf{u}\,\mathbf{u}\,\mathrm{d}t\mathrm{d}x\mathrm{d}\theta$$

and the energy method implies the following

PROPOSITION 7.68. *For all* $\mathbf{h} \in L^2(\omega \times \mathbb{T})$*, there is a unique* $\mathbf{u} \in C^0L^2 \cap L^{p+1}(\Omega \times \mathbb{T}^K)$ *of the form* (7.146)*, and satisfying the profile equations* (7.147) (7.148)*, with initial data* (7.149) (7.150).

Knowing $\mathbf{u}$, one can construct

$$v^\varepsilon(t,x) \sim \mathbf{u}(t,x,\Phi(t,x)/\varepsilon), \tag{7.151}$$

not performing directly the substitution $\theta = \Phi/\varepsilon$ in $\mathbf{u}(t,x,\theta)$, since this function is not necessarily continuous in θ, but in the sense of Definition 7.11: the family (v^ε) is $\mathcal{F}$ oscillating with profile $\mathbf{u}$, where $\mathcal{F}$ is the group generated by the φ_k. The asymptotic in (7.151) is taken in $L^{p+1}(\Omega)$ and in $C^0L^2(\Omega)$.

A key observation is that since $\mathbf{u}$ solves the profile equation, one can select v^ε satisfying approximately the original equation (in the spirit of Section 7.3.6):

PROPOSITION 7.69. *There is a bounded family* v^ε *in* $C^0L^2(\Omega) \cap L^{p+1}(\Omega)$*,* $\mathcal{F}$ *oscillating with profile* $\mathbf{u}$ *and such that*

$$\left\| L(\partial_t,\partial_x)v^\varepsilon + f(v^\varepsilon)\right\|_{L^{1+1/p}(\Omega)} \to 0, \quad \text{as } \varepsilon \to 0.$$

Using the monotonicity of the equation, one can compare v^ε with the exact solution u^ε given by Proposition 7.65:

$$\begin{aligned} &\left\|u^\varepsilon(t) - v^\varepsilon(t)\right\|_{L^2}^2 + c\int_0^t \left\|u^\varepsilon(t') - v^\varepsilon(t')\right\|_{L^{p+1}}^{p+1} \mathrm{d}t' \\ &\quad \le \left\|u^\varepsilon(0) - v^\varepsilon(0)\right\|_{L^2}^2 + c\int_0^t \left\|Lv^\varepsilon + (v^\varepsilon)(t')\right\|_{L^{1+1/p}}^{1+1/p} \mathrm{d}t'. \end{aligned}$$

This implies the following result.

THEOREM 7.70. *Suppose that* $\mathbf{h} \in L^2(\omega)$ *and that the family* h^ε *satisfies* (7.144) *in* L^2*. Let* $\mathbf{u}$ *denote the solution of the profile equation and* (u^ε) *the family of solutions of* (7.141) *with initial data* (h^ε)*. Then* u^ε *satisfies* (7.151) *in* $L^{p+1}(\Omega) \cap C^0L^2(\Omega)$.

This method, inspired from [48], is extended in [79,80] to analyze oscillations in the presence of caustics and in [21] to study oscillations near a diffractive boundary point. For the nonlinear problem, no L^∞ description is known near caustics or diffractive boundary points. The method sketched above applies to energy solutions but provides only asymptotics in L^p spaces with $p < \infty$.

References

[1] G. Alì, J.K. Hunter, *Large amplitude gravitational waves*, J. Math. Phys. **40** (1999) 3035–3052.

[2] D. Alterman, J. Rauch, *Nonlinear geometric optics for short pulses*, J. Differential Equations **178** (2002) 437–465.

[3] D. Alterman, J. Rauch, *Diffractive nonlinear geometric optics for short pulses*, SIAM J. Math. Anal. **34** (2003) 147–1502.
[4] M. Artola, A.J. Majda, *Nonlinear development of instabilities in supersonic vortex sheets I. The basic kink mode*, Physica D **28** (1987) 253–281.
[5] W. Blaschke, G. Bol, *Geometrie der Gewebe*, New York, Springer-Verlag, 1938.
[6] N. Bloembergen, *Nonlinear Optics*, W.A. Benjamin Inc., New York, 1965.
[7] J. Bona, T. Colin, D. Lannes, *Long wave approximations for water waves*, in: Mathematical and Numerical Aspects of Wave Propagation, WAVES 2003, Springer, Berlin, 2003, pp. 489–494.
[8] M. Born, E. Wolf, *Principles of Optics*, Pergamon Press, 1959.
[9] T. Boucherès, T. Colin, B. Nkonga, B. Texier, A. Bourgeade, *Study of a mathematical model for stimulated Raman scattering*, Math. Models Methods Appl. Sci. **14** (2004) 217–252.
[10] R. Boyd, *Nonlinear Optics*, Academic Press, 1992.
[11] G. Browning, H.-O. Kreiss, *Problems with different time scales for non-linear partial differential equations*, SIAM J. Appl. Math. **42** (1982) 704–718.
[12] R. Carles, *Comportements précisés près d'une caustique en optique géométrique non linéaire*, Comm. Partial Differential Equations **23** (1998) 1929–1967.
[13] R. Carles, *Geometric optics with caustic crossing for some nonlinear Schrödinger equations*, Indiana Univ. Math. J. **49** (2000) 475–551.
[14] R. Carles, *Focusing on a line for nonlinear Schrödinger equations in* $\mathbb{R}^2$, Asymptot. Anal. **24** (2000) 255–276.
[15] R. Carles, *Semi-classical Analysis for Nonlinear Schrödinger Equations*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2008.
[16] R. Carles, D. Lannes, *Focusing of a pulse with arbitrary phase shift for a nonlinear wave equation*, Bull. Soc. Math. France **131** (2003) 289–306.
[17] R. Carles, J. Rauch, *Focusing of spherical nonlinear pulses in* $\mathbb{R}^{1+3}$, Proc. Amer. Math. Soc. **130** (2002) 791–804.
[18] R. Carles, J. Rauch, *Focusing of spherical nonlinear pulses in* $\mathbb{R}^{1/3}$ *II. Nonlinear caustic*, Rev. Mat. Iberoamericana **20** (2004) 815–864.
[19] R. Carles, J. Rauch, *Focusing of spherical nonlinear pulses in* $\mathbb{R}^{1+3}$. *III. Sub and supercritical cases*, Tohoku Math. J. **56** (2004) 393–410.
[20] C. Cheverry, *Justification de l'optique géométrique non linéaire pour un systìme de lois de conservation*, Duke Math. J. **87** (1997) 213–263.
[21] C. Cheverry, *Propagation d'oscillations près d'un point diffractif*, J. Maths. Pures et Appl. **75** (1996) 419–467.
[22] C. Cheverry, *Cascade of phases in turbulent flows*, Bull. Soc. Math. France **134** (1) (2006) 33–82.
[23] C. Cheverry, O. Guès, G. Métivier, *Oscillations fortes sur un chamlinéairement dégénéré*, Ann. Sc. Ecol. Norm. Sup. **36** (2003) 691–745.
[24] C. Cheverry, O. Guès, G. Métivier, *Large amplitude high frequency waves for quasilinear hyperbolic equations*, Adv. Differential Equations **9** (2004) 829–890.
[25] J. Chikhi, *Sur la réflexion des oscillations pour un système à deux vitesses*, C. R. Acad. Sci. Paris Sér. I Math. **313** (1991) 675–678.
[26] Y. Choquet-Bruhat, *Construction de solutions radiatives approchées des équations d'Einsein*, Comm. Math. Phys. **12** (1969) 16–35.
[27] Y. Choquet-Bruhat, A. Greco, *High frequency asymptotic solutions of Y.M. and associated fields*, J. Math. Phys **24** (1983) 337–379.
[28] T. Colin, *Rigorous derivation of the nonlinear Schrödinger equation and Davey–Stewartson systems from quadratic hyperbolic systems*, Asymptot. Anal. **31** (2002) 69–91.
[29] M. Colin, T. Colin, *On a quasilinear Zakharov System describing laser-plasma interactions*, Differential Integral Equations **17** (2004) 297–330.
[30] T. Colin, D. Lannes, *Long-wave short-wave resonance for nonlinear geometric optics*, Duke Math. J. **107** (2) (2001) 351–419.
[31] T. Colin, D. Lannes, *Justification of and longwave correction to Davey–Stewartson systems from quadratic hyperbolic systems*, Discrete Contin. Dyn. Syst. **11** (2004) 83–100.
[32] A. Corli, *Weakly non-linear geometric optics for hyperbolic systems of conservation laws with shock waves*, Asymptot. Anal. **10** (1995) 117–172.
[33] A. Corli, *Asymptotic analysis of contact discontinuities*, Ann. Mat. Pura Appl. **173** (1997) 163–202.

[34] A. Corli, O. Guès, *Stratified solutions for systems of conservation laws*, Trans. Amer. Math. Soc. **353** (6) (2001) 2459–2486.

[35] R. Courant, K.O. Friedrichs, *Supersonic Flow and Shock Waves*, Wiley-Interscience, New York, 1948.

[36] R. Courant, D. Hilbert, *Methods of Mathematical Physics*, John Wiley & Sons, New York, 1962.

[37] J.-L. Delcroix, A. Bers, *Physique des plasmas* 1, 2, Inter Editions-Editions du CNRS, 1994 .

[38] R.J. DiPerna, A.J. Majda, *The validity of nonlinear geometric optics for weak solutions of conservation laws*, Comm. Math. Phys. **98** (1985) 313–347.

[39] P. Donnat, Quelques contributions mathématiques en optique non linéaire, Thèse, École Polytechnique, 1994 .

[40] P. Donnat, J.-L. Joly, G. Métivier, J. Rauch, Diffractive nonlinear optics, Séminaire E.D.P., École Polytechnique 1995–96.

[41] P. Donnat, J. Rauch, *Dispersive nonlinear geometric optics*, J. Math. Phys. **38** (1997) 1484–1523.

[42] P. Donnat, J. Rauch, *Global solvability of the Maxwell Bloch equations from nonlinear optics*, Arch. Rational Mech. Anal **136** (3) (1996) 291–303.

[43] J.J. Duistermaat, *Oscillatory integrals, Lagrangian immersions and unfolding of singularities*, Comm. Pure Appl. Math. **27** (1974) 207–281.

[44] E. Dumas, *Periodic multiphase nonlinear diffactive optics with curved phases*, Indiana Univ. Math. J. **52** (2003) 769–810.

[45] E. Dumas, Nonlinear diffactive optics with curved phases: Beam dispersion and transition between light and shadow, Asymptot. Anal. 38 (204) 47–91 .

[46] E. Dumas, *Global existence for Maxwell–Bloch systems*, J. Differential Equations **219** (2005) 484–509.

[47] W.E. *Propagation of oscillations in the solutions of* $1 - D$ *compressible fluid equations*, Comm. Partial Differential Equations **17** (3–4) (1992) 347–370.

[48] L.C. Evans, *Weak Convergence Methods for Non Linear Partial Differential Equations*, AMS, Providence, RI, 1990.

[49] R.P. Feynmann, *Lectures on Physics, vols. I and II*, Addison-Wesley, 1963, and 1964.

[50] K.O. Friedrichs, *Symmetric hyperbolic linear differential equations*, Comm. Pure Appl. Math. **7** (1954) 345–392.

[51] K.O. Friedrichs, *Symmetric positive linear differential equations*, Comm. Pure Appl. Math. **11** (1958) 333–418.

[52] K.O. Friedrichs, *Nonlinear hyperbolic differential equations for functions of two independent variables*, Amer. J. Math. **70** (1948) 555–589.

[53] K.O. Friedrichs, P.D. Lax, *Systems of conservation laws with a convex extension*, Proc. Nat. Acad. Sc. USA **68** (1971) 1686–1688.

[54] S. Friedlander, W. Strauss, M. Vishik, *Nonlinear instability in an ideal fluid*, Ann. Inst. Henri Poincaré **14** (2) (1997) 187–209.

[55] L. Gårding, Problèmes de Cauchy pour des systèmes quasi-linéaires d'ordre un, strictement hyperboliques, dans Les EDP, Colloques Internationaux du CNRS, 117, 1963, pp. 33–40.

[56] P. Godin, *Global oscillatory waves for second order quasilinear wave equations*, Trans. Amer. Math. Soc. **346** (2) (1994) 523–547.

[57] E. Grenier, *On the nonlinear instability of Euler and Prandtl equations*, Comm. Pure Appl. Math. **53** (9) (2000) 1067–1091.

[58] O. Guès, *Problème mixte hyperbolique quasilinéaire caractéristique*, Comm. Partial Differential Equations **15** (1990) 595–645.

[59] O. Guès, *Ondes multidimensionnelles* ε*- stratifiées et oscillations*, Duke Math. J. **68** (3) (1992) 401–446.

[60] O. Guès, *Développement asymptotique de solutions exactes de systèmes hyperboliques quasilinéaires*, Asymptot. Anal. **6** (3) (1993) 241–269.

[61] A. Haraux, *Semilinear Hyperbolic Problems in Bounded Domains*, in: Mathematical Reports, vol. 3, Part 1, Harwood Academic Publishers, 1987.

[62] A. Heibig, *Error estimates for oscillatory solutions to hyperbolic systems of conservation laws*, Comm. Partial Differential Equations **18** (1993) 281–304.

[63] E. Hewitt, K.A. Ross, *Abstract Harmonic Analysis, vol. 1*, Springer-Verlag, Berlin, 1963.

[64] L. Hörmander, *Fourier Integral Operators I*, Acta. Math. **127** (1971) 79–183.

[65] L. Hörmander, *The Analysis of Linear Partial Differential Operators*, Springer, 1990.

[66] L. Hörmander, *Lectures on Nonlinear Hyperbolic Differential Equations*, in: Mathématiques et Applications, vol. 26, Sringer Verlag, 1997.

[67] J.K. Hunter, *Transverse diffraction of nonlinear waves and singular rays*, SIAM J. Appl. Math. **48** (1988) 1–37.
[68] J.K. Hunter, J.B. Keller, *Weakly nonlinear high frequency waves*, Comm. Pure Appl. Math. **36** (1983) 547–569.
[69] J.K. Hunter, J.B. Keller, *Caustics of nonlinear waves*, Wave Motion **9** (1987) 429–443.
[70] J.K. Hunter, A.J. Majda, R. Rosales, *Resonantly interacting weakly nonlinear hyperbolic waves II: Several space variables*, Stud. Appl. Math. **75** (1986) 187–226.
[71] P.-Y. Jeanne, *Optique géométrique pour des systèmes semi-linéaires avec invariance de jauge*, Mém. Soc. Math. France (2002) 9–160.
[72] J.-L. Joly, *Sur la propagations des oscillations semi-lineares en dimension 1 d'espace*, C. R. Acad. Sc. Paris **t. 296** (1983).
[73] J.-L. Joly, G. Métivier, J. Rauch, *Transparent nonlinear geometric optics and Maxwell–Bloch equations*, J. Differential Equations **166** (2000) 175–250.
[74] J.-L. Joly, G. Métivier, J. Rauch, *Generic rigorous asymptotic expansions for weakly nonlinear geometric optics*, Duke. Math. J. **70** (1993) 373–404.
[75] J.-L. Joly, G. Métivier, J. Rauch, *Coherent nonlinear waves and the Wiener algebra*, Ann. Inst. Fourier **44** (1994) 167–196.
[76] J.-L. Joly, G. Métivier, J. Rauch, *Coherent and focusing multidimensional nonlinear geometric optics*, Ann. Sci. Ecole Norm. Sup. (4) **28** (1) (1995) 51–113.
[77] J.-L. Joly, G. Métivier, J. Rauch, *Global existence for anharmonic nonlinear Maxwell equations*, SIAM J. Math. Anal. **27** (1996) 905–913.
[78] J.-L. Joly, G. Métivier, J. Rauch, *Focusing at a point and absorption of nonlinear oscillations*, Trans. Amer. Math. Soc. **347** (1995) 3921–3971.
[79] J.-L. Joly, G. Métivier, J. Rauch, *Nonlinear oscillations beyond caustics*, Comm. Pure Appl. Math. **48** (1996) 443–529.
[80] J.-L. Joly, G. Métivier, J. Rauch, *Caustics for dissipative semilinear oscillations*, Mem. Amer. Math. Soc. (685) (2000).
[81] J.-L. Joly, G. Métivier, J. Rauch, *Diffractive geometric optics with rectification*, Indiana Math. J. **47** (1998) 1167–1241.
[82] J.-L. Joly, G. Métivier, J. Rauch, *Resonant one dimensional nonlinear geometric optics*, J. Funct. Anal. **114** (1993) 106–231.
[83] J.-L. Joly, G. Métivier, J. Rauch, *Global solutions to Maxwell equations in a ferromagnetic medium*, Ann. Henri Poincaré **1** (2000) 307–340.
[84] J.-L. Joly, G. Métivier, J. Rauch, *Hyperbolic domains of determinacy and Hamilton–Jacobi equations*, J. Hyperbolic 1 Differential Equations **2** (2005) 1–32.
[85] L.A. Kalyakin, *Long wave asymptotics integrable equations as asymptotic limit of nonlinear systems*, Russian Math. Surveys **44** (1) (1989) 3–42.
[86] S. Klainerman, A.J. Majda, *Singular perturbations of quasilinear hyperbolic systems with large parameters and the incompressible limit of compressible fluids*, Comm. Pure Appl. Math. **34** (1981) 481–524.
[87] S. Klainerman, A.J. Majda, *Compressible and incompressible fluids*, Comm. Pure Appl. Math. **35** (1982) 629–653.
[88] H.-O. Kreiss, J. Lorenz, M.J. Naughton, *Convergence of the solutions of the compressible to the solutions of the incompressible Navier–Stokes equations*, Adv. in Appl. Math. **12** (1991) 187–214.
[89] L.D. Landau, E.M. Lifshitz, *Électrodynamique des milieux continus Physique théorique (008)*, Mir, New York, 1969.
[90] D. Lannes, *Dispersive effects for nonlinear geometrical optics with rectification*, Asympt. Anal. **18** (1998) 111–116.
[91] D. Lannes, *Consistency of the KP approximation*, Discrete Contin. Dyn. Syst. (2003) 517–525.
[92] P.D. Lax, *Asymptotic solutions of oscillatory initial value problems*, Duke Math. J. **24** (1957) 627–645.
[93] P.D. Lax, *Lectures on Hyperbolic Partial Differential Equations*, Stanford, 1963.
[94] P.D. Lax, R. Phillips, *Local boundary conditions for dissipative symmetric linear differential operators*, Comm. Pure and Appl. Math. **13** (1960) 427–455.
[95] G. Lebeau, *Nonlinear optic and supercritical wave equation*, Bull. Soc. Math. Liège **70** (2001) 267–306.
[96] J. Leray, *Hyperbolic Differential Equations*, Institute for Advanced Study, 1953.
[97] V. Lescaret, *Wave transmission in dispersive media*, Math. Models Methods Appl. Sci. **17** (2007) 485–535.

[98] V. Lescaret, *Diffractive wave transmission in dispersive media*, J. Differential Equations **244** (2008) 972–1010.

[99] J.-L. Lions, W. Strauss, *Some nonlinear evolution equations*, Bull. Soc. Math. France **93** (1965) 43–96.

[100] H.A. Lorentz, *The Theory of Electrons*, 2nd edition, Teubner, 1908, Reprinted by Dover, 1952.

[101] D. Ludwig, *Uniform asymptotic expansions at a caustic*, Comm. Pure Appl. Math **19** (1966) 215–250.

[102] D. Ludwig, *Conical refraction in crystal optics and hydromagnetics*, Comm. Pure Appl. Math **14** (1961) 113–124.

[103] A.J. Majda, *Compressible Fluid Flow and Systems of Conservation Laws in Several Space Variables*, Springer-Verlag, New York, 1984.

[104] A.J. Majda, M. Artola, *Nonlinear geometric optics for hyperbolic mixed problems*, in: Analyse Mathématique et Applications, Gauthuier-Villars, Paris, 1988, pp. 319–356.

[105] A.J. Majda, R. Rosales, *Resonantly interacting weakly nonlinear hyperbolic waves I: A single space variable*, Stud. Appl. Math. **71** (1984) 149–179.

[106] S. Mizohata, *On the Cauchy Problem*, in: Notes and Reports in Mathematics in Science an Engineering, vol. 3, Academic Press, 1985.

[107] A. Newell, J. Moloney, *Nonlinear Optics*, Addison-Wesley, Reading Mass, 1992.

[108] A. Owyoung, *The Origin of the nonlinear refractive indices of liquids and glasses*, Ph.D Dissertation Caltech, 1971 .

[109] R. Pantell, H. Puthoff, *Fundamentals of Quantum Electronics*, J. Wiley & Sons Inc., New York, 1973; Electronique quantitique en vue des applications, Dunod.

[110] Y.-J. Peng, *Solutions faibles globales pour l'équation d'Euler d'un fluide compressible avec de grandes données initiales*, Comm. Partial Differential Equations **17** (1992) 161–187.

[111] *Sur les surfaces de translation et les fonctions Abeliennes*, Bull. Soc. Math. France **29** (1901) 61–86.

[112] J. Rauch, *BV estimates fail for most quasilinear hyperbolic systems in dimensions greater than one*, Comm. Math. Phys. **106** (1986) 481–484.

[113] J. Rauch, *Symmetric positive systems with boundary characteristic of constant multiplicity*, Trans. Amer. Math. Soc. **291** (1985) 167–187.

[114] J. Rauch, *Lectures on geometric optics*, IAS-Park City Summer School on Nonlinear Waves, July, 1995 .

[115] S. Schochet, *The compressible Euler equations in a bounded domain: existence of solutions and the incompressible limit*, Commun. Math. Phys **104** (1986) 49–75.

[116] S. Schochet, *Asymptotics for symmetric hyperbolic systems with a large parameter*, J. Differential Equations **75** (1988) 1–27.

[117] S. Schochet, *Fast singular limits of hyperbolic PDEs*, J. Differential Equations **114** (1994) 476–512.

[118] S. Schochet, *Resonant nonlinear geometric optics for weak solutions of conservation laws*, J. Differential Equations **113** (1994) 473–504.

[119] D. Serre, *Oscillations nonlinéaires de haute fréquence. Dim* ≥ 2, in: A. Marino, M.K.V. Murthy (Eds.), Nonlinear Variational Problems and Partial Differential Equations, in: Pitman Res. Notes in Math., vol. 320, Longman, London, 1995, pp. 245–294, Springer-Verlag, New-York, 1983.

[120] B. Sévennec, *Géométrie des systèmes hyperboliques de lois de conservation*, Société mathématique de France, Mémoire 56, Suppl. Bull. Soc. Math. France **122** (1) (1994).

[121] W. Strauss, *The energy method in nonlinear partial differential equations,* Notas de matematica 47, Instituto de Matematica Pura e Applicata, Rio de Janeiro 1969 .

[122] L. Tartar, *Solutions oscillantes des équations de Carleman,* Séminaire Goulaouic–Schwartz, École Polytechnique, Paris, année 1980-0981 .

[123] M. Taylor, *Partial Differential Equations III*, in: Applied Mathematical Sciences, vol. 117, Springer, 1996.

[124] B. Texier, *The short-wave limit for nonlinear symmetric hyperbolic systems*, Adv. Differential Equations **9** (2004) 1–52.

[125] B. Texier, *WKB asymptotics for the Euler–Maxwell equations*, Asymptot. Anal. **42** (2005) 211–250.

[126] A. Weil, *L'intégration dans les groupes topologiques et ses applications*, Hermann, Paris, 1940.

[127] G.B. Witham, *Nonlinear Waves*, in: Wiley-Intersciences Series of Texts, John Wiley and Sons, 1993.

[128] M. Williams, *Resonant reflection of multidimensional semilinear oscillations*, Comm. in Partial Diff. Equ. **18** (1993) 1901–1959.

[129] M. Williams, *Boundary layers and glancing blow-up in nonlinear geometric optics*, Ann. Sci. École Norm. Sup. **33** (2000) 383–432.

CHAPTER 4

Thermoelasticity

Reinhard Racke

Department of Mathematics and Statistics, University of Konstanz, 78457 Konstanz, Germany
E-mail: reinhard.racke@uni-konstanz.de

Contents

Abstract

This work presents an extensive discussion of initial boundary value problems in thermoelasticity. First, the classical hyperbolic–parabolic model with Fourier's law of heat conduction is considered giving results on linear systems, nonlinear systems, in one or more space dimensions, in particular discussing the asymptotic behavior of solutions as time tends to infinity. Second, recent developments for hyperbolic models using Cattaneo's law for heat conduction are described, including the comparison to classical thermoelasticity.

Keywords: Thermoelasticity, initial boundary value problem, Fourier law, Cattaneo law, decay rate

HANDBOOK OF DIFFERENTIAL EQUATIONS
Evolutionary Equations, Volume 5
Edited by C.M. Dafermos and M. Pokorný

DOI: 10.1016/S1874-5717(08)00211-9

1. Introduction

The equations of thermoelasticity describe the behavior of an elastic heat conducting body. In the classical model the hyperbolic system of elasticity is combined with the classical parabolic model of heat conduction. This leads to a hyperbolic–parabolic coupled system that is interesting because of its importance in mechanics, physics and engineering, but also, and not least, because of its mathematical features: as a hyperbolic–parabolic system it shares properties that are typical for hyperbolic and parabolic systems, respectively. The main question is which feature is predominant. It turns out that the answer depends on the space dimension and also on the type of question raised. This will be demonstrated in an extensive discussion of the following aspects: linear systems, nonlinear systems, one-dimensional models, higher-dimensional models, local and global well-posedness in different function classes (strong, weak solutions), asymptotic behavior as time tends to infinity (exponential or polynomial decay to equilibria). This discussion is mainly based on our book with Jiang [47]. We also remark that as a hyperbolic–parabolic system, a small part already appeared in the article by Hsiao and Jiang [33] in volume I of this series.

The common modeling of heat conduction using the Fourier law, essentially saying that the heat flux is a certain function of the gradient of the temperature, leads to the well-known paradox of infinite propagation of signals, in particular of heat signals. To overcome this drawback, different models of hyperbolic heat conduction have been developed, one being the Cattaneo[1] law replacing the Fourier law. We report on the development in recent years in this field – well-posedness, asymptotics of solutions – where the damped hyperbolic system of thermoelasticity arising from Cattaneo's law is used. This leads to the *second sound* effect, meaning the wave propagation of heat in this model. A comparison of this system with the classical one above will show that many features are the same both quantitatively and qualitatively, while this might be not true for related hyperbolic-hyperbolic/-parabolic thermoelastic systems like Timoshenko-type ones.

The material in Section 2 is mainly taken from [47], while the rest of this Chapter has previously only been published in original papers or has not yet been published at all. The list of references is long but does not claim to be exhaustive, cf. [47] for further references on different topics.

The paper is organized as follows: in Section 2 we present results for the system of classical hyperbolic–parabolic thermoelasticity. Section 2.1 gives the basic setting for the linearized system together with first descriptions of the asymptotic behavior as time tends to infinity. In Section 2.2 this long-term behavior is investigated in detail for one space dimension, while Section 2.3 provides the same for dimensions two or three. Section 2.4 presents the discussion of nonlinear systems giving global well-posedness results for smooth or weak solutions and blow-up results, respectively, first in one space dimension and then in higher dimensions.

Section 3 is devoted to the discussion of thermoelasticity, where heat conduction is governed by Cattaneo's law, leading to a fully hyperbolic system. In Section 3.1, linearized models are studied in particular with respect to the time asymptotic behavior, together with a comparison to Section 2 for real materials. Section 3.2 presents recent results for

[1] It is also referred to as Maxwell's law, or Cattaneo-Vernotte law, or Lord-Shulman model, see [8,7,66].

nonlinear systems. Finally, remarks on other hyperbolic thermoelastic models are given in Section 3.3.

We use standard notations, e.g. see [1] for Lebesgue spaces $L^p(\Omega)$ and Sobolev spaces $H^{m,p}(\Omega)$ with norms $\|\cdot\|_p$ and $\|\cdot\|_{m,p}$, respectively.

2. Classical thermoelasticity

In this second section we consider initial-boundary value problems for the system of classical thermoelasticity, where the term *classical* refers to the most common system where a coupling of a hyperbolic system from elasticity with a parabolic equation for the heat conduction is given; the latter arises from Fourier's law of heat conduction and leads to a formally infinite propagation of signals. In contrast to this system we shall look at different models for heat conduction in later sections. In particular, in Section 3 we shall study, via Cattaneo's law replacing Fourier's law, a fully hyperbolic model of thermoelasticity with finite propagation also, of heat signals in this model.

The aim of this section is to give a survey of results of existence, both for linearized models and for the fully nonlinear basic system, together with a study of the asymptotics of solutions as time evolves. This is done in a variety of settings — domains, bounded or unbounded, systems nonlinear or linearized, global existence or blow-up of smooth solutions versus global existence of weak solutions and different boundary conditions. By the type of the coupling, topics from various fields of mathematics are involved (hyperbolic, parabolic, elliptic partial differential equations, dynamical systems) making the system interesting, both from an applied point of view looking at a specific system, and within a mathematical framework of partial differential equations and functional analysis. The presentation in Section 2 is essentially based on [47].

The basic unknowns to describe the elastic and the thermal behavior of a body with undistorted reference configuration Ω, Ω being a domain in $\mathbb{R}^n$, $n = 1, 2$ or 3, are the displacement (vector) $U = U(t,x) := X(t,x) - x$, where $X(t,x)$ denotes the position of the reference point x at time t, and the temperature difference $\theta(t,x) := T(t,x) - T_0$, where T denotes the absolute temperature and T_0 is a chosen fixed reference temperature. A deformation goes along with a change of temperature causing the coupling of heat conduction and of elasticity. The two basic nonlinear differential equations arise from the balance of linear momentum and the balance of energy. The former reads as

$$\rho U_{tt} - \nabla' \tilde{S} = \rho b, \tag{2.1}$$

where ρ is the material density in Ω, $\tilde{S}$ is the Piola–Kirchhoff stress tensor, and b is the specific external body force, while B' denotes the transposition of a matrix B; in particular, with ∇ denoting the gradient operator, ∇' denotes the divergence operator. This system of equations essentially describes the elastic part; actually, if $\tilde{S}$ does not depend on the temperature, it represents the (hyperbolic) partial differential equations in pure elasticity. The balance of energy is given by

$$\varepsilon_t - \operatorname{tr}\{\tilde{S} F_t\} + \nabla' q = r, \tag{2.2}$$

where ε is the internal energy, q is the heat flux, r is the external heat supply, $\mathrm{tr}B$ denotes the trace of a matrix B, and F is the deformation gradient,

$$F = 1 + \nabla U.$$

By η we denote the entropy and by

$$\psi := \varepsilon - (\theta + T_0)\eta$$

the Helmholtz free energy.

The constitutive assumptions in *classical* thermoelasticity are that $\tilde{S}$, q, ψ and η are functions of the present values of $\nabla U, \theta$ and $\nabla\theta$ (and x); cp. Section 3 for different assumptions in thermoelasticity with *second sound*.

It is always assumed that these functions are smooth and that

$$\det F \neq 0, \quad T > 0.$$

With the help of the second law of thermodynamics it turns out that

$$\psi = \psi(\nabla U, \theta), \qquad \tilde{S} = \tilde{S}(\nabla U, \theta) = \frac{\partial \psi}{\partial(\nabla U)}(\nabla U, \theta),$$

$$\eta(\nabla U, \theta) = -\frac{\partial \psi}{\partial \theta}(\nabla U, \theta),$$

$$q(\nabla U, \theta, \nabla\theta)\nabla\theta \leq 0.$$

Using these relations we rewrite (2.2) as

$$(\theta + T_0)\eta_t + \nabla' q = r, \tag{2.3}$$

or

$$(\theta + T_0)\left\{-\frac{\partial^2\psi}{\partial\theta^2}\theta_t - \frac{\partial^2\psi}{\partial(\nabla U)\partial\theta}\nabla U_t\right\} + \nabla' q = r. \tag{2.4}$$

Equation (2.1) is mainly a hyperbolic system for U; Eq. (2.4) is mainly a parabolic equation for θ.

The problem of finding U and θ will become well-posed if additionally *initial conditions*

$$U(t=0) = U^0, \qquad U_t(t=0) = U^1, \qquad \theta(t=0) = \theta^0, \tag{2.5}$$

and, if $\Omega \neq \mathbb{R}^n$, *boundary conditions*, are prescribed, for example “rigidly clamped, constant temperature”,

$$U = 0, \quad \theta = 0 \quad \text{on } \partial\Omega, \tag{2.6}$$

or "traction free, insulated",

$$\tilde{S}\,\nu = 0, \quad \nu' q = 0, \tag{2.7}$$

or other combinations of the boundary conditions for U and θ. Here $\nu = \nu(x)$ denotes the exterior normal in $x \in \partial\Omega$, $\partial\Omega$ being the boundary of Ω.

The investigation of the *linearized* equations will play an important role. The linearized equations arise from (2.1), (2.4) by assuming that

$$|\nabla U|, |\nabla U_t|, |\theta|, |\theta_t|, |\nabla\theta|$$

are small. We arrive at ($T_0 = 1$ without loss of generality)

$$\rho U_{tt} - \mathcal{D}' S \mathcal{D} U + \mathcal{D}' \Gamma \theta = \rho b, \tag{2.8}$$

$$\tilde{\delta}\theta_t - \nabla' K \nabla\theta + \Gamma' \mathcal{D} U_t = r, \tag{2.9}$$

where $\rho = \rho(x)$ can be regarded as a symmetric density matrix, $S = S(x)$ is an $M \times M$ symmetric, positive definite matrix containing the elastic moduli ($M = 6$ in $\mathbb{R}^3$), $\Gamma = \Gamma(x)$ is a vector with coefficients determining the so-called stress–temperature tensor, $\tilde{\delta} = \tilde{\delta}(x)$ is the specific heat and $K = K(x)$ is the heat conductivity tensor. All functions are assumed to be sufficiently smooth. $\mathcal{D}$ is an abbreviation for a gradient,

$$\mathcal{D} := \begin{pmatrix} \partial_1 & 0 & 0 \\ 0 & \partial_2 & 0 \\ 0 & 0 & \partial_3 \\ 0 & \partial_3 & \partial_2 \\ \partial_3 & 0 & \partial_1 \\ \partial_2 & \partial_1 & 0 \end{pmatrix} \text{ in } \mathbb{R}^3, \quad \mathcal{D} := \begin{pmatrix} \partial_1 & 0 \\ 0 & \partial_2 \\ \partial_2 & \partial_1 \end{pmatrix} \text{ in } \mathbb{R}^2, \quad \mathcal{D} := \partial_1 \text{ in } \mathbb{R}^1.$$

In this way the general (linear) non-homogeneous, anisotropic case is described. The linear counterpart of the boundary conditions (2.7) reads

$$\mathcal{N}' \left(S\mathcal{D}U - \Gamma\theta\right) = 0, \quad \nu' K \nabla\theta = 0,$$

where $\mathcal{N}$ arises from the normal vector ν in the same way as $\mathcal{D}$ arises from the gradient vector ∇.

The elastic moduli C_{ijkl}, $i, j, k, l = 1 \ldots, n$, which are given in general by

$$C_{ijkl} = \frac{\partial^2 \psi(0, 0, x)}{\partial(\partial_j U_i)\partial(\partial_k U_l)},$$

satisfy in the linear case

$$C_{ijkl} = C_{klij} = C_{jikl} = C_{jikl}(x).$$

The assumption of positive definiteness of $(C_{ijkl})_{ijkl}$, in the sense that

$$\exists\, k_0 > 0\ \forall\, \xi_{ij} = \xi_{ji} \in \mathbb{R} \quad \forall x \in \Omega: \quad \xi_{ij} C_{ijkl} \xi_{kl} \geq k_0 \sum_{j,k=1}^{n} |\xi_{jk}|^2,$$

where the Einstein summation convention is used, implies that the matrix $S = S(x)$ is uniformly positive definite, since

$$S = \begin{pmatrix} C_{1111} & C_{1122} & C_{1133} & C_{1123} & C_{1131} & C_{1112} \\ \cdot & C_{2222} & C_{2233} & C_{2223} & C_{2231} & C_{2212} \\ \cdot & \cdot & C_{3333} & C_{3323} & C_{3331} & C_{3312} \\ \cdot & \cdot & \cdot & C_{2323} & C_{2331} & C_{2312} \\ \cdot & \cdot & \cdot & \cdot & C_{3131} & C_{3112} \\ \cdot & \cdot & \cdot & \cdot & \cdot & C_{1212} \end{pmatrix} \quad \text{in } \mathbb{R}^3,$$

$$S = \begin{pmatrix} C_{1111} & C_{1122} & C_{1112} \\ \cdot & C_{2222} & C_{2212} \\ \cdot & \cdot & C_{1212} \end{pmatrix} \quad \text{in } \mathbb{R}^2, \qquad S = C_{1111} \quad \text{in } \mathbb{R}^1.$$

In the simplest case of a homogeneous medium which is isotropic we have

$$S = \begin{pmatrix} 2\mu+\lambda & \lambda & \lambda & 0 & 0 & 0 \\ \cdot & 2\mu+\lambda & \lambda & 0 & 0 & 0 \\ \cdot & \cdot & 2\mu+\lambda & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \mu & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \mu & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \mu \end{pmatrix} \quad \text{in } \mathbb{R}^3,$$

$$S = \begin{pmatrix} 2\mu+\lambda & \lambda & 0 \\ \cdot & 2\mu+\lambda & 0 \\ \cdot & \cdot & \mu \end{pmatrix} \quad \text{in } \mathbb{R}^2, \qquad S = 2\mu+\lambda =: \alpha \quad \text{in } \mathbb{R}^1,$$

and the equations reduce in two or three space dimensions to

$$U_{tt} - ((2\mu+\lambda)\nabla\nabla' - \mu\nabla\times\nabla\times)U + \tilde{\gamma}\nabla\theta = b, \tag{2.10}$$

$$\tilde{\delta}\theta_t - \kappa\Delta\theta + \tilde{\gamma}\nabla' U_t = r, \tag{2.11}$$

where the density $\rho = 1$, without loss of generality, and $\mu, \lambda, \tilde{\gamma}, \tilde{\delta}$ and κ are constants; μ, λ are the Lamé moduli,

$$\mu > 0, \quad 2\mu + n\lambda > 0,$$

moreover,

$$\tilde{\delta}, \kappa > 0, \quad \tilde{\gamma} \neq 0.$$

In one space dimension, the basic equations for the homogeneous (and necessarily isotropic) case are:

$$U_{tt} - \alpha U_{xx} + \tilde{\gamma}\theta_x = b, \tag{2.12}$$

$$\tilde{\delta}\theta_t - \kappa\theta_{xx} + \tilde{\gamma}U_{tx} = r, \tag{2.13}$$

and we shall often write in this case u instead of U. In the following subsections we shall present results on the well-posedness of the linearized and of the nonlinear systems as well as a description of the time-asymptotic behavior in both bounded and unbounded domains.

Notes: For an extended derivation of the equations see Carlson [4]; for the representation (2.8) and (2.9) see Leis [61].

2.1. *Linear systems — well-posedness*

We discuss the well-posedness of the linear initial boundary value problem (2.5), (2.6), (2.8) and (2.9), i.e., the linearized differential equations

$$\rho U_{tt} - \mathcal{D}' S \mathcal{D} U + \mathcal{D}' \Gamma \theta = \rho b, \tag{2.14}$$

$$\tilde{\delta}\theta_t - \nabla' K \nabla\theta + \Gamma' \mathcal{D} U_t = r, \tag{2.15}$$

together with initial conditions

$$U(t=0) = U^0, \quad U_t(t=0) = U^1, \quad \theta(t=0) = \theta^0, \tag{2.16}$$

and Dirichlet–Dirichlet boundary conditions (if $\Omega \neq \mathbb{R}^n$)

$$U = 0, \quad \theta = 0 \quad \text{on } \partial\Omega. \tag{2.17}$$

Moreover, we shall present the first results of the asymptotic behavior of the solutions as time tends to infinity. We consider the initial boundary value problem (2.14)–(2.17) for $U = U(t,x) \in \mathbb{R}^n, \theta = \theta(t,x) \in \mathbb{R}, t \geq 0, x \in \Omega \subset \mathbb{R}^n$, where Ω has a smooth boundary $\partial\Omega$ — the smoothness only being necessary for later higher regularity of the solution. The matrices $S = S(x)$, $K = K(x)$, and $\rho = \rho(x)$ are assumed to be positive definite, uniformly in x, $\tilde{\delta} = \tilde{\delta}(x)$ and the norm $|\Gamma|$ of the vector $\Gamma = \Gamma(x)$ are assumed to be positive, uniformly in x, too.

The domain Ω will be either bounded or exterior, i.e., a domain with a bounded complement; $\Omega = \mathbb{R}^n$ is then allowed with void condition (2.17), and in one space dimension we also consider the half space $\Omega = (0, \infty)$.

Solutions in appropriate Sobolev spaces will be obtained using semigroup theory. For this purpose we transform Eqs. (2.14) and (2.15) into an evolution equation of first order in time. Let

$$V := \begin{pmatrix} S\mathcal{D}U \\ U_t \\ \theta \end{pmatrix} \equiv \begin{pmatrix} V^1 \\ V^2 \\ V^3 \end{pmatrix}.$$

Using the definitions

$$Q := \begin{pmatrix} S^{-1} & 0 & 0 \\ 0 & \rho & 0 \\ 0 & 0 & \tilde{\delta} \end{pmatrix} \qquad N := \begin{pmatrix} 0 & -\mathcal{D} & 0 \\ -\mathcal{D}' & 0 & \mathcal{D}'\Gamma \\ 0 & \Gamma'\mathcal{D} & -\nabla' K \nabla \end{pmatrix},$$

V satisfies the differential equation

$$V_t + Q^{-1}NV = F := \begin{pmatrix} 0 \\ b \\ r/\tilde{\delta} \end{pmatrix} \tag{2.18}$$

with initial conditions

$$V^0 := V(t=0) = \begin{pmatrix} S\mathcal{D}U^0 \\ U^1 \\ \theta^0 \end{pmatrix}. \tag{2.19}$$

Let

$$\mathcal{H} := L^2(\Omega)$$

be the underlying Hilbert space with inner product

$$\langle W, Z \rangle_{\mathcal{H}} := \langle W, QZ \rangle,$$

where $\langle \cdot, \cdot \rangle$ will denote the inner product in $L^2(\Omega)$ with corresponding norm $\| \cdot \|$.

The norm

$$E(t) := \|V(t)\|^2_{\mathcal{H}} \equiv \langle V(t), V(t) \rangle_{\mathcal{H}}$$

corresponds to the "energy" of (U, θ), since

$$E(t) = \int_\Omega S(x)\mathcal{D}U(t,x) \cdot \mathcal{D}U(t,x) + \rho(x)U_t(t,x) \cdot U_t(t,x) + \tilde{\delta}(x)|\theta(t,x)|^2 \mathrm{dx}.$$

We call $E = E(t)$ the "first-order energy" of (U, θ) or of V.

The initial-boundary value problem (2.14)–(2.17) is now reformulated in terms of V with the help of the following differential operator A with domain $D(A)$ in $\mathcal{H}$.

Let

$$\begin{aligned}
&A : D(A) \subset \mathcal{H} \longrightarrow \mathcal{H},\\
&D(A) := \{V \in \mathcal{H} \mid V^2 \in H_0^1(\Omega), V^3 \in H_0^1(\Omega), NV \in \mathcal{H}\}\\
&\qquad = \{V \in \mathcal{H} \mid V \in D'(\Omega) \times H_0^1(\Omega) \times H_0^1(\Omega), \nabla' K \nabla V^3 \in L^2(\Omega)\},\\
&AV := Q^{-1} NV,
\end{aligned}$$

where $H_0^1(\Omega)$ $(= W_0^{1,2}(\Omega))$ denotes the usual Sobolev space generalizing the Dirichlet boundary condition (cf. [1]) and

$$D'(\Omega) := \{W \in L^2(\Omega) \mid \mathcal{D}'\mathcal{W} \in \mathcal{L}^{\in}(\otimes)\}.$$

Then V should satisfy (cf. (2.18) and (2.19))

$$V_t(t) + AV(t) = F(t), \qquad V(t = 0) = V^0, \tag{2.20}$$

and we look for a solution

$$V \in C^1([0, \infty), \mathcal{H}) \cap C^0([0, \infty), D(A)).$$

Using the Hille–Yosida characterization it is not difficult to show that $-A$ generates a contraction semigroup, in particular $D(A)$ is a Hilbert space with inner product

$$\langle \cdot, \cdot \rangle_A := \langle \cdot, \cdot \rangle_{\mathcal{H}} + \langle A\cdot, A\cdot \rangle_{\mathcal{H}}.$$

Here the following expression is used

$$D^0(\Omega) = H_0^1(\Omega), \tag{2.21}$$

where

$$D^0(\Omega) := \{U \in L^2(\Omega) \mid \exists \mathcal{D}V^2 \text{ and } \forall \psi \in D'(\Omega) : \langle U, \mathcal{D}'\psi \rangle = -\langle \mathcal{D}V^2, \psi \rangle\}.$$

Thus we have

THEOREM 2.1. *The operator $-A$ is the generator of a contraction semigroup $\{T(t)\}_{t \geq 0}$, $T(t) = \mathrm{e}^{-tA}$, and for $F \in C^1((0, \infty), \mathcal{H})$ and $V^0 \in D(A)$ there exists a unique solution*

$$V \in C^1([0, \infty), \mathcal{H}) \cap C^0([0, \infty), D(A))$$

to (2.20). V can be represented as

$$V(t) = T(t)V^0 + \int_0^t T(t-s)F(s)\mathrm{d}s.$$

The transformation to a first-order system in time has produced an artificial null space. The asymptotic behavior of the solution is nontrivial only on the orthogonal complement of the null space $N(A) = N(A^*)$, which reduces A.

Let

$$D_0'(\Omega) := \{W \in L^2(\Omega) \mid \mathcal{D}'W = 0\}.$$

LEMMA 2.2. (i) $N(A) = D_0'(\Omega) \times \{0\} \times \{0\}$.

(ii) $N(A)^\perp = \overline{S\mathcal{D}H_0^1(\Omega)} \times L^2(\Omega) \times L^2(\Omega)$.

PROOF. (i) easily follows using

$$\|\mathcal{D}V^2\|^2 \geq \frac{1}{2}\|\nabla V^2\|^2 \tag{2.22}$$

for $V^2 \in H_0^1(\Omega)$. Then (ii) follows from the projection theorem

$$L^2(\Omega) = \overline{\mathcal{D}H_0^1(\Omega)} \oplus D_0'(\Omega). \qquad \square$$

For bounded domains we conclude from the Korn-type inequality (2.22) and the Poincaré inequality in $H_0^1(\Omega)$, that

$$\overline{\mathcal{D}H_0^1(\Omega)} = \mathcal{D}H_0^1(\Omega). \tag{2.23}$$

In the sequel we shall denote by A the given operator A reduced to the $N(A)^\perp$ described in Lemma 2.2(ii), and in the following we denote $N(A)^\perp$ by $\mathcal{H}$.

THEOREM 2.3. (i) *A^{-1} is compact in a bounded domain.*

(ii) *A^{-1} is "locally compact" in an exterior domain, i.e., every sequence which is bounded in the graph norm $\|\cdot\|_A$ has a subsequence that converges in $L^2(\Omega_R)$ for every $R > 0$, where*

$$\Omega_R := \Omega \cap B(0, R).$$

PROOF. (i): Claim:

$$\exists d > 0 \quad \forall V \in D(A) : \|V\| \leq d\|AV\|. \tag{2.24}$$

Proof of (2.24), by contradiction. Assume

$$\exists (V_n)_n \subset D(A) \quad \forall n : \|V_n\| = 1 \quad \text{and} \quad \|AV_n\| < \frac{1}{n}.$$

Using (2.22), this implies the existence of a subsequence of $(V_n^2)_n$ that converges in $L^2(\Omega)$ by Rellich's theorem. Since

$$\|\nabla V_n^3\|^2 \leq C(\|V_n^3\|^2 + \|\nabla' K \nabla V_n^3\|^2),$$

with a constant $C > 0$, also $(V_n^3)_n$ has an $L^2(\Omega)$-convergent subsequence. We know from (2.23) that

$$V_n^1 = S\mathcal{D}W_n$$

for some $W_n \in H_0^1(\Omega)$ implying

$$\|\nabla W_n\| \leq C,$$

hence $(W_n)_n$ has an $L^2(\Omega)$-convergent subsequence. Denoting subsequences by the same symbol, we conclude

$$\|V_n^1 - V_m^1\|^2 \leq C\|W_n - W_m\| \, \|\mathcal{D}'(V_n^1 - V_m^1)\| \longrightarrow 0 \quad \text{as } n, m \to \infty.$$

Altogether we obtain

$$\exists V \in \mathcal{H}: \quad V_n \xrightarrow{\mathcal{H}} V \quad \text{as } n \to \infty.$$

Since $AV_n \to 0$ and A is closed we get $V = 0$, which is a contradiction to $\|V_n\| = 1$ for $n \in \mathbb{N}$. This proves (2.24).

Now let $(F_n = AV_n)_n$ be bounded. By (2.24) we conclude the boundedness of $(V_n)_n$ (and $(AV_n)_n$). Thus, the $L^2(\Omega)$-convergence of a subsequence of $(V_n)_n$ follows by arguments as in the proof of (2.24). This proves (i).

(ii): As in the proof of (i), the boundedness of $(V_n^2)_n$ and (V_n^3) in $H_0^1(\Omega)$ implies the convergence of subsequences in $L^2(\Omega_R)$ for any $R > 0$.

In order to conclude the convergence of a subsequence of $(V_n^1)_n$ we need the following lemma, which will be proved below.

LEMMA 2.4. *Let the domain G be bounded and have the strict cone property. Then we have*

$$\exists k > 0 \quad \forall W \in H^1(G) = W^{1,2}(G) : \|W\| \leq k \sup_{\substack{V \in D_0(G) \\ \|V\|=1}} \{\|\mathcal{D}W\| + |\langle W, V\rangle|\},$$

where

$$D_0(G) := \{W \in L^2(G) \mid \mathcal{D}W = 0\}.$$

By virtue of (2.23),

$$V_n^1 = L^2(\Omega)\text{–}\lim_{j \to \infty} S\mathcal{D}W_{n,j}$$

for some $(W_{n,j})_j \subset H_0^1(\Omega)$.

Lemma 2.4 is applicable in Ω_R since $W_{n,j|\partial\Omega} = 0$ and $\partial B(0, R)$ has the strict cone property. Without loss of generality,

$$W_{n,j} \in D_0(\Omega_R)^{\perp}.$$

Thus, $(W_{n,j})_j$ converges in $L^2(\Omega_R)$, as well as $(S\mathcal{D}W_{n,j})_j$.

Korn's inequality in $H^1(G)$ for domains G with the strict cone property (cf. [61]), saying that

$$\exists p > 0 \quad \forall V \in H^1(G) : \|\mathcal{D}V\|^2 + \|V\|^2 \geq p\|V\|_{1,2}^2, \tag{2.25}$$

where $\|\cdot\|_{1,2}$ denotes the norm in $H^1(G)$, yields $W_n \in H^1(\Omega_R)$, such that $W_{n,j} \to W_n$ in $H^1(\Omega_R)$ and $W_n \in D_0(\Omega_R)^{\perp}$. Hence

$$V_n^1 = S\mathcal{D}W_n,$$

and, applying Lemma 2.4 and (2.25), we observe that $(W_n)_n$ is uniformly bounded in $H^1(\Omega_R)$, implying the convergence of a subsequence in $L^2(\Omega_R)$.

Let $\psi \in C^\infty(\mathbb{R}^n)$, $0 \leq \psi \leq 1$, $\psi_{|\Omega_{R-1}} = 1$, $\psi_{|\mathbb{R}^n \setminus \Omega_R} = 0$. Then we get

$$\begin{aligned}
\|V_n^1 - V_m^1\|_{L^2(\Omega_{R-1})}^2 &= \|S\mathcal{D}\psi(W_n - W_m)\|_{L^2(\Omega_{R-1})}^2 \\
&\leq C\langle \mathcal{D}\psi(W_n - W_m), S\mathcal{D}\psi(W_n - W_m)\rangle_{L^2(\Omega_R)} \\
&\leq C\|W_n - W_m\|_{L^2(\Omega_R)}\|\psi\mathcal{D}'S\mathcal{D}\psi(W_n - W_m)\|_{L^2(\Omega_R)} \\
&\leq C\|W_n - W_m\|_{L^2(\Omega_R)} \to 0 \quad \text{as } n \to \infty. \qquad \square
\end{aligned}$$

PROOF OF LEMMA 2.4. Suppose the existence of $(W_n)_n \subset H^1(G)$ such that

$$\|W_n\| = 1 \quad \text{and} \quad \sup_{\substack{V \in D_0 \\ \|V\|=1}} \{\|DW_n\| + |\langle W_n, V\rangle\} < 1/n.$$

Then

$$\|\mathcal{D}W_n\| \to 0 \quad \text{as } n \to \infty$$

and (2.25) implies the boundedness of $(W_n)_n$ in $H^1(G)$, therefore the convergence of a subsequence of $(W_n)_n$ in $L^2(G)$ by Rellich's theorem. Hence by virtue of (2.25) there exists a subsequence of $(W_n)_n$, still denoted by $(W_n)_n$, and a $W \in H^1(G)$ such that

$$\|W_n - W\|_{1,2} \to 0, \qquad W \in D_0, \qquad \|W\| = 1,$$

which is a contradiction since

$$\|W\|^2 = \lim_n \langle W_n, W\rangle = 0. \qquad \square$$

As a consequence of the compactness of A^{-1} in bounded domains we obtain

THEOREM 2.5. *If Ω is bounded, the spectrum $\sigma(A)$ of A consists of isolated eigenvalues with finite multiplicity.*

For the case $\Omega = \mathbb{R}^n$ and homogeneous isotropic media we can describe the spectrum in detail. For bounded domains we look for the existence of purely imaginary eigenvalues of A, since they describe pure oscillations for the time-dependent problem, excluding a damping through heat conduction.

Let $AV = \lambda V$, $\lambda \in \mathbb{C}$, then

$$\mathrm{Re}\,\lambda\|V\|_{\mathcal{H}}^2 = \langle\nabla V^3, K\nabla V^3\rangle, \qquad \mathrm{Im}\,\lambda\|V\|_{\mathcal{H}}^2 = 2\mathrm{Im}\,\langle\mathcal{D}'(\Gamma V^3 - V^1), V^2\rangle.$$

$\lambda = \mathrm{i}b$ with $b \in \mathbb{R} \setminus \{0\}$ leads to

$$V^3 = 0, \qquad -S\mathcal{D}V^2 = \mathrm{i}bV^1, \qquad -\rho^{-1}\mathcal{D}'V^1 = \mathrm{i}bV^2, \qquad \tilde{\delta}^{-1}\Gamma'\mathcal{D}V^2 = 0,$$

hence

$$-\rho^{-1}\mathcal{D}'S\mathcal{D}V^2 = b^2V^2, \qquad V^2 \in H_0^1(\Omega), \tag{2.26}$$

$$\Gamma'\mathcal{D}V^2 = 0, \tag{2.27}$$

and we immediately have

THEOREM 2.6. *There are purely imaginary eigenvalues if and only if the eigenvalue problem (2.26) with side condition (2.27) has solutions.*

Since $V^3 = 0$ for eigenvectors $V = (V^1, V^2, V^3)'$ to purely imaginary eigenvalues, the following lemma is obvious.

LEMMA 2.7. (i) *$\lambda \in \mathrm{i}\mathbb{R}$ is an eigenvalue of $A \Leftrightarrow \overline{\lambda} \in \mathrm{i}\mathbb{R}$ is an eigenvalue of A^*.*
(ii) *$AV_j = \lambda_j V_j$, $\lambda_j \in \mathrm{i}\mathbb{R}$, $j = 1, 2$, $\lambda_1 \neq \lambda_2 \Rightarrow \langle V_1, V_2\rangle_{\mathcal{H}} = 0$.*

The question of existence of purely imaginary eigenvalues is easy to answer in one space dimension.

THEOREM 2.8. *There are no purely imaginary eigenvalues if $\Omega \subset \mathbb{R}^1$.*

PROOF. The side condition (2.27) turns into

$$\Gamma(x)\frac{\mathrm{d}}{\mathrm{dx}}V^2(x) = 0.$$

This implies $V^2 = 0$, since the function Γ is pointwise different from zero. □

In more than one space dimension the situation becomes more complicated, with consequences for the time asymptotic behavior. If we consider a homogeneous, isotropic medium in $\mathbb{R}^2$,

$$\rho = \mathrm{Id}, \qquad \tilde{\delta} = 1, \qquad K = \mathrm{Id}, \qquad \Gamma = \tilde{\gamma}\begin{pmatrix}1\\1\\0\end{pmatrix}, \qquad \tilde{\gamma} \neq 0,$$

and (2.26), (2.27) become

$$\Delta V^2 + \mu^{-1} b^2 V^2 = 0, \qquad V^2 \in H_0^1(\Omega), \tag{2.28}$$

$$\nabla' V^2 = 0, \tag{2.29}$$

1. and if $\Omega = B(0,1)$, the unit ball in $\mathbb{R}^2$, there are solutions to (2.28), (2.29). All solutions are given by, in polar coordinates (r, ψ),

   $$V_n^2(r,\varphi) = \begin{pmatrix}-J_1(\chi_n r)\sin\varphi\\ J_1(\chi_n r)\cos\varphi\end{pmatrix}, \quad n \in \mathbb{N},$$

   where J_1 is the Bessel function of first kind, χ_n equals the nth zero of J_1, and $b^2 = b_n^2 = \mu\chi_n^2$.
2. If Ω is a rectangle, there are no solutions to (2.28), (2.29).
3. A *sufficient* condition for the non-existence of solutions to (2.28), (2.29) is that the *scalar* eigenvalue problem

   $$-\Delta v + \lambda v = 0, \quad v_{|\partial\Omega} = 0, \quad \Omega \subset \mathbb{R}^n, \tag{2.30}$$

   has only distinct eigenvalues, i.e., all eigenvalues are simple (cf. [12]).
   - Domains $\Omega \subset \mathbb{R}^2$, for which $\partial\Omega$ is smooth and, additionally, a straight line in a neighborhood of a point, are examples, where (2.30) has only simple eigenvalues (cf. [13]).
   - In every C^3-neighborhood of a domain Ω, there is a domain Ω^*, such that (2.30) has only simple eigenvalues. With respect to an appropriate C^3-metric, the set of domains Ω leading to multiple eigenvalues in (2.30) is of the first Baire category (cf. [68–70,117]).

This complicated picture of the existence of purely imaginary eigenvalues of A is not only of interest for itself in the stationary case but will have its implications for the asymptotic behavior for solutions to the time-dependent problem.

For homogeneous, isotropic media and $\Omega = \mathbb{R}^n$, i.e., no boundary, we can use the Fourier transform to describe the spectrum explicitly.

Let, without loss of generality,

$$\rho = \mathrm{Id}, \qquad \tilde{\delta} = 1, \qquad K = \kappa\mathrm{Id},$$
$$\mathcal{D}'\Gamma = \tilde{\gamma}\nabla, \qquad \Gamma\mathcal{D} = \tilde{\gamma}\nabla', \qquad \tilde{\gamma} \neq 0,$$

and S as described above.

Denoting A, before reducing it to $N(A)^{\perp}$, by A_0 for the moment and letting $\hat{A}_0(p)$ denote the symbol of A after Fourier transform, we have

$$\sigma(A_0) = \{\beta \in \mathbb{C} \mid \exists p \in \mathbb{R}^n : \det(\hat{A}_0(p) - \beta) = 0\}.$$

In $\underline{\mathbb{R}^1}$ we have

$$\begin{aligned}\det(\hat{A}_0(p) - \beta) &= \begin{vmatrix} -\beta & -\alpha \mathrm{i}p & 0 \\ -\mathrm{i}p & -\beta & \tilde{\gamma}\mathrm{i}p \\ 0 & \tilde{\gamma}\mathrm{i}p & \kappa|p|^2 - \beta \end{vmatrix} \\ &= -\beta^3 + \kappa|p|^2\beta^2 - (\alpha + \tilde{\gamma}^2)|p|^2\beta + \kappa\alpha|p|^4 \\ &=: \Delta(\beta, p, \alpha),\end{aligned}$$

to be read as a definition for any $\beta \in \mathbb{C}$, $p \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$.

REMARK 2.9. The case $\tilde{\gamma} = 0$, i.e., a decoupling of heat conduction and elasticity, leads to

$$\Delta(\beta, p, \alpha) = -(\beta - \kappa|p|^2)(\beta^2 + \alpha|p|^2),$$

i.e., we have three branches $\beta_j(p)$, $j = 1, 2, 3$,

$$\beta_1(p) = \kappa|p|^2$$

from pure heat conduction, and

$$\beta_{2,3}(p) = \pm\mathrm{i}\sqrt{\alpha}|p|,$$

from pure elasticity, after the transformation to a first-order system.

Now let $\tilde{\gamma} \neq 0$ again. The zeros $\beta_j(p)$ of $\Delta(\beta, p, \alpha)$, $j = 1, 2, 3$, can be computed explicitly and we obtain that $\{\beta_1(p), p \in \mathbb{R}\}$ equals the nonnegative real axis ("heat conduction"), and $\{\beta_j(p), p \in \mathbb{R}\}$ for $j = 2, 3$ consists of two parts which are symmetric to the real axis and for which the real part is bounded by some $\alpha_0 = \alpha_0(\tilde{\gamma}^2, \alpha)$.

For the qualitative behavior in the case

$$\frac{\tilde{\gamma}^2}{\alpha} < 1 \tag{2.31}$$

see Figure 2.1.

The condition (2.31) is the natural one from the physical point of view, see [61,86].

For $\underline{\Omega = \mathbb{R}^2}$ we obtain

$$\det(\hat{A}_0(p) - \beta) = -\Delta(\beta, p, 2\mu + \lambda)(\beta^2 + \mu|p|^2)\beta,$$

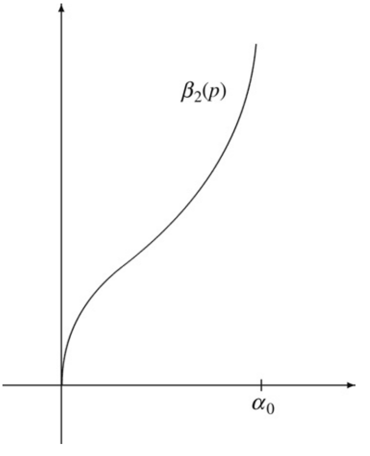

Fig. 2.1. $\{\beta_2(p) \mid p \in \mathbb{R}\}$.

and for $\underline{\Omega = \mathbb{R}^3}$

$$\det(\hat{A}_0(p) - \beta) = -\Delta(\beta, p, 2\mu + \lambda)(\beta^2 + \mu|p|^2)^2\beta^3.$$

In comparison to $\mathbb{R}^1$ we obtain, additionally, a null space (artificial for the original problem) and a "purely elastic" part with parametrized zeros $\beta_{4,5}(p)$,

$$\beta_{4,5}(p) = \pm \mathrm{i}\sqrt{\mu}|p|.$$

Altogether we obtain for $\Omega = \mathbb{R}^n$ and homogeneous, isotropic media:

THEOREM 2.10. (i) $\Omega = \mathbb{R}^1$:
$\sigma(A_0) = C_\sigma(A_0) = \Lambda^{po} := \{\beta \in \mathbb{C} \mid \exists p \in \mathbb{R}^n : \Delta(\beta, p, \alpha) = 0\}$.
(C_σ: continuous spectrum)
(ii) $\Omega = \mathbb{R}^2, \mathbb{R}^3$:
$\sigma(A_0) = P_\sigma(A_0) \cup C_\sigma(A_0)$,
$P_\sigma(A_0) = \{0\}$, $C_\sigma(A_0) = \Lambda^{po} \cup \Lambda^{so}$, where
$\Lambda^{so} := \mathrm{i}\mathbb{R}$. ($P_\sigma$: point spectrum), $\alpha = 2\mu + \lambda$ in Λ^{po}.

REMARK 2.11. The notation "Λ^{po}, Λ^{so}" is naturally connected to the decomposition of vector fields into potential and solenoidal fields.

Finally we remark that for exterior domains $\Omega \neq \mathbb{R}^n$, the spectrum has not yet been characterized in detail. It is known that, e.g.,

$$\sigma(A) \supset \sigma(A_0)$$

or that

$$\dim N(A - \lambda) < \infty \quad \text{if } \lambda \in \mathbb{C} \setminus \sigma(A_0),$$

cf. [61].

We now turn to the first results of the asymptotic behavior as $t \to \infty$ of solutions to

$$V_t(t) + AV(t) = 0, \qquad V(t=0) = V^0 \in D(A),$$

i.e., to the thermoelastic system with zero exterior forces and heat supply. This is not only of interest in itself as a description of the dynamical system in its linearized form, but also important for the later discussion of nonlinear systems, where knowledge of the behavior of the linearized system will play a crucial role.

If there are purely imaginary eigenvalues of A, then the initial data V^0, spanned by corresponding eigenvectors, produce a solution $V = V(t) = T(t)V^0$, which oscillates and preserves its norm in $\mathcal{H}$, the energy remains constant:

$$E(t) = \|V(t)\|_{\mathcal{H}}^2 = \|V^0\|_{\mathcal{H}}^2, \quad t \geq 0.$$

The aim will be to show that on the orthogonal complement in $D(A)$ of those eigenvectors, the solution will always tend to zero. Before doing so, we easily get from

$$\|V(t)\|_{\mathcal{H}}^2 = \|V^0\|_{\mathcal{H}}^2 - 2\int_0^t \langle \nabla V^3(r), K\nabla V^3(r)\rangle \mathrm{d}r. \tag{2.32}$$

THEOREM 2.12. (i) $\lim_{t\to\infty} \|V(t)\|_{\mathcal{H}}$ *exists.*
(ii) $\lim_{t\to\infty} \|\nabla V^3(t)\| = 0$.

Now we consider the asymptotic behavior in *bounded* domains. Let

$$L := \{W \in D(A) \mid \forall t \geq 0 : \|T(t)W\|_{\mathcal{H}} = \|T^*(t)W\|_{\mathcal{H}} = \|W\|_{\mathcal{H}}\},$$

where $T^*(t) = \mathrm{e}^{-tA^*}$ is the adjoint operator to $T(t)$ and the generator of $\{T^*(t)\}_{t\geq 0}$ is A^*. It turns out that L is spanned by the eigenvectors to purely imaginary eigenvalues. The operator L is closed in $D(A)$, and the decomposition

$$D(A) = L \oplus L^{\perp} \quad \text{in } D(A) \tag{2.33}$$

is used in the following main theorem on the general asymptotic behavior for bounded domains.

THEOREM 2.13. *Let Ω be bounded and let $V(t) = T(t)V^0$ be the solution of*

$$V_t(t) + AV(t) = 0, \qquad V(t=0) = V^0 \in D(A),$$

where $V^0 = V_1^0 + V_2^0$ according to (2.33). Then we have

(i) $\forall t \geq 0 : \|T(t)V_1^0\|_{\mathcal{H}} = \|V_1^0\|_{\mathcal{H}}$.
(ii) $\lim_{t\to\infty} \|V(t) - T(t)V_1^0\| = 0$.
Moreover,

(iii) *L is the closure in $D(A)$ of the span of eigenvectors to purely imaginary eigenvalues of A.*

PROOF. (i): This is obvious by the definition of L.

(ii): Let $V_2(t) := T(t)V_2^0$ and $(t_n)_n \subset (0, \infty)$ with $\lim_{n\to\infty} t_n = \infty$. Since

$$\|AV_2(t_n)\|_{\mathcal{H}} \le \|AV_2^0\|_{\mathcal{H}},$$

there is $Z \in \mathcal{H}$ such that for a subsequence, again denoted by $(AV_2(t_n))_n$,

$$AV_2(t_n) \rightharpoonup Z(\text{weakly in } \mathcal{H}).$$

The compactness of A^{-1} (Theorem 2.3) implies the existence of $W \in L^\perp$ such that

$$\lim_{n\to\infty} \|V_2(t_n) - W\|_{\mathcal{H}} = 0 \quad \text{and} \quad AW = Z.$$

For any $s \ge 0$ we get, using Theorem 2.12,

$$\|W\|_{\mathcal{H}} = \lim_{n\to\infty} \|T(s)T(t_n)V_2^0\|_{\mathcal{H}} = \|T(s)W\|_{\mathcal{H}} \tag{2.34}$$

implying

$$T^*(s)T(s)W = W. \tag{2.35}$$

Let $t_n = s + r_n$, $r_n \ge 0$.

$$\|T^*(s)W\|_{\mathcal{H}} = \lim_{n\to\infty} \|T^*(s)T(s)T(r_n)V_2^0\|_{\mathcal{H}} = \|T^*(s)T(s)W_1\|_{\mathcal{H}}$$

with, having chosen a subsequence,

$$T(s)W_1 = W,$$

hence, by (2.34) and (2.35) for W_1,

$$\|T^*(s)W\|_{\mathcal{H}} = \|W_1\|_{\mathcal{H}} = \|T(s)W_1\|_{\mathcal{H}} = \|W\|_{\mathcal{H}}.$$

The equalities (2.34) and (2.35) imply $W \in L$, hence $W = 0$, which proves (ii).

(iii): The relations

$$AL \subset \tilde{L} := \overline{L}^{\|\cdot\|_{\mathcal{H}}}, \qquad A^*L \subset \tilde{L}, \qquad A_{|\tilde{L}} = -A^*_{|\tilde{L}}$$

imply that

$$B := \mathrm{i}A_{|\tilde{L}}, \qquad D(B) := \tilde{L},$$

is a self-adjoint operator with compact inverse.

First let $W = \sum_{j=1}^{N} \alpha_j V_j, \alpha_j \in \mathbb{C}, AV_j = \lambda_j V_j, \lambda_j = \mathrm{i}b_j, b_j \in \mathbb{R}$. Then, for any $t \geq 0$,

$$T(t)W = \sum_{j=1}^{N} \alpha_j \mathrm{e}^{-\mathrm{i}b_j t} V_j, \qquad T^*(t) = \sum_{j=1}^{N} \alpha_j \mathrm{e}^{\mathrm{i}b_j t} V_j$$

and, using Lemma 2.7,

$$\|T(t)W\|_{\mathcal{H}} = \|T^*(t)W\|_{\mathcal{H}} = \|W\|_{\mathcal{H}},$$

which implies $W \in L$. Since A is closed and $T(t)$, $T^*(t)$ are continuous, we see that the $D(A)$-closure of these eigenvectors is a subset of L.

Now, since B is self-adjoint and B^{-1} is compact, there exists a complete orthonormal system $(V_j)_j$ of eigenvectors of B in $\tilde{L}$ for real eigenvalues $(d_j)_j$,

$$0 < |d_1| \leq \ldots \leq |d_j| \to \infty \quad \text{as } j \to \infty,$$
$$BV_j = d_j V_j.$$

Thus

$$AV_j = -\mathrm{i}d_j V_j$$

and for any $V^0 \in L$ we have

$$V^0 = \sum_{j=1}^{\infty} \langle V^0, V_j \rangle_{\mathcal{H}} V_j, \qquad AV^0 = \sum_{j=1}^{\infty} -\mathrm{i}d_j \langle V^0, V_j \rangle_{\mathcal{H}} V_j,$$

that is,

$$V^0 = \| \cdot \|_A - \lim_{N \to \infty} \sum_{j=1}^{N} \langle V^0, V_j \rangle_{\mathcal{H}} V_j. \qquad \square$$

COROLLARY 2.14. *A has no purely imaginary eigenvalues if and only if every solution, for arbitrary initial data from $D(A)$, tends to zero.*

This gives the first general description of the asymptotic behavior as $t \to \infty$. The discussion of the existence of purely imaginary eigenvalues in the previous section shows the complexity of the situation. There are configurations (domains, data) such that

- there are pure oscillations (sphere in $\mathbb{R}^2$),
- there is decay to zero (rectangle),
- solutions always decay to zero ($\mathbb{R}^1$),

and it is not simple to describe the geometry of domains always having solutions that decay to zero. The next questions would be if there is a rate of decay — if there is decay at all — and if a description of generic bounded domains for non-decay is possible. These questions will be addressed in the next chapters.

For *exterior* domains $\Omega \neq \mathbb{R}^n$, one can prove, with similar arguments using the "local compactness" expressed in Theorem 2.3, that the following analogue to Theorem 2.13 (ii) holds:

$$\forall R > 0 : \lim_{t\to\infty} \|V(t) - T(t)V_1^0\|_{L^2(\Omega_R)} = 0,$$

where $V_1^0 \in L$ and L are defined as before.

For $\underline{\Omega = \mathbb{R}^n}$ and homogeneous, isotropic media, an analogue to Theorem 2.3 (i), (ii) exists and can be proved using the Fourier transform.

We assume, without loss of generality,

$$\varrho = \mathrm{Id}, \quad \tilde{\delta} = 1.$$

The equations for (U, θ) given in (2.10), (2.11) are

$$U_{tt} - ((2\mu + \lambda)\nabla\nabla' - \mu\nabla \times \nabla\times)U + \tilde{\gamma}\nabla\theta = 0,$$
$$\theta_t - \kappa\Delta\theta + \tilde{\gamma}\nabla' U_t = 0.$$

$V := (S\mathcal{D}U, U_t, \theta)'$ solves

$$V_t + A_0 V = 0, \qquad V(t = 0) = V^0$$

as before, where A_0 denotes A before reducing it to $N(A)^\perp$ (cp. page 14).

A decomposition of U into its curl-free part U^{po} and its divergence-free part U^{so}, according to the orthogonal decomposition

$$L^2(\mathbb{R}^n) = \overline{\nabla H^1(\mathbb{R}^n)} \oplus \mathcal{D}_0'(\mathbb{R}^n),$$
$$U = U^{po} + U^{so},$$

where $\mathcal{D}_0'(\mathbb{R}^n)$ denotes vector fields with vanishing divergence, implies a decomposition of V into

$$V = V^{po} + V^{so},$$

where

$$V^{po} := (S\mathcal{D}U^{po}, U_t^{po}, \theta)',$$
$$V^{so} := (S\mathcal{D}U^{so}, U_t^{so}, 0)',$$

and

$$U_{tt}^{po} - (2\mu + \lambda)\Delta U^{po} + \tilde{\gamma}\nabla\theta = 0,$$
$$\theta_t - \kappa\Delta\theta + \tilde{\gamma}\nabla' U_t^{po} = 0,$$

$$U_{tt}^{so} - \mu \Delta U^{so} = 0,$$

i.e., the coupling is just between U^{po} and θ. The part U^{so}, independent of θ, satisfies a wave equation.

The space $\mathcal{H}_0 := L^2(\mathbb{R}^n)$ correspondingly can be decomposed into

$$\mathcal{H}_0 = \mathcal{H}_0^{po} \oplus \mathcal{H}_0^{so} \oplus N(A_0), \tag{2.36}$$

where

$$\begin{aligned}\mathcal{H}_0^{po} &:= \overline{S\mathcal{D}(H^1(\mathbb{R}^n) \cap \overline{\nabla H^1(\mathbb{R}^n)})} \times \overline{\nabla H^1(\mathbb{R}^n)} \times L^2(\mathbb{R}^n),\\ \mathcal{H}_0^{so} &:= \overline{S\mathcal{D}(H^1(\mathbb{R}^n) \cap \mathcal{D}_0'(\mathbb{R}^n))} \times \mathcal{D}_0'(\mathbb{R}^n) \times \{0\}.\end{aligned}$$

Looking at the characteristic polynomials in the Fourier space, $\det(\hat{A}_0(p) - \beta)$, from Section 2.1, we realize that in one space dimension $\mathcal{H}^{so}$ and $N(A_0)$ are trivial; while in two and three space dimensions, an artificial null space $N(A_0)$ – artificial in the same sense that V^{po}, V^{so} are in $\mathcal{H}_0^{po}$ and $\mathcal{H}_0^{so}$, respectively, for data $V^0 = (S\mathcal{D}U^0, U^1, \theta^0)'$ – is represented by the factor $\beta (n = 2)$ and $\beta^3 (n = 3)$, respectively, moreover the part V^{so} is reflected in the factor $(\beta^2 + \mu |p|^2)^{(2)}$, while V^{p_0} is determined via $\Delta(\beta, \tilde{\gamma}, 2\mu + \lambda)$ as in the one-dimensional case.

Transforming the equations for V^{po} under the Fourier transform,

$$\hat{V}_t^{po}(t, p) + \hat{A}_0(p)\hat{V}^{po}(t, p) = 0, \qquad \hat{V}^{po}(t = 0) = \hat{V}^{0,po},$$

we obtain a system of ordinary differential equations with parameter $p \in \mathbb{R}^n$. This can be solved explicitly, and the solution is of the form

$$\hat{V}^{po}(t, p) = \sum_{j=1}^{3} \mathrm{e}^{-\beta_j(p)t} H_j(p)$$

with appropriate $H_j(p)$ depending on $\hat{V}^{0,po} = \hat{V}^{0,po}(p)$, where $\beta_j(p)$, $j = 1, 2, 3$, are the zeros of $\Delta(\beta, p, \alpha)$, here $\alpha := 2\mu + \lambda$ and the asymptotic behavior of $\beta_j(p)$ can be computed.

LEMMA 2.15. (i) $p \neq 0 \Rightarrow \operatorname{Re} \beta_j(p) > 0$, $j = 1, 2, 3$.

(ii) $\forall r > 0 \exists C_1, C_2, C_3 > 0 :$

$$\begin{aligned} |p| \le r &\;\Rightarrow\; C_1|p|^2 \le \operatorname{Re} \beta_j(p) \le C_2|p|^2,\\ |p| \ge r &\;\Rightarrow\; \operatorname{Re} \beta_j(p) \ge C_3.\end{aligned}$$

(iii) *As* $|p| \to 0$*:*

$$\beta_1(p) = \frac{\kappa \alpha}{\alpha + \tilde{\gamma}^2}|p|^2 + \mathcal{O}(|p|^3),$$

$$\beta_{2,3}(p) = \frac{\kappa\tilde{\gamma}^2}{2(\alpha+\tilde{\gamma}^2)}|p|^2 \pm \mathrm{i}\sqrt{\alpha+\tilde{\gamma}^2}|p| + \mathcal{O}(|p|^3).$$

(iv) *As* $|p| \to \infty$*:*

$$\beta_1(p) = \kappa|p|^2 - \frac{\tilde{\gamma}^2}{\kappa} - \frac{\alpha_1}{\kappa^3}|p|^{-2} + \mathcal{O}(|p|^{-3}),$$

$$\beta_{2,3}(p) = \frac{\tilde{\gamma}^2}{2\kappa} + \frac{\alpha_1}{2\kappa^3}|p|^{-2} + \mathcal{O}(|p|^{-4}) \pm \mathrm{i}\left(\sqrt{\alpha}|p| + \frac{\alpha_2}{\kappa^2}|p|^{-1} + \mathcal{O}|p|^{-3}\right),$$

where

$$\alpha_1 := \tilde{\gamma}^2(\tilde{\gamma}^2-\alpha), \quad \alpha_2 := \frac{\tilde{\gamma}^2(4\alpha-\tilde{\gamma}^2)}{8\sqrt{\alpha}}.$$

Then it is not difficult to prove (determining $H_j(p)$ above and splitting the integral over $\mathbb{R}^n$ into two parts: $|p| \le r$, r small enough, and $|p| > r$) that $V^{po}(t)$ tends to zero as $t \to \infty$. With this sketch of the proof we conclude

THEOREM 2.16. *Let* $\Omega = \mathbb{R}^n$ *and let* $V(t) = T(t)V^0$ *be the solution to*

$$V_t + A_0 V = 0, \qquad V(t=0) = V^0,$$

and let V^0 *be decomposed as follows*

$$V^0 = V^{0,po} + V^{0,so} + V^{0,nu}$$

according to (2.36). *Then*

$$V(t) = V^{po}(t) + V^{so}(t) + V^{nu}(t) = T(t)V^{0,po} + T(t)V^{0,so} + T(t)V^{0,nu},$$

and we have

(i) $\forall t \ge 0 : V^{nu}(t) = V^{0,nu}$
(ii) $\forall t \ge 0 : \|V^{so}(t)\|_{\mathcal{H}_0} = \|V^{0,so}\|_{\mathcal{H}_0}$
(iii) $\lim_{t\to\infty} \|V^{po}(t)\|_{\mathcal{H}_0} = 0.$

This is the complete analogue to Theorem 2.13 (i), (ii), where the reduction to the orthogonal complement of the null space had already taken place. The space L is essentially replaced by divergence-free fields, in terms of the displacement vector U.

REMARK 2.17. For the discussion of the Dirichlet-type boundary condition, as done in this chapter, the boundary of the domain Ω could be arbitrary, i.e., no smoothness was necessary.

Notes: For a comprehensive treatment of linear well-posedness and asymptotic stability in bounded domains see Dafermos [12,13] and Racke [87]; for the case $\Omega = \mathbb{R}^n$ or arbitrary exterior domains see Leis [59–61] and Racke [86,87]; compare Kupradze [56] for an approach via integral equations.

2.2. *Linear asymptotic behavior in one dimension*

We consider the case of a homogeneous medium, where the differential equations are given in (2.12), (2.13), that is, assuming $\tilde{\delta} = 1$, without loss of generality, as well as zero forces and zero heat supply,

$$u_{tt} - \alpha u_{xx} + \tilde{\gamma}\theta_x = 0, \tag{2.37}$$
$$\theta_t - \kappa\theta_{xx} + \tilde{\gamma} u_{tx} = 0. \tag{2.38}$$

Additionally, we have initial conditions

$$u(t=0) = u^0, \quad u_t(t=0) = u^1, \quad \theta(t=0) = \theta^0 \quad \text{in } \Omega \tag{2.39}$$

and the following boundary conditions, for example,

$$u = \theta = 0 \quad \text{on } \partial\Omega, \tag{2.40}$$

where $\Omega = (0, 1)$, or $\Omega = (0, \infty)$ for initial boundary value problems, or $\Omega = \mathbb{R}^1$ for the Cauchy problem. We shall prove the exponential decay of solutions to the problem (2.37)–(2.40). If we apply the energy method in the usual way, some ill-behaved boundary terms arise because of the boundary conditions (2.40). To control such boundary terms we utilize a technique from the theory of boundary control (cf. (2.46) below). Let

$$E_k(t) := \frac{1}{2}\int_0^1 \left(\partial_t^k u + \alpha\partial_t^{k-1}u_x + \partial_t^{k-1}\theta\right)(t,x)\mathrm{d}x$$

denote the energy of order k. We differentiate (2.37), (2.38) $k-1$ times ($k = 1, 2$) with respect to t, then multiply the resulting equations by $\partial_t^k u$ and $\partial_t^{k-1}\theta$, respectively, integrate over (0, 1) and sum up the resulting equations. Recalling the boundary conditions (2.40), we integrate by parts to obtain

$$\frac{\mathrm{d}}{\mathrm{dt}}E_k(t) + \kappa\int_0^1 |\partial_t^{k-1}\theta_x(x,t)|^2\mathrm{d}x \le 0, \quad t \ge 0, k = 1, 2. \tag{2.41}$$

Multiplying (2.37) by $\alpha^{-1}u_{xx}$ in $L^2((0,1))$ we infer

$$\begin{aligned}\|u_{xx}(t)\|^2 &= \frac{1}{\alpha}\langle u_{tt}, u_{xx}\rangle + \frac{\tilde{\gamma}}{\alpha}\langle\theta_x, u_{xx}\rangle \\ &\le -\frac{1}{\alpha}\frac{\mathrm{d}}{\mathrm{dt}}\langle u_{tx}, u_x\rangle + \frac{1}{\alpha}\|u_{tx}\|^2 + C\|\theta_x\|^2 + \frac{1}{3}\|u_{xx}\|^2,\end{aligned}$$

whence

$$\frac{2}{3}\|u_{xx}(t)\|^2+\frac{1}{\alpha}\frac{\mathrm{d}}{\mathrm{dt}}\langle u_{tx},u_x\rangle\leq\frac{1}{\alpha}\|u_{tx}(t)\|^2+C\|\theta_x\|^2,\quad t\geq 0. \tag{2.42}$$

Similarly, we multiply (2.38) by $3(\alpha\tilde{\gamma})^{-1}u_{tx}$ in $L^2((0,1))$, utilize Eq. (2.37) and Poincaré's inequality for u_t to obtain

$$\begin{aligned}\frac{3}{\alpha}\|u_{tx}(t)\|^2&=-\frac{3\kappa}{\alpha\tilde{\gamma}}\langle\theta_x,u_{txx}\rangle+\frac{3}{\alpha\tilde{\gamma}}\langle\theta_{tx},u_t\rangle+\frac{3\kappa}{\alpha\tilde{\gamma}}\theta_x u_{tx}\,|_{x=0}^{x=1}\\&\leq-\frac{3\kappa}{\alpha\tilde{\gamma}}\frac{\mathrm{d}}{\mathrm{dt}}\langle\theta_x,u_{xx}\rangle+\frac{3\kappa}{\alpha\tilde{\gamma}}\langle\theta_{tx},u_{xx}\rangle+C\|\theta_{tx}\|\,\|u_t\|\\&\quad+C|\theta_x|_{\partial\Omega}|u_{tx}|_{\partial\Omega}\\&\leq-\frac{3\kappa}{\alpha^2\tilde{\gamma}}\frac{\mathrm{d}}{\mathrm{dt}}\langle\theta_x,u_{tt}\rangle-\frac{3\kappa}{\alpha^2}\frac{\mathrm{d}}{\mathrm{dt}}\|\theta_x\|^2+\frac{1}{6}\|u_{xx}\|^2+\frac{1}{\alpha}\|u_{tx}\|^2\\&\quad+C\|\theta_{tx}\|^2+C|\theta_x|_{\partial\Omega}|u_{tx}|_{\partial\Omega},\end{aligned}$$

where $|\cdot|_{\partial\Omega}$ denotes the L^∞-norm on $\partial\Omega$. Adding the above inequality to (2.42), one infers

$$\begin{aligned}&\frac{1}{\alpha}\|u_{tx}(t)\|^2+\frac{1}{2}\|u_{xx}(t)\|^2+\frac{\mathrm{d}}{\mathrm{dt}}\left\{\frac{1}{\alpha}\langle u_{tx},u_x\rangle+\frac{3\kappa}{\alpha^2\tilde{\gamma}}\langle\theta_x,u_{tt}\rangle+\frac{3\kappa}{\alpha^2}\|\theta_x\|^2\right\}(t)\\&\quad\leq C(\|\theta_x\|^2+\|\theta_{tx}\|^2)+C\epsilon^{-1}|\theta_x|^2_{\partial\Omega}+\epsilon|u_{tx}|^2_{\partial\Omega}\qquad(0<\epsilon<1).\end{aligned} \tag{2.43}$$

If we multiply (2.38) by $\kappa^{-1}\theta_{xx}$ in $L^2((0,1))$ and integrate by parts, we have

$$\begin{aligned}\frac{1}{2}\|\theta_{xx}(t)\|^2&\leq-\frac{1}{\kappa}\langle\theta_{tx},\theta_x\rangle+\frac{\tilde{\gamma}^2}{\kappa^2}\|u_{tx}(t)\|^2\\&\leq C(\|\theta_{tx}\|^2+\|\theta_x\|^2+\|u_{tx}\|^2).\end{aligned} \tag{2.44}$$

To bound the boundary term $|\theta_x|_{\partial\Omega}$ on the right-hand side of (2.43) we apply Sobolev's imbedding theorem ($W^{1,1}\hookrightarrow L^\infty$) and (2.44) to arrive at

$$\begin{aligned}|\theta_x|^2_{\partial\Omega}&\leq C(\|\theta_x\|^2+\|\theta_x\|\|\theta_{xx}\|)\\&\leq C\epsilon^{-2}(\|\theta_{tx}\|^2+\|\theta_x\|^2)+C\epsilon^2\|u_{tx}\|^2.\end{aligned}$$

Inserting the above estimate into (2.43) and letting ϵ be appropriately small, we obtain

$$\begin{aligned}&\frac{1}{2\alpha}\|u_{tx}(t)\|^2+\frac{1}{2}\|u_{xx}(t)\|^2+\frac{\mathrm{d}}{\mathrm{dt}}\left\{\frac{1}{\alpha}\langle u_{tx},u_x\rangle+\frac{3\kappa}{\alpha^2\tilde{\gamma}}\langle\theta_x,u_{tt}\rangle+\frac{3\kappa}{\alpha^2}\|\theta_x\|^2\right\}\\&\quad\leq C\epsilon^{-3}(\|\theta_x\|^2+\|\theta_{tx}\|^2)+C\epsilon|u_{tx}|^2_{\partial\Omega},\quad t\geq 0,\ 0<\epsilon<\epsilon_0\end{aligned} \tag{2.45}$$

for a suitable $\epsilon_0 < 1$. Now we make use of a technique from boundary control to estimate the boundary term on the right-hand side of (2.45). To this end let $\phi(x) := \frac{1}{2} - x$ for $x \in \mathbb{R}$. We differentiate (2.37) with respect to t and then multiply the resulting equation by ϕu_{tx}. We deduce that

$$\frac{\mathrm{d}}{\mathrm{dt}}\langle u_{tt}, \phi u_{tx}\rangle - \frac{1}{2}\int_0^1 \phi \partial_x u_{tt}^2 \mathrm{dx} - \frac{\alpha}{2}\int_0^1 \phi \partial_x u_{tx}^2 \mathrm{dx} + \tilde{\gamma}\langle \theta_{tx}, \phi u_{tx}\rangle = 0,$$

which, by integration by parts and Eq. (2.37), implies

$$\begin{aligned}\frac{1}{4}|u_{tt}|^2_{\partial\Omega} + \frac{\alpha}{4}|u_{tx}|^2_{\partial\Omega} + \frac{\mathrm{d}}{\mathrm{dt}}\langle u_{tt}, \phi u_{tx}\rangle &\le C(\|u_{tt}\|^2 + \|u_{tx}\|^2 + \|\theta_{tx}\|^2)\\ &\le C(\|u_{xx}\|^2 + \|u_{tx}\|^2 + \|\theta_{tx}\|^2 + \|\theta_x\|^2).\end{aligned} \tag{2.46}$$

Substituting (2.46) into (2.45) and choosing ϵ suitably small but fixed, we obtain

$$\begin{aligned}\frac{1}{4\alpha}\|u_{tx}\|^2 + \frac{1}{4}\|u_{xx}\|^2 + \frac{\mathrm{d}}{\mathrm{dt}}\left\{\frac{1}{\alpha}\langle u_{tx}, u_x\rangle + \frac{3\kappa}{\alpha^2\tilde{\gamma}}\langle \theta_x, u_{tt}\rangle + \frac{3\kappa}{\alpha^2}\|\theta_x\|^2\right\}\\ + C_0\frac{\mathrm{d}}{\mathrm{dt}}\langle u_{tt}, \phi u_{tx}\rangle \le C(\|\theta_x\|^2 + \|\theta_{tx}\|^2), \quad t \ge 0\end{aligned} \tag{2.47}$$

with C_0 being a constant independent of t.

From Eq. (2.37) we get

$$\begin{aligned}&(\|\theta\|^2 + \|\theta_t\|^2 + \|u_t\|^2 + \|u_{tt}\|^2)(t)\\ &\quad\le C(\|\theta_x\|^2 + \|\theta_{tx}\|^2 + \|u_{tx}\|^2 + \|u_{xx}\|^2)(t)\end{aligned} \tag{2.48}$$

for $t \ge 0$. Moreover, if we multiply (2.37) by u in $L^2((0,1))$ we obtain

$$\begin{aligned}\alpha\|u_x(t)\|^2 &\le C(\|u_{tt}\| + \|\theta_x\|)\|u\|\\ &\le C(\|u_{xx}\|^2 + \|\theta_x\|^2) + (\alpha/2)\|u_x\|^2,\end{aligned}$$

whence

$$\|u_x(t)\|^2 \le C(\|u_{xx}(t)\|^2 + \|\theta_x(t)\|^2), \quad t \ge 0. \tag{2.49}$$

Denote

$$\left.\begin{aligned}H(t) :&= \frac{1}{\epsilon}(E_1(t) + E_2(t)) + \frac{1}{\alpha}\langle u_{tx}, u_x\rangle + \frac{3\kappa}{\alpha^2\tilde{\gamma}}\langle \theta_x, u_{tt}\rangle\\ &\quad + \frac{3\kappa}{\alpha^2}\|\theta_x\|^2 + C_0\langle u_{tt}, \phi u_{tx}\rangle,\\ E(t) :&= E_1(t) + E_2(t) + \|\theta_x(t)\|^2,\end{aligned}\right\} \tag{2.50}$$

where $\epsilon > 0$ is a small constant to be determined later on.

Thus, (2.41) $\times\epsilon^{-1}$ (with $k = 1, 2$) + (2.47)+ (2.48) $\times\epsilon$+ (2.49) $\times\epsilon$ yields

$$\frac{\mathrm{d}}{\mathrm{dt}}H(t) + C_1\epsilon E(t) \le 0, \quad t \ge 0, 0 < \epsilon < \epsilon_0, \tag{2.51}$$

where C_1 is a positive constant independent of t, ϵ_0 some small but fixed constant. Now, in view of the definitions of $H(t)$ and $E(t)$, we choose $\epsilon \in (0, \epsilon_0)$ small enough but fixed such that

$$\frac{1}{C_2}E(t) \le H(t) \le C_2 E(t) \tag{2.52}$$

for some positive constant C_2, which does not depend on t. Combining (2.52) with (2.51), we obtain

THEOREM 2.18. *Let (u, θ) be a solution of (2.37)–(2.40). Then there are positive constants C, α independent of t, such that*

$$E(t) \le C\mathrm{e}^{-\alpha t}E(0), \quad t \ge 0,$$

where $E(t)$ is given in (2.50).

We remark that explicit rates of decay (the discussion of the optimal, the largest α in the previous theorem) for various real materials are given in [38] and will be presented in Section 3 in comparison to hyperbolic thermoelastic models.

The problem (2.37)–(2.39) with Neumann boundary conditions,

$$\sigma := \alpha u_x - \tilde{\gamma}\theta = 0, \quad \theta_x = 0 \quad \text{on } \partial\Omega, \tag{2.53}$$

has also been considered. If $u_\infty, \theta_\infty \in \mathbb{R}^1$ be defined by solving the following algebraic equations:

$$\left.\begin{aligned} \alpha u_\infty - \tilde{\gamma}\theta_\infty &= 0, \\ \tilde{\gamma}u_\infty + \theta_\infty &= \int_0^1 (\theta^0 + \tilde{\gamma}u_x^0)\mathrm{dx}, \end{aligned}\right\} \tag{2.54}$$

then $(xu_\infty, \theta_\infty)$ is a stationary solution of the system (2.37)–(2.39), (2.53) and one can show that solutions of (2.37)–(2.39), (2.53) converge to a constant state exponentially as $t \to \infty$.

For the Cauchy problem it is convenient to transform the equations (2.37), (2.38) into a symmetric system of the first order. Denote $w := \sqrt{\alpha}u_x$, $v := u_t$. Then $V := (w, v, \theta)$ satisfies

$$A^0 V_t + A^1 V_x - BV_{xx} = 0, \tag{2.55}$$

$$V(t = 0) = V^0 \equiv (w^0, v^0, \theta^0), \tag{2.56}$$

where

$$A^0 := \mathrm{Id}, \quad A^1 := \begin{pmatrix} 0 & -\sqrt{\alpha} & 0 \\ -\sqrt{\alpha} & 0 & \tilde{\gamma} \\ 0 & \tilde{\gamma} & 0 \end{pmatrix}, \quad B := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \kappa \end{pmatrix}.$$

If we take the Fourier transform of (2.55), we obtain

$$A^0 \hat{V}_t + \left\{ \mathrm{i}|\xi| A(\omega) + |\xi|^2 B \right\} \hat{V} = 0, \tag{2.57}$$

where $A(\omega) := A^1 \omega$ and $\omega := \xi/|\xi|$, $\xi \in \mathbb{R} \setminus \{0\}$.

Note that A^0, $A(\omega)$, B are all real and symmetric, and B is positive semi-definite. We take the inner product (in $\mathbb{C}^3$) of (2.57) with $\hat{V}$, and then the real part of both sides of the resulting equation to deduce that

$$\frac{1}{2} \frac{\mathrm{d}}{\mathrm{dt}} \langle A^0 \hat{V}, \hat{V} \rangle + |\xi|^2 \langle B \hat{V}, \hat{V} \rangle = 0, \tag{2.58}$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product in $\mathbb{C}^3$. Let

$$K := \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 2\sqrt{\alpha}\tilde{\gamma}^{-1} \\ 0 & -2\sqrt{\alpha}\tilde{\gamma}^{-1} & 0 \end{pmatrix}.$$

Then it is easy to see that the matrix KA^0 is real skew-symmetric and the symmetric part of the matrix $\beta K A^1 + B$ is positive semi-definite for β appropriately small, since

$$\mathrm{sym}(\beta K A^1 + B) = \begin{pmatrix} \beta\sqrt{\alpha} & 0 & \frac{\beta}{2\tilde{\gamma}}(2\alpha - \tilde{\gamma}^2) \\ 0 & \beta\sqrt{\alpha} & 0 \\ \frac{\beta}{2\tilde{\gamma}}(2\alpha - \tilde{\gamma}^2) & 0 & \kappa - 2\beta\sqrt{\alpha} \end{pmatrix}.$$

So, multiplying (2.57) by $-\mathrm{i}|\xi| K(\omega)$ ($K(\omega) := K\omega$), and then taking the inner product with $\hat{V}$, noting that $\mathrm{i}K(\omega)A^0$ is Hermitian and B positive semi-definite, we obtain, after taking the real part of the resulting equality, that

$$\begin{aligned} &-\frac{|\xi|}{2} \frac{\mathrm{d}}{\mathrm{dt}} \langle i K(\omega) A^0 \hat{V}, \hat{V} \rangle + |\xi|^2 \langle \mathrm{sym}[K(\omega)A(\omega)] \hat{V}, \hat{V} \rangle \\ &\quad = \mathrm{Re}\{\mathrm{i}|\xi|^3 \langle K(\omega) B \hat{V}, \hat{V} \rangle\} \le \epsilon |\xi|^2 |\hat{V}|^2 + C(\epsilon)|\xi|^4 \langle B\hat{V}, \hat{V} \rangle, \end{aligned} \tag{2.59}$$

where $\mathrm{sym}[K(\omega)A(\omega)]$ denotes the symmetric part of $K(\omega)A(\omega)$, and $0 < \epsilon < 1$ is to be determined below. Now set

$$E^{\beta}(t) := \frac{1}{2}\langle A^0\hat{V}, \hat{V}\rangle - \frac{\beta}{2}\frac{|\xi|}{(1+|\xi|^2)}\langle \mathrm{i}K(\omega)A^0\hat{V}, \hat{V}\rangle,$$

where β is a small positive constant to be determined later on. Then, (2.58) $\times(1+|\xi|^2)+$ (2.59) $\times\beta$ yields

$$(1+|\xi|^2)\frac{\mathrm{d}}{\mathrm{dt}}E^{\beta} + |\xi|^2\langle\{\mathrm{sym}[\beta K(\omega)A(\omega)] + B\}\hat{V}, \hat{V}\rangle + |\xi|^4\langle B\hat{V}, \hat{V}\rangle$$
$$\leq \beta\epsilon|\xi|^2|\hat{V}|^2 + \beta C(\epsilon)|\xi|^4\langle B\hat{V}, \hat{V}\rangle. \tag{2.60}$$

It can easily be seen that there is a small constant $\beta_0 > 0$ such that E^{β} is equivalent to $|\hat{V}|^2$ and that

$$\mathrm{sym}[\beta K(\omega)A(\omega)] + B = \mathrm{sym}[\beta K A^1] + B = \mathrm{sym}[\beta K A^1 + B]$$

is positive definite for any $\beta \in (0, \beta_0]$. Thus the second term on the left-hand side of (2.60) is bounded from below by $\beta C_0|\xi|^2|\hat{V}|^2$ for some constant $C_0 > 0$. Now choosing ϵ and β so that $\epsilon = C_0/2$ and $\beta := \min\{1, \beta_0, 1/C(\epsilon)\}$, the estimate (2.60) implies $E^{\beta}_t(t) + C_1\rho(|\xi|)E^{\beta}(t) \leq 0$ with $\rho(r) := r^2(1+r^2)^{-1}$. Multiplying this inequality by $\mathrm{e}^{C_1\rho(|\xi|)t}$ and integrating with respect to t, we conclude

LEMMA 2.19. *There are positive constants C, C_1 such that solutions of (2.57) satisfy*

$$|\hat{V}(t,\xi)|^2 \leq C\mathrm{e}^{-C_1\rho(|\xi|)t}|\hat{V}(0,\xi)|^2 \qquad \text{for } (t,\xi) \in \mathbb{R}_+ \times \mathbb{R}, \tag{2.61}$$

where $\rho(r) := r^2(1+r^2)^{-1}$.

As a consequence of Lemma 2.19 we have the decay rates of the solutions to the Cauchy problem (2.55), (2.56).

THEOREM 2.20. *Let $\ell \geq 0$ and $0 \leq k \leq \ell$ be integers, and let $p \in [1,2]$. Assume that $V(0) \in H^{\ell}(\mathbb{R}) \cap L^p(\mathbb{R})$. Then solutions of (2.55), (2.56) satisfy for $t \geq 0$ the estimate*

$$\|D_x^{\ell}V(t)\|^2 \leq C\left\{\mathrm{e}^{-C_1 t}\|D_x^{\ell}V(0)\|^2 + (1+t)^{-(2\lambda+\ell-k)}\|D_x^k V(0)\|_p^2\right\},$$

where $\lambda = 1/(2p) - 1/4$ and C_1 is the same constant as in Lemma 2.19.

The norm $\|\cdot\|_p$ denotes the norm in L^p.

REMARK 2.21. We define e^{-tS} by

$$(\mathrm{e}^{-tS}f)(x) = \mathcal{F}^{-1}(\mathrm{e}^{-tS(\xi)}\hat{f}(\xi))(x) \quad \text{for } f \in L^2(\mathbb{R}).$$

Then $V(t,\cdot) = (A^0)^{-1/2}\left(\mathrm{e}^{-tS}(A_0)^{1/2}V(0)\right)$ is a solution of (2.55), (2.56). Thus, Theorem 2.20 implies

$$\|D_x^{\ell}\mathrm{e}^{-tS}f\|^2 \leq C\left\{\mathrm{e}^{-C_1 t}\|D_x^{\ell}f\|^2 + (1+t)^{-(2\lambda+\ell-k)}\|D_x^k f\|_p^2\right\}.$$

Furthermore, we see that the decay rate of solutions to (2.55), (2.56) coincides with that of the linear heat equation and the equations of compressible viscous fluids.

PROOF. We multiply (2.61) with $|\xi|^{2\ell}$ and integrate over $\mathbb{R}_\xi$. We use Plancherel's theorem to infer

$$\begin{aligned}\|D_x^\ell V(t)\|^2 &\leq C\int_{\mathbb{R}} |\xi|^{2\ell} \mathrm{e}^{-C_1 t\rho(|\xi|)} |\hat{V}(0,\xi)|^2 \mathrm{d}\xi \\ &= C\left\{\int_{|\xi|\leq 1} + \int_{|\xi|\geq 1}\right\} \equiv I_1(t) + I_2(t). \end{aligned} \tag{2.62}$$

Next we estimate the integrals on the right-hand side of (2.62) in the regions $\{|\xi| \leq 1\}$ and $\{|\xi| \geq 1\}$, respectively. By virtue of Hölder's inequality and the Hausdorff–Young inequality ($\|\hat{f}\|_q \leq C\|f\|_{q'}$, $q' \in [1,2]$, $q'^{-1} + q^{-1} = 1$),

$$\begin{aligned} I_1(t) &\leq \int_{|\xi|\leq 1} |\xi|^{2\ell-2k} \mathrm{e}^{-C_1|\xi|^2 t/2} |\xi|^{2k} |\hat{V}(0,\xi)|^2 \mathrm{d}\xi \\ &\leq \left(\int_{|\xi|\leq 1} |\xi|^{2(\ell-k)r} \mathrm{e}^{-C_1 r|\xi|^2 t/2} \mathrm{d}\xi\right)^{1/r} \left(\int_{|\xi|\leq 1} |\widehat{D_x^k V(0,}\xi)|^{2r'} \mathrm{d}\xi\right)^{1/r'} \\ &\leq (1+t)^{-[(2r)^{-1}+\ell-k]} \|\widehat{D_x^k V(0)}\|_{2r'}^2 \\ &\leq C(1+t)^{-(2\lambda+\ell-k)} \|D_x^k V(0)\|_p^2, \end{aligned} \tag{2.63}$$

where $r = p/(2-p) \in [1,\infty]$, and r' is the dual number of r. For $I_2(t)$ we have

$$I_2(t) \leq C\mathrm{e}^{-C_1 t/2} \int_{|\xi|\geq 1} |\xi|^{2\ell} |\hat{V}(0,\xi)|^2 \mathrm{d}\xi \leq C\mathrm{e}^{-C_1 t/2} \|D_x^\ell V(0)\|^2. \tag{2.64}$$

Inserting (2.63), (2.64) into (2.62), we obtain the theorem. □

REMARK 2.22. We may perform a detailed analysis of the spectrum for the differential operator $A^1\partial_x - B\partial_x^2$ in (2.55), and represent the solution V of (2.55), (2.56) using the projection on the eigenspaces. In this way we can obtain the decay rates of $V(t,\cdot)$ in the L^1- and L^∞-norms. We refer to [126] for the details. We can also investigate the decay rates of solutions to initial boundary value problems in the half line ($\Omega = (0,\infty)$). In this case the arguments are similar to those in [126] for the Cauchy problem.

The decay results that have been given earlier, show that in one-dimensional space the decay, as time tends to infinity, is essentially determined by the dissipation through heat conduction. This parabolic part is the dominant one in contrast to the hyperbolic one arising from elasticity. But the hyperbolic part still has an essential impact, namely, if we seek smoothing effects of the system. The typical behavior of the solution to a pure heat equation would be that it smoothes infinitely strongly, creating arbitrarily smooth solutions for $t > 0$, also for non-smooth data at $t = 0$. This is in complete contrast to the typical hyperbolic phenomenon, where singularities in the initial data will propagate along characteristics. It turns out that this property of propagation of singularities is still present in the thermoelastic system, and in this sense it typically behaves as a hyperbolic.

Let us consider the case of a homogeneous medium, where the differential equations are given in (2.37), (2.38). Additionally, we consider the initial conditions (2.39) and the following boundary conditions,

$$u = \theta_x = 0 \quad \text{on } \partial\Omega,$$

where $\Omega = (0, 1)$, or $\Omega = (0, \infty)$, or $\Omega = \mathbb{R}^1$.

From Eqs. (2.37) and (2.38) we easily derive the following equation for $v := u$:

$$v_{ttt} - \kappa\Delta v_{tt} - (\tilde{\gamma}^2 + \alpha)\Delta v_t + \kappa\alpha\Delta^2 v = 0, \tag{2.65}$$

with initial data

$$v(t = 0) = v_0 := u^0, v_t(t = 0) = v_1 := u^1, v_{tt}(t = 0) = v_2 := \alpha\Delta u^0 - \tilde{\gamma}\nabla\theta^0,$$

and boundary conditions

$$v = \Delta v = 0 \quad \text{on } \partial\Omega.$$

We note that θ satisfies the same differential equation as v with appropriate boundary conditions. Denoting by A the Laplace operator with domain

$$D(A) := H^2(\Omega) \cap H_0^1(\Omega) \subset L^2(\Omega), \quad Av := -\Delta v,$$

we see that v satisfies

$$v_{ttt} + \kappa A v_{tt} + (\tilde{\gamma}^2 + \alpha)A v_t + \kappa\alpha A^2 v = 0, \tag{2.66}$$

$$v(t = 0) = v_0, \quad v_t(t = 0) = v_1, \quad v_{tt}(t = 0) = v_2, \tag{2.67}$$

$$v(t) \in D(A^2), \quad t \geq 0. \tag{2.68}$$

We can prove a result on the propagation of singularities for a general system of equations (2.66)–(2.68) with A being a non-negative self-adjoint operator in a separable Hilbert space $\mathcal{H}$, $v : [0, \infty) \longrightarrow \mathcal{H}$. In the formulation of the theorem and in the proof we shall use the diagonalization theorem (or spectral theorem) for self-adjoint operators in the form given in [34], cf. [63]: There exists a Hilbert space

$$\tilde{\mathcal{H}} = \int_{\oplus} \mathcal{H}(\lambda) \mathrm{d}\mu(\lambda),$$

a direct integral of Hilbert spaces $\mathcal{H}(\lambda)$, $\lambda \in \mathbb{R}$, with respect to a pointwise measure μ, and a unitary operator $\mathcal{U} : \mathcal{H} \to \tilde{\mathcal{H}}$ such that

$$D(A^m) = \{v \in \mathcal{H} \mid \lambda \mapsto \lambda^m \mathcal{U}v(\lambda) \in \tilde{\mathcal{H}}\}, \quad m \in \mathbb{N}_0,$$

and

$$\mathcal{U}(A^m v)(\lambda) = \lambda^m \mathcal{U} v(\lambda).$$

Moreover,

$$\|A^m v\|_{\mathcal{H}}^2 = \int_0^\infty \lambda^{2m} |\mathcal{U} v(\lambda)|^2_{\mathcal{H}(\lambda)} \mathrm{d}\mu(\lambda).$$

THEOREM 2.23. *Let $A \geq 0$ be self-adjoint in a separable Hilbert space $\mathcal{H}$, let v be a solution to (2.66)–(2.68). Then we have for $v_0 = v_2 = 0$ and for all $s \geq 0$ that*

$$v_1 \notin D(A^{s+2}) \Rightarrow \forall t \geq 0 : \lambda \mapsto \lambda^{s+2} \left(\mathcal{U} v_t(t, \lambda), \lambda^{1/2} \mathcal{U} v(t, \lambda) \right) \notin \mathcal{H} \times \mathcal{H}.$$

The proof of Theorem 2.23 defines $w(t, \lambda) := \mathcal{U} v(t, \lambda)$ which satisfies

$$w_{ttt} + \kappa \lambda w_{tt} + (\tilde{\gamma}^2 + \alpha) \lambda w_t + \kappa \alpha \lambda^2 w = 0, \tag{2.69}$$

$$\begin{aligned} &w(t = 0) = w_0 := \mathcal{U} v_0, \quad w_t(t = 0) = w_1 := \mathcal{U} v_1, \\ &w_{tt}(t = 0) = w_2 := \mathcal{U} v_2. \end{aligned} \tag{2.70}$$

The solution w of (2.69), (2.70) is given by

$$w(t, \lambda) = \sum_{j=1}^{3} b_j(\lambda) \mathrm{e}^{-\beta_j(\lambda) t},$$

where $\beta_j(\lambda)$, $j = 1, 2, 3$, are the roots of the characteristic equation

$$-\beta^3 + \kappa \lambda \beta^2 - (\tilde{\gamma}^2 + \alpha) \lambda \beta + \kappa \alpha \lambda^2 = 0,$$

i.e., the β_j are the roots of the characteristic polynomial $\Delta(\beta, p, \alpha)$ defined in Section 2.1 with $\lambda = |p|^2$. Moreover, the asymptotic behavior of the $b_j(\lambda)$ can be obtained using Lemma 2.15. Then it is shown by contradiction that w is not smoother than w_1.

For the Cauchy problem, i.e., $\Omega = \mathbb{R}$ for the equations (2.37)–(2.39), we can describe the propagation of singularities and the distribution of regular domains in the space-time region, respectively, if the initial data have different regularity in different parts of the real line.

We obtain from the equations (2.37), (2.38) that u and θ satisfy the following equations (cf. (2.65))

$$\begin{aligned} &P(\partial) u = 0, \\ &P(\partial) \theta = 0, \\ &u(t = 0) = u^0, \qquad u_t(t = 0) = u^1, \qquad u_{tt}(t = 0) = u^2, \end{aligned}$$

$$\theta(t=0)=\theta^0, \qquad \theta_t(t=0)=\theta^1, \qquad \theta_{tt}(t=0)=\theta^2,$$

where

$$P(\partial) := \partial_t^3 - \kappa\partial_t^2\partial_x^2 - (\alpha+\tilde{\gamma}^2)\partial_t\partial_x^2 + \kappa\alpha\partial_x^4,$$

and

$$\begin{aligned} u^2 &:= \alpha u^{0''} - \tilde{\gamma}\theta^{0'}, \\ \theta^1 &:= \kappa\theta^{0''} - \tilde{\gamma}u^{1'}, \\ \theta^2 &:= \kappa(\kappa\theta^{0''''} - \tilde{\gamma}u^{1'''}) - \tilde{\gamma}(\alpha u^{0'''} - \tilde{\gamma}\theta^{0''}). \end{aligned}$$

Here, a prime ($'$) denotes differentiation with respect to a single variable. The initial data (u^0, u^1, θ^0) are assumed to satisfy

$$\begin{aligned} &u^0, \theta^0 \in H^s(\mathbb{R}) \cap H^{s+1}(\mathbb{R}\setminus[a,b]), \\ &u^1 \in H^{s-1}(\mathbb{R}) \cap H^s(\mathbb{R}\setminus[a,b]), \end{aligned}$$

where s will be in $\mathbb{R}_0^+$ and $0 < a < b < \infty$ are fixed. The tools used to study this problem will be Fourier analysis, and explicit representations of solutions in the Fourier space, together with the knowledge about the asymptotic behavior of coefficients as the Fourier variable tends to zero or to infinity. In order to formulate the result, let us denote by I, II and III, respectively, the following three regions:

$$\begin{aligned} \mathrm{I} :=\ & \{(x,t) \mid -\infty < x < a - \sqrt{\alpha}t, 0 < t < \infty\} \\ & \cup \{(x,t) \mid \sqrt{\alpha}t + b < x < \infty, 0 < t < \infty\} \\ & \cup \left\{(x,t) \mid b - \sqrt{\alpha}t < x < \sqrt{\alpha}t + a, \frac{b-a}{2\sqrt{\alpha}} < t < \infty\right\}, \\ \mathrm{II} :=\ & \left\{(x,t) \mid a - \sqrt{\alpha}t \le x < \sqrt{\alpha}t + a, 0 < t < \frac{b-a}{2\sqrt{\alpha}}\right\} \\ & \cup \left\{(x,t) \mid a - \sqrt{\alpha}t \le x \le b - \sqrt{\alpha}t, \frac{b-a}{2\sqrt{\alpha}} \le t\right\}, \\ \mathrm{III} :=\ & a\left\{(x,t) \mid b - \sqrt{\alpha}t < x \le \sqrt{\alpha}t + b, 0 < t < \frac{b-a}{2\sqrt{\alpha}}\right\} \\ & \cup \left\{(x,t) \mid \sqrt{\alpha}t + a \le x \le \sqrt{\alpha}t + b, \frac{b-a}{2\sqrt{\alpha}} \le t\right\}. \end{aligned}$$

THEOREM 2.24. *Consider the linear Cauchy problem*

$$\left.\begin{aligned} &P(\partial)u = 0, \\ &u(t=0) = u^0, u_t(t=0) = u^1, u_{tt}(t=0) = u^2, \end{aligned}\right\} \tag{2.71}$$

and let $s \geq 4$ and $T > 0$ be fixed. If the initial data (u^0, u^1, u^2) satisfy
$u^0 \in H^s(\mathbb{R}) \cap H^{s+1}(\mathbb{R} \setminus [a,b])$, $u^1 \in H^{s-1}(\mathbb{R}) \cap H^s(\mathbb{R} \setminus [a,b])$,
$u^2 \in H^{s-3}(\mathbb{R}) \cap H^{s-2}(\mathbb{R} \setminus [a,b])$,
then the solution u to (2.71) has the following regularity:

$$
\begin{aligned}
u \in & \bigcap_{j=0}^{1} C^j([0,T], H^{s-j}(\mathbb{R})) \cap C^2([0,T], H^{s-3}(\mathbb{R})) \cap \\
& \bigcap_{j=2}^{3} H^j([0,T], H^{s+2-2j}(\mathbb{R})), \\
& (\partial_t + \sqrt{\alpha}\partial_x)u \in L^2([0,T], H^s(\mathrm{I} \cup \mathrm{III})), \\
& (\partial_t - \sqrt{\alpha}\partial_x)u \in L^2([0,T], H^s(\mathrm{I} \cup \mathrm{II})).
\end{aligned}
$$

Moreover, there is a constant $c = c(T) > 0$, depending only upon T, such that the following estimates are valid:

$$
\begin{aligned}
& \|u\|_{\bigcap_{j=0}^{1} C^j([0,T], H^{s-j}(\mathbb{R}))} + \|u\|_{C^2([0,T], H^{s-3}(\mathbb{R}))} + \|u\|_{\bigcap_{j=2}^{3} H^j([0,T], H^{s+2-2j}(\mathbb{R}))} \\
& \quad \leq c(T) \left(\|u^0\|_{s,2} + \|u^1\|_{s-1,2} + \|u^2\|_{s-3,2} \right)
\end{aligned}
$$

and

$$
\begin{aligned}
& \|(\partial_t + \sqrt{\alpha}\partial_x)u\|_{L^2([0,T], H^s(\mathrm{I}\cup\mathrm{III}))} + \|(\partial_t - \sqrt{\alpha}\partial_x)u\|_{L^2([0,T], H^s(\mathrm{I}\cup\mathrm{II}))} \\
& \quad \leq c(T) \Big(\|u^0\|_{s,2} + \|u^1\|_{s-1,2} + \|u^2\|_{s-3,2} + \|u^0\|_{H^{s+1}(\mathbb{R}\setminus[a,b])} \\
& \qquad + \|u^1\|_{H^s(\mathbb{R}\setminus[a,b])} + \|u^2\|_{H^{s-2}(\mathbb{R}\setminus[a,b])} \Big).
\end{aligned}
$$

Here we have used the following notation: For any $\Omega \subset \mathbb{R}_0^+ \times \mathbb{R}$, let $\Omega_r := \Omega \cap \{t = r\}$; we define $L^2, H^j([0,T], H^s(\Omega_t))$ as the spaces of functions belonging to $L^2, H^j([T_1,T_2], H^s([x_1,x_2]))$ for any rectangle $[T_1,T_2] \times [x_1,x_2] \subset \overline{\Omega} \cap \{0 \leq t \leq T\}$, and we omit the index t of Ω_t for simplicity.

For the proof, the problem can be divided into three problems in each of which only one of u^0, u^1 and u^2 is non-zero. Consider for example the case
$u^0 = u^1 = 0$, i.e.,

$$
\left.
\begin{aligned}
& P(\partial)u = 0, \\
& u(t=0) = u_t(t=0) = 0, u_{tt}(t=0) = u^2.
\end{aligned}
\right\} \tag{2.72}
$$

Applying the Fourier transform $\mathcal{F}$, we can express the solution u to (2.72) as

$$
u(t,\cdot) = \mathcal{F}^{-1}_{\xi \to x} \left(\sum_{j=1}^{3} b_j^2(\xi) \mathrm{e}^{-\beta_j(\xi)t} \hat{u}^2(\xi) \right), \tag{2.73}
$$

where $\beta_j(\xi)(j = 1, 2, 3)$ are again the roots of the algebraic equation

$$-\beta^3 + \kappa\xi^2\beta^2 - (\alpha + \tilde{\gamma}^2)\xi^2\beta + \kappa\alpha\xi^4 = 0$$

and

$$b_j^2(\xi) := \left(\prod_{l\neq j}(\beta_j(\xi) - \beta_l(\xi))\right)^{-1}, \qquad j = 1, 2, 3. \tag{2.74}$$

Inserting the expansions of the characteristic roots given in Lemma 2.15 into formula (2.74), we obtain the behavior of the $b_j^2(\xi)$ for both small and large $|\xi|$, respectively. Then one uses cut-off techniques to separate the behavior near zero from that near infinity and inserts the expansions into the representation of the solution. From the expansion, the leading terms in the "hyperbolic" part are perceived to be the same as those for the pure wave equation, which, after some lengthy, but straightforward calculations, will prove Theorem 2.24.

Notes: The exponential decay result is mainly taken from Muñoz Rivera [72], the part on the decay for the Cauchy problem is adapted from Kawashima's thesis [50]. The semi-axis problem was studied by Jiang [39,40]. The propagation of singularities was discussed in the paper by Racke and Wang [97] and extended by Reissig and Wang [101].

Using different methods, e.g. from semigroup theory and spectral analysis, several authors have proved the exponential decay of solutions in the (first-order) energy norm, to initial boundary value problems in bounded domains, see the papers of Haraux, S.W. Hansen, Kim, Henry, Perissinitto, Lopes, Burns, Liu, Zheng, Muñoz Rivera, also for non-homogeneous media see [27,26,54,29,3,64,65,74]. The decay of solutions in the (first-order) energy norm has an important impact in control theory (see e.g. [64]). Recently, Shibata [106] obtained decay rates of solutions to the Dirichlet problem in the half-line. There are related results by Racke, Shibata [95,105] for initial boundary value problems in bounded domains, where the method of spectral analysis is applied to get polynomial decay rates of solutions.

2.3. *Linear asymptotic behavior in several dimensions*

In one space dimension the asymptotic behavior is dominated by dissipation through heat conduction. In two or three space dimensions the hyperbolic part predominates. In general there will exist oscillations that are not damped to zero. Only for situations like the radially symmetric one, where the rotation of the displacement vector vanishes, an exponential decay result in bounded domains can be proved. For the homogeneous isotropic Cauchy problem, the system splits up into a pure hyperbolic part of the divergence-free part of the displacement vector, and into a coupling of the curl-free component of the displacement and the temperature. This coupled system behaves like a solution to a heat equation.

We start with the case of a bounded domain and then shall discuss the Cauchy problem for the isotropic case. Remarks on anistropic cases complete this sections.

Let us consider the homogeneous, isotropic linear case (2.10), (2.11) with zero forces and zero heat supply, i.e.,

$$U_{tt} - ((2\mu + \lambda)\nabla\nabla' - \mu\nabla \times \nabla\times)U + \tilde{\gamma}\nabla\theta = 0, \tag{2.75}$$

$$\tilde{\delta}\theta_t - \kappa\Delta\theta + \tilde{\gamma}\nabla' U_t = 0, \tag{2.76}$$

$$U(t=0) = U^0, \qquad U_t(t=0) = U^1, \qquad \theta(t=0) = \theta^0, \tag{2.77}$$

$$U_{|\partial\Omega} = 0, \qquad \theta_{|\partial\Omega} = 0. \tag{2.78}$$

Ω is assumed to have a smooth boundary (Lipschitz would be sufficient). In the case that the rotation of U vanishes identically, i.e., $\operatorname{rot} U \equiv \nabla \times U = 0$, we shall show that the solution (U, θ) tends to zero exponentially; this holds in particular for radially symmetric situations. The proof will use energy methods; in order to deal with certain boundary integrals we shall make use of the following well-known lemma (cp. [47]).

LEMMA 2.25. (a) *Let $v = (v_1, v_2, v_3)$ be a solution to the equations of elasticity*

$$v_{tt} - \mu\Delta v - (\mu + \lambda)\nabla\nabla' v = h_1 \quad \text{in } [0,\infty) \times \Omega,$$
$$v_{|\partial\Omega} = 0 \quad \text{in } [0,\infty).$$

Then

$$\mu \int_{\partial\Omega} \left|\frac{\partial v}{\partial \nu}\right|^2 + (\mu + \lambda)\int_{\partial\Omega} |\nabla' v|^2 = 2\frac{\mathrm{d}}{\mathrm{dt}}\int_\Omega v_t \sigma_k \partial_k v$$
$$+ \int_\Omega \nabla'\sigma \mid v_t \mid^2 + 2\mu \int_\Omega \partial_j v_i \partial_j \sigma_k \partial_k v_i - \mu \int_\Omega \nabla'\sigma \mid \nabla v|^2$$
$$+ 2(\mu + \lambda)\int_\Omega \nabla' v \nabla\sigma_k \partial_k v - (\mu + \lambda)\int_\Omega \nabla'\sigma |\nabla' v|^2 - 2\int_\Omega h_1 \sigma_k \partial_k v, \tag{2.79}$$

where $\sigma \in (C^1(\overline{\Omega}))^3$ such that $\sigma_i = \nu_i$ on $\partial\Omega$, $i = 1, 2, 3$.

(b) *Let θ be a solution to the heat equation*

$$\tilde{\delta}\theta_t - \kappa\Delta\theta = h_2 \quad \text{in } [0,\infty) \times \Omega,$$
$$\theta_{|\partial\Omega} = 0 \quad \text{in } [0,\infty).$$

Then

$$\kappa \int_{\partial\Omega} \left|\frac{\partial \theta}{\partial \nu}\right|^2 = 2\tilde{\delta}\int_\Omega \theta_t \sigma\nabla\theta + 2\kappa\int_\Omega \nabla\theta\nabla\sigma_k\partial_k\theta$$
$$- \kappa\int_\Omega \nabla'\sigma \mid \nabla\theta \mid^2 - \int_\Omega h_2 \sigma\nabla\theta. \tag{2.80}$$

Then we have the following result on exponential decay for solutions with vanishing rotation.

THEOREM 2.26. *Let* (U, θ) *be the solution to* (2.75)–(2.78) *and assume*

$$\nabla \times U = 0 \quad \text{in } [0, \infty) \times \Omega.$$

Then there are constants $\Gamma \geq 1$ *and* $d > 0$ *such that*

$$\forall t \geq 0 : E(t) + \int_0^t \mathrm{e}^{ds} \|\nabla \theta_t(s)\|^2 \mathrm{ds} \leq \Gamma E(0),$$

where

$$E(t) := \mathrm{e}^{dt} \left\{ \sum_{k=0}^{2} \|\partial_t^k U(t)\|_{2-k,2}^2 + \|\theta_t(t)\|^2 + \|\theta(t)\|_{2,2}^2 \right\}.$$

PROOF. Since U has rotation zero we have

$$\nabla \nabla' U = \Delta U \quad \text{and} \quad \|\nabla U\| = \|\nabla' U\|. \tag{2.81}$$

Let $\alpha := 2\mu + \lambda$ and

$$F_1(t) := \frac{1}{2} \left(\|U_t\|^2 + \alpha \|\nabla U\|^2 + \tilde{\delta} \|\theta\|^2 \right)(t),$$
$$F_2(t) := \frac{1}{2} \left(\|U_{tt}\|^2 + \alpha \|\nabla U_t\|^2 + \tilde{\delta} \|\theta_t\|^2 \right)(t),$$
$$F_3(t) := \frac{1}{2} \left(\|\nabla U_t\|^2 + \alpha \|\nabla \nabla' U\|^2 + \tilde{\delta} \|\nabla \theta\|^2 \right)(t),$$
$$F(t) := \sum_{j=1}^{3} F_j(t).$$

Then

$$\frac{\mathrm{d}}{\mathrm{dt}} F_1(t) = -\kappa \|\nabla \theta\|^2, \qquad \frac{\mathrm{d}}{\mathrm{dt}} F_2(t) = -\kappa \|\nabla \theta_t\|^2, \tag{2.82}$$

$$\frac{\mathrm{d}}{\mathrm{dt}} F_3(t) = -\kappa \|\Delta \theta\|^2 + \tilde{\gamma} \int_{\partial\Omega} \frac{\partial \theta}{\partial \nu} \nabla' U_t. \tag{2.83}$$

Furthermore,

$$\left| \tilde{\gamma} \int_{\partial\Omega} \frac{\partial \theta}{\partial \nu} \nabla' U_t \right| \leq \epsilon \alpha \int_{\partial\Omega} |\nabla' U_t|^2 + \frac{C}{\epsilon} \int_{\partial\Omega} \left| \frac{\partial \theta}{\partial \nu} \right|^2 \equiv I_1(t) + I_2(t), \tag{2.84}$$

where $0 < \epsilon < 1$ will be chosen below. Using Lemma 2.25, we get

$$I_1 \leq \frac{\epsilon 2\alpha}{\mu + \lambda} \frac{\mathrm{d}}{\mathrm{dt}} \int_\Omega U_{tt} \sigma_k \partial_k U_t + C\epsilon (\|U_{tt}\|^2 + \|\nabla U_t\|^2 + \|\nabla \theta_t\|^2)$$
$$- \frac{\epsilon \mu \alpha}{\mu + \lambda} \int_{\partial\Omega} \left| \frac{\partial U_t}{\partial \nu} \right|^2, \tag{2.85}$$

$$
\begin{aligned}
I_2 &\le \epsilon\|\nabla' U_t\|^2 + \frac{C}{\epsilon^2}(\|\theta_t\|^2 + \|\nabla\theta\|^2) \\
&\le \epsilon\|\nabla' U_t\|^2 + \frac{C}{\epsilon^2}(\|\nabla\theta_t\|^2 + \|\nabla\theta\|^2)
\end{aligned}
\tag{2.86}
$$

implying

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{dt}}F_3(t) \le -\kappa\|\Delta\theta\|^2 - \frac{\epsilon\mu\alpha}{\mu+\lambda}\int_{\partial\Omega}\left|\frac{\partial U_t}{\partial\nu}\right|^2 + \frac{2\epsilon\alpha}{\mu+\lambda}\frac{\mathrm{d}}{\mathrm{dt}}\int_\Omega U_{tt}\sigma_k\partial_k U_t \\
+ C\epsilon(\|U_{tt}\|^2 + \|\nabla U_t\|^2) + \frac{C}{\epsilon^2}(\|\nabla\theta_t\|^2 + \|\nabla\theta\|^2).
\end{aligned}
\tag{2.87}
$$

Choosing $\eta \ge C/(\kappa\epsilon^2) + 1$ we conclude from (2.82), (2.87), (2.81) that

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{dt}}(\eta F_1 + \eta F_2 + F_3) \le -C_1\|(\theta, \nabla\theta, \theta_t, \nabla\theta_t, \Delta\theta, U_t, \nabla U_t)\|^2 \\
&\quad - \frac{\epsilon\mu\alpha}{\mu+\lambda}\int_{\partial\Omega}\left|\frac{\partial U_t}{\partial\nu}\right|^2 + \frac{2\alpha\epsilon}{\mu+\lambda}\frac{\mathrm{d}}{\mathrm{dt}}\int_\Omega U_{tt}\sigma_k\partial_k U_t + C\epsilon\|(U_{tt}, \nabla U_t)\|^2.
\end{aligned}
\tag{2.88}
$$

Then we get for

$$
\begin{aligned}
H(t) &:= \eta F_1 + \eta F_2 + F_3 - \frac{2\epsilon\alpha}{\mu+\lambda}\int_\Omega U_{tt}\sigma_k\partial_k U_t + \sqrt{\epsilon}\int_\Omega U_t U \\
&\quad + \sqrt{\epsilon}\int_\Omega \nabla' U\,\nabla' U_t, \\
\frac{\mathrm{d}}{\mathrm{dt}}H(t) &\le -C_1\|(\theta, \nabla\theta, \theta_t, \nabla\theta_t, \Delta\theta, U_t, \nabla U_t)\|^2 \\
&\quad - C_1\sqrt{\epsilon}\|(U, \nabla U, \Delta U, U_{tt})\|^2 \\
&\quad - \frac{\epsilon\mu\alpha}{\mu+\lambda}\int_{\partial\Omega}\left|\frac{\partial U_t}{\partial\nu}\right|^2 + C\left(\epsilon\|U_{tt}\|^2 + \sqrt{\epsilon}\|(\theta, U_t, \nabla\theta, \nabla U_t)\|^2\right).
\end{aligned}
$$

Choosing ϵ such that $C\sqrt{\epsilon} = C_1/2$ we get

$$
\forall\, t \ge 0: \quad \frac{\mathrm{d}}{\mathrm{dt}}H(t) + C\left(\|\nabla\theta_t(t)\|^2 + \int_{\partial\Omega}\left|\frac{\partial U_t}{\partial\nu}\right|^2 \mathrm{dx} + F(t)\right) \le 0.
\tag{2.89}
$$

Observing (choosing η large enough) that

$$
\exists k_1, k_2 > 0 \quad \forall t \ge 0: k_1 F(t) \le H(t) \le k_2 F(t),
$$

we conclude from (2.89), with $d := C/k_2$, for all $t \ge 0$,

$$
\begin{aligned}
&\mathrm{e}^{dt}F(t) + \int_0^t \mathrm{e}^{ds}\|\nabla\theta_t(s)\|^2 \mathrm{ds} \\
&\quad + \int_0^t \mathrm{e}^{ds}\int_{\partial\Omega}\left|\frac{\partial U_t}{\partial\nu}\right|^2 \mathrm{dxds} \le CF(0) \le CE(0).
\end{aligned}
\tag{2.90}
$$

□

As an application and example we consider the equations of linear thermoelasticity for a homogeneous and isotropic medium with unit reference density in a smoothly bounded, radially symmetric domain Ω in $\mathbb{R}^n$, $n = 2, 3$; i.e.,

$$x \in \Omega \Rightarrow \forall\, R \in O(2) \text{ (if } n = 2) \text{ resp. } SO(3)(\text{if } n = 3): \quad Rx \in \Omega.$$

Here we denote by $O(n)$ the set of orthogonal $n \times n$ real matrices and by $SO(n)$ the set of matrices in $O(n)$ having determinant 1. The typical examples are spheres and annular domains.

We recall the definition of radially symmetric vector fields and functions, respectively:

DEFINITION 2.27. *A vector field $U : \Omega \longrightarrow \mathbb{R}^n$ [a function $\theta : \Omega \longrightarrow R$] is called radially symmetric, if*

$$\forall\, R \in O(2)(\textit{if } n = 2) \textit{ or } SO(3)(\textit{if } n = 3)\ \forall\, x \in \Omega :$$
$$U(Rx) = RU(x) \quad [\theta(Rx) = \theta(x)].$$

Radially symmetric functions are characterized by the following (folklore) Lemma.

LEMMA 2.28.
- (i) *A function θ is a radially symmetric function $\Longleftrightarrow$ There exists a function $\psi : \mathbb{R}_0^+ \longrightarrow \mathbb{R}$ with $\theta(x) = \psi(r)$, $r = |x|$, $x \in \Omega$.*
- (ii) *U is a radially symmetric vector field $\Longleftrightarrow$ There exists a function $\phi : \mathbb{R}_0^+ \longrightarrow \mathbb{R}$ with $U(x) = x\phi(r)$.*
- (iii) *In* (ii) *one has* $0 \in \Omega \Rightarrow U(0) = 0$.
- (iv) *A radially symmetric vector field has vanishing rotation.*

As a consequence, we can apply Theorem 2.26 and we obtain the exponential decay for radially symmetric situations:

THEOREM 2.29. *Assume that the domain Ω, having a smooth boundary, is radially symmetric, and that the initial data U^0, U^1, θ^0 are radially symmetric. Let (U, θ) be the solution to (2.75)–(2.78). Then there are constants $\Gamma \geq 1$ and $d > 0$ independent of the initial data and of t such that*

$$E(t) + \int_0^t \mathrm{e}^{ds} \|\nabla\theta_t(s)\|^2 \mathrm{ds} \leq \Gamma E(0),$$

where

$$E(t) := \mathrm{e}^{dt} \left\{ \sum_{k=0}^{2} \|\partial_t^k U(t)\|_{2-k,2}^2 + \|\theta_t(t)\|^2 + \|\theta(t)\|_{2,2}^2 \right\}.$$

In the general situation, exponential stability cannot be expected. As soon as reflecting rays exist, the decay can be arbitrarily slow. For the main result on the absence of decay rates, Theorem 2.32 below, we shall assume on the geometry of the domain:

$$\textit{There exists a two-periodic orbit of the billiard in } \Omega, \tag{2.91}$$

where an orbit of the billiard consists of lines which are reflected transversally at the boundary, and a two-periodic orbit consists of a single line which intersects the boundary perpendicularly. It is clear that such an orbit exists if the boundary $\partial\Omega$ is convex. The proof of the theorem will construct waves, the energy of which is concentrated near the two-periodic orbit of the billiard and is reduced to a similar problem for the scalar wave equation. The construction requires general results on the propagation of singularities.

First we consider the following auxiliary problem, the differential equations (2.75), (2.76) and the initial conditions (2.77), together with a non-homogeneous boundary condition on $\Gamma := \partial\Omega$,

$$U_{|\Gamma} = \Phi, \quad \theta_{|\Gamma} = 0, \quad \text{with } U^0_{|\Gamma} = \Phi(t=0), \qquad U^1_{|\Gamma} = \Phi_t(t=0). \tag{2.92}$$

We shall denote by Ω_t and Γ_t the cylinders $(0,t)\times\Omega$ and $(0,t)\times\Gamma$, respectively, $t>0$. The first-order energy is represented by

$$\|(U^0, U^1, \theta^0)\|_E^2 \equiv \frac{1}{2}\int_\Omega \Big\{\mu|\nabla U^0(x)|^2 + (\lambda+\mu)|\nabla' U^0(x)|^2 + |U^1(x)|^2 + \tilde{\delta}|\theta^0(x)|^2\Big\}\,\mathrm{d}x.$$

LEMMA 2.30. *Let $U^0 \in H^1(\Omega)$, $U^1 \in L^2(\Omega)$, $\theta^0 \in L^2(\Omega)$, $\Phi \in H^1(\Gamma_T)$ for $T>0$ and (U,θ) be a solution to (2.75)–(2.77), (2.92). Then there exists a constant $c = c(T) > 0$ such that for all $t \in [0,T]$*

$$\|(U(t), U_t(t), \theta(t))\|_E^2 + \left\|\frac{\partial U}{\partial \nu}\right\|^2_{L^2(\Gamma_t)} \le c\left\{\|(U^0,U^1,\theta^0)\|_E^2 + \|\Phi\|^2_{H^1(\Gamma_t)}\right\}.$$

PROOF. Multiplying (2.75) by U_t and (2.76) by θ in $L^2(\Omega)$, using the vector field σ from Lemma 2.25 which equals ν on Γ, and multiplying (2.75) by $\sigma_k\partial_k U$, we get

$$\left\|\frac{\partial U(t)}{\partial\nu}\right\|^2_{L^2(\Gamma)} - \frac{2}{\mu}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega U_t(t)\sigma_k\partial_k U(t)\mathrm{d}t \le C\left\{\|(U,U_t,\theta)(t)\|_E^2 + \|\Phi(t)\|^2_{H^1(\Gamma)}\right\}. \tag{2.93}$$

Multiplying (2.93) by a small $\beta>0$, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\{\mathrm{e}^{-Ct}\|(U,U_t,\theta)(t)\|_E^2 - \frac{2\beta}{\mu}\int_\Omega U_t\sigma_k\partial_k U\mathrm{d}x\right\} + \frac{\beta}{2}\int_\Gamma\left|\frac{\partial U}{\partial\nu}\right|^2\mathrm{d}\Gamma \le C(\beta,t)\|\Phi(t)\|^2_{H^1(\Gamma)} \tag{2.94}$$

for $\beta = \beta(T)$ small enough.

Integration of (2.94) from 0 to t and choosing $\beta = \beta(T)$ small enough, the claim of Lemma 2.30 follows. □

The key lemma, the proof of which will be sketched below, will be the following. Choosing appropriate coordinates we may assume, without loss of generality, that under the assumption (2.91), there exists a two-periodic orbit of the billiard which hits the boundary at $P = (0, \dots, 0, 1)$ and $Q = (0, \dots, 0, -1)$.

LEMMA 2.31. *Assume (2.91) and let $\epsilon > 0, T > 0$. There exists a closed infinite dimensional subspace $V \subset C^0([0,T], H^2(\Omega)) \cap C^1([0,T], H^1(\Omega))$ of scalar functions v satisfying*

$$v_{tt} - \mu\Delta v = 0, \qquad \frac{\partial v}{\partial \nu}|_{\partial\Omega} = 0, \qquad v(t=0) \in H_0^2(\Omega),$$

and

$$\begin{aligned}\|\partial_1 v\|^2_{H^1(\Gamma_T)} + \|\partial_n v\|^2_{H^1(\Gamma_T)} &\leq \epsilon(1+T)\{\|\partial_n v(t=0)\|^2_{1,2} \\ &+ \|\partial_1 v(t=0)\|^2_{1,2} + \|\partial_t\partial_n v(t=0)\|^2 + \|\partial_t\partial_1 v(t=0)\|^2\}.\end{aligned}$$

Now we can formulate and prove the main result on *slow* decay. By $\{T(t)\}_{t\geq 0}$ we denote the semigroup corresponding to solutions of (2.75)–(2.78).

THEOREM 2.32. *Assume (2.91). Then*

(i) $\forall t \geq 0 : \|T(t)\| = 1$.
(ii) *For any function $w : [0,\infty) \longrightarrow (0,\infty)$ with $\lim_{t\to\infty} w(t) = 0$ there exists a solution (U,θ) to (4.1)–(4.4) such that*

$$\overline{\lim_{t\to\infty}} \frac{1}{w(t)} \|(U, U_t, \theta)(t)\|_E = \infty.$$

PROOF. (ii) follows from (i) and the theorem of Banach and Steinhaus applied to the family of operators $\left\{\frac{1}{w(t)}T(t)\right\}_{t\geq 0}$ (by contradiction).

The assertion (i) is equivalent to saying that for any $\epsilon > 0$ and $t > 0$ there exists a solution $(U,\theta) \neq (0,0)$ to (2.75)–(2.78) which satisfies

$$\|(U, U_t, \theta)(t)\|_E \geq (1-\epsilon)\|(U^0, U^1, \theta^0)\|_E, \tag{2.95}$$

and we shall prove (2.95) in the sequel. Let $\epsilon > 0$ and $t > 0$ be given, let $\epsilon_1 > 0$ and for this ϵ_1, let V be as in Lemma 2.31, choose $v \in V$, $v \neq 0$, and define

$$W_0 := (\partial_n v, 0, \dots, 0, -\partial_1 v).$$

Let (U,θ) be the solution to (2.75)–(2.78) for

$$U^0 := W_0(t=0), \qquad U^1 := \partial_t W_0(t=0), \qquad \theta^0 := 0.$$

Then (W_1, θ) with

$$W_1 := U - W_0$$

is a solution to (2.75)–(2.77) and (2.92) with

$$W_1(t=0)=0, \qquad \partial_t W_1(t=0)=0, \qquad \theta(t=0)=0$$

and

$$\Phi = -W_0.$$

Let $(W_2, 0)$ solve (2.75)–(2.78) with initial data U^0, U^1, θ^0 and with $\tilde{\gamma}=0$, i.e., W_2 solves the corresponding elastic equation.

Let

$$W_3 := W - W_0.$$

By Lemmas 2.30 and 2.31 we have

$$\begin{aligned}\|(W_1, \partial_t W_1, \theta)(t)\|_E + \|(W_3, \partial_t W_3, 0)(t)\|_E &\le c(t)\|W_0\|^2_{H^1(\Gamma_t)}\\ &\le c(t)\sqrt{\epsilon_1(1+t)}\left(\|W_0(t=0)\|^2_{1,2} + \|\partial_t W_0(t=0)\|^2_{1,2}\right)^{1/2},\end{aligned}$$

hence

$$\begin{aligned}\|(U, U_t, \theta)(t)\|_E &\ge \|(W_2, \partial_t W_2, 0)(t)\|_E - (\|(W_1, \partial_t W_1, \theta)(t)\|_E\\ &\quad + \|(W_3, \partial_t W_3, 0)(t)\|_E)\\ &\ge (1-\sqrt{\epsilon_1}c(t)\sqrt{1+t})\|(U^0, U^1, 0)\|_E.\end{aligned}$$

Choosing

$$\epsilon_1 := \frac{\epsilon^2}{(c(t))^2(1+t)},$$

the proof of (2.95) is complete. □

Finally, we have to present the main task, i.e., the

PROOF OF LEMMA 2.31. In the proof we need the following lemma, which can be obtained using the energy method.

LEMMA 2.33. *Let v be a solution to the wave equation*

$$v_{tt} - \Delta v = 0 \quad in\ \Omega \times [0,\infty). \tag{2.96}$$

Then

$$\begin{aligned}\int_{\partial\Omega}\{v_t^2 + |\partial_\nu v|^2 - |\nabla_\Gamma v|^2\}\mathrm{dx} = 2\frac{\mathrm{d}}{\mathrm{dt}}\int_\Omega v_t\sigma_k\partial_k v\mathrm{dx} + \int_\Omega(\nabla'\sigma|v_t|^2\\ + 2\partial_j v\partial_j\sigma_k\partial_k v - \nabla'\sigma|\nabla v|^2)\mathrm{dx}.\end{aligned} \tag{2.97}$$

If in addition $\frac{\partial v}{\partial \nu} = 0$ *on* $\partial\Omega$*, then*

$$\int_J \int_{\partial\Omega} \{v_t^2 - |\nabla_\Gamma v|^2\} \mathrm{dx} \le C(J)(\|v_t\|^2_{L^2(\Omega_J)} + \|\nabla v\|^2_{L^2(\Omega_J)}), \tag{2.98}$$

where J *is an arbitrary connected interval in* $\mathbb{R}^1$*,* ∇_Γ *denotes the tangential gradient with respect to* Γ*, and* $\sigma \in (C^1(\overline{\Omega}))^3$ *such that* $\sigma_i = \nu_i$ *on* $\partial\Omega$*,* $i = 1, 2, 3$*.*

We recall that the boundary is given as the graph of two functions $\psi_\pm$ in a neighborhood of $P = (0, \dots, 1)$ and $Q = (0, \dots, -1) \in \mathbb{R}^n$ and that the gradient of $\psi_\pm$ is zero at 0.

Consider the problem:

$$\left.\begin{array}{ll} v_{tt} - \Delta v = 0 & \text{in } \Omega, \\ \dfrac{\partial v}{\partial \nu} = 0 & \text{on } \Gamma. \end{array}\right\} \tag{2.99}$$

The wavefront set of v is a conical set in the cotangent bundle. Suppose that we have a solution v with the wavefront set contained in the cone generated by $((t; 0, \dots, 0, t), \pm(e_0 - e_n))$ for $-\beta < t < \beta$, where e_0 is the unit cotangent vector pointing into the t direction and e_n is the cotangent vector pointing into the x_n direction. Then the wavefront set of the solution is contained in the cone generated by the broken bicharacteristics $\tilde{\gamma}$ starting at $((0, 0), \pm(e_0 - e_n))$. The wavefront set S of the restrictions to the boundary is contained in the cone generated by the union of the points

$$((4k + 1, 0, \dots, 1), (\pm 1, 0, \dots, 0)), \qquad ((4k - 1, 0, \dots, -1), (\pm 1, 0, \dots, 0)),$$

where $k \in \mathbb{Z}$. This follows, for example, from Theorem 24.2.1 of [30]. Let $\tilde{Q}$ be the singular support of $v_{|\Gamma_\mathbb{R}}$. It is contained in the union of the points $(4k + 1, 0, \dots, 1)$ and $(4k - 1, 0, \dots, -1)$.

Locally the boundary Γ near $(0, \dots, \pm 1)$ is the graph of $\psi_\pm$ over $\mathbb{R}^{n-1}$ in a neighborhood of 0 in $\mathbb{R}^{n-1}$. Having parameterized Γ in a neighborhood Z of $(0, \dots, \pm 1)$, a neighborhood U of 0 in $\mathbb{R}^{n-1}$ is determined, which parameterizes Z. In what follows we use these local coordinates. We restrict ourselves to the upper part. We introduce the symbol

$$\left.\begin{array}{l} a(t, x, s, y, \alpha, \xi) := a_1(t, x) a_2(s, y) \chi(t - s) a_3(\alpha, \xi), \\ a_1(t, x) := \eta_\beta(d^2((t, x), \tilde{Q})), a_2(s, y) := \eta_\beta(d^2((s, y), \tilde{Q})), \\ a_3(\alpha, \xi) := \eta_\beta\left(\dfrac{|\xi|^2}{|\xi|^2 + \alpha^2}\right)\left(1 - \eta(|\xi|^2 + \alpha^2)\right), \\ \quad \chi(t - s) := \eta_\beta((s - t)^6), \end{array}\right\} \tag{2.100}$$

where $\eta \in C_0^\infty$ is 1 near zero and 0 for arguments larger than 1, $\eta_\beta(s) := \eta(s/\beta^2)$, β is a small number and $d((t, x), W)$ is the Euclidian distance between (t, x) and W in $\mathbb{R} \times \mathbb{R}^{n-1}$.

Let $f \in C^\infty(\mathbb{R} \times U)$ and $z := (y, s, \xi, \alpha)$, we define the operator P_β, using the parameterization of the boundary, by

$$(P_\beta f)(t,x) := \frac{1}{(2\pi)^n} \int_{(\mathbb{R}\times\mathbb{R}^{n-1})^2} a(t,x,s,y,\alpha,\xi) e^{i((x-y,\xi)+(t-s)\alpha)} f(s,y) dz,$$

the principal symbol of which is given by a evaluated at $x = y$ and $t = s$. The relation between a and standard symbols is described, for example, in Theorem II.3.8 of Taylor's book [114]. P_β is a symmetric pseudo-differential operator with the symbol a supported in a conical β-neighborhood of $\tilde{S} = \{(t,x,t,x,\alpha,\xi) | (t,x,\alpha,\xi) \in S\}$. Moreover, by virtue of the definition (2.100), the symbol a is 1 in a conical β-neighborhood of $\tilde{S}$ intersected by the complement of the unit sphere in the cotangent variables.

Let J be an open interval in $\mathbb{R}^1$, and assume that the distance between the two ends of J and odd integers is at least $2\beta^{1/3}$. It is easy to see that the symbol $a \in S^0_{1,0}$. Hence (also recalling the parameterization of the boundary) $P_\beta : L^2(\Gamma_J) \longrightarrow L^2(\Gamma_J)$ is bounded (cf. [115]).

We shall construct an infinite-dimensional closed subspace V of solutions to (2.99), whose elements restricted to the boundary have wavefront sets contained in S. First we give a subspace W, then we shall choose V to be a suitable subspace of W at the end of the proof.

Let $\psi \in C_0^\infty(\mathbb{R}^n)$ satisfy $\int_{\mathbb{R}^n} \psi(x) dx = 1$. Then $\hat{\psi}(0) = 1$ and if $\phi := \psi * \bar{\tilde{\psi}}$, $\tilde{\psi}(x) := (-1)^n \psi(-x)$, then $\phi \in C_0^\infty(\mathbb{R}^n)$ and $\hat{\phi} = |\hat{\psi}|^2$. Thus $\hat{\phi} \geq 0$ and $\hat{\phi}(0) = 1$. Now, we put for $k \in \mathbb{N}$

$$f_k(x) := k^{n/2-8} e^{ik^4 x_n} \phi(kx).$$

Clearly

$$\|f_k\| = \|\phi\| k^{-8}, \qquad \|f_k\|_{H^2} = \|\phi\| + \mathcal{O}(k^{-3}) = \|\partial_n^2 f_k\| + \mathcal{O}(k^{-3})$$

and

$$\langle f_k, f_j \rangle_{H^2} = \mathcal{O}((|k| + |j|)^{-2})$$

if $k \neq j$, obtained by integration by parts. Now we define

$$\widetilde{W} := \left\{ \sum_{k=1}^\infty a_k f_k \in H_0^2(\Omega) \mid a_k \in \mathbb{C}, \sum_{k=1}^\infty |a_k|^2 < \infty \right\} \subset H_0^2(\Omega). \tag{2.101}$$

Then $WF(f) \subset \{(0, te_n) \mid t > 0\}$ for any $f \in \widetilde{W}$. There are indeed functions in $\widetilde{W}$ with their wavefront set being exactly equal to $\{(0, te_n) \mid t > 0\}$. If we take $a_k = k^{-n/2}$ (notice $\sum_{k=1}^\infty |a_k|^2 < \infty$), then $f = \sum_{k=1}^\infty a_k f_k = \sum_{k=1}^\infty k^{-8} e^{ik^4 x_n} \phi(kx)$ has the wavefront set $WF(f) = \{(0, te_n) \mid t > 0\}$. In fact, it is easy to see that $f \in C_0^\infty(\mathbb{R}^n \setminus \{0\})$. Hence,

$WF(f) \subset \{0\} \times (\mathbb{R}^n \setminus \{0\})$. Now

$$\hat{f}(\xi) = \sum_{k=1}^{\infty} k^{-8-n} \hat{\phi}(k^{-1}\xi - k^3 e_n),$$

whence

$$\hat{f}(m^4 e_n) = \sum_{k=1}^{\infty} k^{-8-n} \hat{\phi}((k^{-1}m^4 - k^3)e_n) \geq m^{-8-n}\hat{\phi}(0) = m^{-8-n},$$

which implies $\{(0, te_n) \mid t > 0\} \subset WF(f)$. Suppose now U_W is a closed cone disjoint from $\{te_n \mid t > 0\}$. Then there exists a constant $c > 0$ such that $|\xi - te_n| \geq c(|\xi| + t)$ for $\xi \in U_W$ and $t > 0$. It follows that $|k^{-1}\xi - k^3 e_n| = k^{-1}|\xi - k^4 e_n| \geq ck^{-1}(|\xi| + k^4) \geq 2ck|\xi|^{1/2}$. Since $|\hat{\phi}(\xi)| \leq C_N(1 + |\xi|^2)^{-N}$ for any N, we have

$$|\hat{\phi}(k^{-1}\xi - k^3 e_n)| \leq \sup_{|\omega| \geq 2ck\sqrt{|\xi|}} |\hat{\phi}(\omega)| \leq C_N(1 + 4c^2k^2|\xi|)^{-N} \quad \text{for } \xi \in U_W.$$

Then

$$|\hat{f}(\xi)| \leq C_N \sum_{k=1}^{\infty} k^{-8-n}(1 + 4c^2k^2|\xi|)^{-N} \leq C_N(1 + |\xi|)^{-N}$$

for $\xi \in U_W$. It follows that $WF(f) = \{(0, te_n) \mid t > 0\}$. Now, let $v_0 \in \widetilde{W}$ and $v_1 = 0$. We denote by v the solution of the homogeneous Neumann problem (2.99) with the initial data $(v, v_t)|_{t=0} = (v_0, v_1)$. We define W as the space of these solutions. Thus, by virtue of the theory of propagation of singularities (see [114, Chapter IX, Theorem 3.3]), we have that $WF(v)|_{\Gamma_{\mathbb{R}}} \subset S$ for $v \in W$.

For $v \in W$ we set $u = \partial_t v$ or $u = \partial_i v$ for $1 \leq i \leq n-1$. Denote

$$B_0(u) := \int_{\Gamma_J} (\partial_\nu u)^2 + u_t^2 - |\nabla_\Gamma u|^2, \tag{2.102}$$

where ∇_Γ denotes the tangential part of the gradient. Then by Lemma 2.33,

$$|B_0(u)| \leq C\|v\|_{H^2(\Omega_J)}. \tag{2.103}$$

We write $u_{|\Gamma_J} = P_\beta u_{|\Gamma_J} + (1 - P_\beta)u_{|\Gamma_J}$ with $u = \partial_t v$ or $u = \partial_i v$ for $1 \leq i \leq n-1$. Because the wavefront set of u is contained in S and $1 - a = 0$ in a conical neighborhood of S, the map

$$W \ni v \mapsto (1 - P_\beta)u_{|\Gamma_J} \in C^k(\Gamma_J) \quad \text{for } k \in \mathbb{N}$$

is well defined (cf. [114, Chapter VI, Proposition 1.4]). We know that $1 - P_\beta : L^2(\Gamma_J) \longrightarrow L^2(\Gamma_J)$ is bounded. Let $\mathcal{D}((1 - P_\beta)^*)$ denote the definition range of the adjoint operator $(1 - P_\beta)^*$ of $1 - P_\beta$. It is easy to check that $1 - P_\beta : L^2(\Gamma_J) \longrightarrow C^k(\Gamma_J)$ is closed. In fact, let $u_n \to u$ in $L^2(\Gamma_J)$ resp. $(1 - P_\beta)u_n \to h$ in $C^k(\Gamma_J)$ as $n \to \infty$. Then for any test function $\phi \in C_0^\infty(\Gamma_J) \subset \mathcal{D}((1-P_\beta)^*)$, one has

$$\begin{aligned}\langle h, \phi\rangle_{L^2(\Gamma_J)} &\leftarrow \langle (1-P_\beta)u_n, \phi\rangle_{L^2(\Gamma_J)} = \langle u_n, (1-P_\beta)^*\phi\rangle_{L^2(\Gamma_J)} \\ &\to \langle u, (1-P_\beta)^*\phi\rangle_{L^2(\Gamma_J)} = \langle (1-P_\beta)u, \phi\rangle_{L^2(\Gamma_J)}.\end{aligned}$$

Hence, $h = (1-P_\beta)u$, and $1-P_\beta$ is closed. It follows from the closed graph theorem that for arbitrary but fixed β the map $1-P_\beta : L^2(\Gamma_J) \longrightarrow C^k(\Gamma_J)$ is bounded. Furthermore, $1-P_\beta : L^2(\Gamma_J) \longrightarrow H^1(\Gamma_J)$ is compact, since it factors over compact imbeddings.

For $t \in J$ we have supp $\chi(t-\cdot) \subset [\inf J - \beta^{1/3}, \sup J + \beta^{1/3}]$. We notice that the support of $a_2(x,s)$, with respect to s, lies in the β-neighborhoods of odd integers. Hence, recalling the choice of J, $a_2(s,x)\chi(t-s) = 0$, if $s \notin J$. Now, let $\phi \in L^\infty(\mathbb{R})$ with $\phi(t) = 1$ for $t \in J$ and $\phi(t) = 0$ for $t \notin J$. Thus, recalling the definition of $P_\beta f$ and (2.100), one finds that

$$(2\pi)^{n/2}(P_\beta u)(t,x) = a_1(t,x)\left((a_2 u\phi) * h\right)(t,x) \quad \text{for } t \in J, \tag{2.104}$$

where $h(s,y) := \chi(s)\widehat{a_3}(s,y)$, whence

$$\begin{aligned}(2\pi)^n \int_J \|P_\beta u\|^2 \mathrm{dt} &= \|a_1(t,x)((a_2 u\phi) * h)\|^2_{L^2(\Gamma_J)} \\ &\le \|((a_2 u\phi) * h)\|^2_{L^2(\mathbb{R}\times\mathbb{R}^{n-1})} \\ &= \|\widehat{a_2 u\phi} \cdot \hat{h}\|^2_{L^2(\mathbb{R}\times\mathbb{R}^{n-1})}.\end{aligned} \tag{2.105}$$

On the other hand,

$$\begin{aligned}&(2\pi)^{n/2}|\hat{h}(\alpha,\xi)| \\ &\quad = \left|\int_{\mathbb{R}\times\mathbb{R}^{n-1}} \chi(s)\mathrm{e}^{-\mathrm{i}s\alpha} \int_{\mathbb{R}\times\mathbb{R}^{n-1}} a_3(t,x)\mathrm{e}^{-\mathrm{i}(xy+ts)}\mathrm{dxdte}^{-\mathrm{i}y\xi}\mathrm{dyds}\right| \\ &\quad = \left|\int_{\mathbb{R}} \hat{\chi}(t+\alpha)a_3(t,-\xi)\mathrm{dt}\right| \\ &\quad \le C\beta^{-1/3}\max_t |a_3(t,-\xi)|.\end{aligned} \tag{2.106}$$

Inserting (2.106) into (2.105), we obtain

$$\|P_\beta u\|_{L^2(\Gamma_J)} \le C\beta^{-1/3}\max_{t,\xi}|a_3(t,\xi)|\|u\|_{L^2(J\times\mathbb{R}^{n-1})}. \tag{2.107}$$

Next we estimate the tangential derivative of $P_\beta u$. In view of the definition of $P_\beta u$ we may write

$$\begin{aligned}(2\pi)^{n/2}\partial_j P_\beta u = \int_{(\mathbb{R}\times\mathbb{R}^{n-1})^2} &\left\{\partial_j a(t,x,s,y,\alpha,\xi)\right. \\ &\left. + \ \mathrm{i}\xi_j a(t,x,s,y,\alpha,\xi)\right\} \mathrm{e}^{\mathrm{i}((x-y,\xi)+(t-s)\alpha)}u(y,s)\mathrm{dydsd}\xi\mathrm{d}\alpha \\ \equiv I_1(t,x) + I_2(t,x).&\end{aligned} \tag{2.108}$$

Denote $K := (\inf J - \beta^{1/3}, \sup J + \beta^{1/3})$. Let $\psi \in C^1(\mathbb{R})$ with $\psi(t) = 1$ for $t \in J$ and $\psi = 0$ for $t \in \mathbb{R} \setminus K$. Analogously to (2.104), I_2 can be written in the form

$$I_2(t,x) = a_1(t,x)\,(a_2 u \psi) * h_1, \quad t \in J,$$

where $h_1(s,y) := \chi(s)\widehat{b_1}(s,y)$ and $b_1(\alpha,\xi) := \mathrm{i}\xi_j a_3(\alpha,\xi)$.

Recalling that $\operatorname{supp} a_3(t,\xi) \subset \{(t,\xi) \,||\xi| \le C\beta|t|\}$, we have in the same manner as for (2.106) that,

$$|\widehat{h_1}(\alpha,\xi)| \le C(1+|\alpha|)\beta^{1/3}. \tag{2.109}$$

By virtue of the definition of a_2, J and K, we know $\operatorname{supp} a_2(\cdot,x) \cap (K \setminus J) = \phi$. Thus we have $\operatorname{supp}[a_2(\cdot,x)\psi(\cdot)] \subset J$. So, using (2.109) and (2.107), keeping in mind that $1 - P_\beta : L^2(\Gamma_J) \longrightarrow H^1(\Gamma_J)$ is bounded, we infer

$$\begin{aligned}\|I_2\|^2_{L^2(\Gamma_J)} &\le C\|(a_2 u\psi) * h_1\|^2_{L^2(\Gamma_{\mathbb{R}})}\\ &\le C\beta^{2/3}\|(1+|\alpha|)\widehat{a_2 u\psi}\|^2_{L^2(\Gamma_{\mathbb{R}})}\\ &\le C\beta^{2/3}\|\partial_t u\|^2_{L^2(\Gamma_J)} + C(\beta)\|u\|^2_{L^2(\Gamma_J)}\\ &\le C\beta^{2/3}\|\partial_t P_\beta u\|^2_{L^2(\Gamma_J)} + C(\beta)\|u\|^2_{L^2(\Gamma_J)}. \end{aligned} \tag{2.110}$$

From (2.107) we easily get $\|I_1\|^2_{L^2(\Gamma_J)} \le C(\beta)\|u\|^2_{L^2(\Gamma_J)}$. Substituting this and (2.110) into (2.108), we obtain

$$\|\nabla_\Gamma P_\beta u\|^2_{L^2(\Gamma_J)} \le C\beta^{2/3}\|\partial_t P_\beta u\|^2_{L^2(\Gamma_J)} + C(\beta)\|u\|^2_{L^2(\Gamma_J)}. \tag{2.111}$$

On the other hand, utilizing (2.102), (2.103), taking into account $\partial_\nu u_{|\partial\Omega} = 0$ and the boundedness of the map $1 - P_\beta : L^2(\Gamma_J) \longrightarrow C^k(\Gamma_J)$, we deduce that

$$\begin{aligned}\|\partial_t P_\beta u\|^2_{L^2(\Gamma_J)} &\le C\{B_0(u) + \|\nabla_\Gamma u\|^2_{L^2(\Gamma_J)}\} + C(\beta)\|u\|^2_{L^2(\Gamma_J)}\\ &\le C\left\{B_0(u) + \|\nabla_\Gamma P_\beta u\|^2_{L^2(\Gamma_J)} + \|\nabla_\Gamma(1-P_\beta)u\|^2_{L^2(\Gamma_J)}\right\} + C(\beta)\|u\|^2_{L^2(\Gamma_J)}\\ &\le C\{\|v\|^2_{H^2(\Omega_J)} + \|\nabla_\Gamma P_\beta u\|^2_{L^2(\Gamma_J)}\} + C(\beta)\|u\|^2_{L^2(\Gamma_J)}. \end{aligned} \tag{2.112}$$

Inserting (2.112) into (2.111) and choosing βappropriately small, we obtain

$$\|\nabla_\Gamma P_\beta u\|^2_{L^2(\Gamma_J)} \le C\beta^{2/3}\|v\|^2_{H^2(\Omega_J)} + C(\beta)\|u\|^2_{L^2(\Gamma_J)}. \tag{2.113}$$

To make the second term on the right-hand side of (2.113) small we need the following lemma, which can easily be obtained using the spectral representation for compact operators.

LEMMA 2.34. *Let H_1 and H_2 be Hilbert spaces, let $T : H_1 \longrightarrow H_2$ be a compact linear operator, and $\{h_i\}$ a basis in H_1. Then for any $\epsilon > 0$ there exists N such that the restriction $T : \operatorname{span}\{h_N, h_{N+1}, \ldots\} \longrightarrow H_2$ has norm smaller than ϵ.*

By Lemma 2.34 we can choose a subspace $\tilde{V}$ of W with finite codimension such that the L^2-norm in (2.113) is controlled by the first term on the right-hand side. Hence the operator

$$\tilde{V} \ni v \mapsto u_{|\Gamma_J} \in H^1(\Gamma_J)$$

is compact because it can be written as the sum of an operator of size $O(\beta^{1/3})$ and a compact operator (recall $v \mapsto u = (1 - P_\beta)u + P_\beta u$). Hence, in view of Lemma 2.34, we may take V to be a suitable subspace $V \subset \tilde{V} \subset W$ with finite codimension such that

$$\begin{aligned}
&\|\partial_1 v\|^2_{H^1(\Gamma_J)} + \|\partial_n v\|^2_{H^1(\Gamma_J)} \le \|u\|^2_{H^1(\Gamma_J)} + \|\partial_n^2 v\|^2_{L^2(\Gamma_J)} \\
&\quad \le \|u\|^2_{H^1(\Gamma_J)} + C\|\partial_t^2 v\|^2_{L^2(\Gamma_J)} + C\sum_{k=1}^{n-1} \|\partial_k^2 v\|^2_{L^2(\Gamma_J)} \\
&\quad \le \|u\|^2_{H^1(\Gamma_J)} \le C\beta\|v\|^2_{H^2(\Omega_J)} \le C\beta(\|v(0)\|^2_{2,2} + \|v_t(0)\|^2_{1,2}) \\
&\quad \le C\beta\|\partial_n v(0)\|^2_{1,2} \quad \text{for } v \in V,
\end{aligned}$$

where we have also used (2.99). The last inequality holds at least if $a_k = 0$ for $k \le N$ and some large number N, using the definition of $\widetilde{W}$ in (2.101). So, we can choose β appropriately small to obtain the assertion of Lemma 2.31. □

Similar and extended results on slow decay were obtained by Lebeau and Zuazua [58], partially, also going back to the work of Henry, Perissinitto, and Lopes [29]. They used the theory of geometrical optics and propagation of singularities to prove the non-uniform decay of solutions of (2.75)–(2.78) for a large class of domains:

Let $\Omega \subset \mathbb{R}^n$ ($n = 2$ or $n = 3$) be a bounded domain with smooth boundary. Assume that Ω is such that there exists a ray of geometrical optics in Ω of arbitrary length which is always reflected perpendicularly on the boundary. Then the decay of solutions to (2.75)–(2.78) is not uniform.

Furthermore, among other points, they proved in the case of $n = 2$, the following:

Let Ω be a bounded domain in $\mathbb{R}^2$ with smooth boundary and without contacts of infinite order with its tangents. Assume that Ω is neither a sphere nor an annulus and satisfies the following condition

$$\begin{cases}
\text{If } \varphi \in \left(H_0^1(\Omega)\right)^n & \text{is such that} \\
-\Delta\varphi = b^2\varphi & \text{in } \Omega, \\
\nabla'\varphi = 0 & \text{in } \Omega, \\
\varphi = 0 & \text{on } \partial\Omega \\
\text{for some } b \in \mathbb{R}, & \text{then } \varphi = 0.
\end{cases}$$

Then there exists $C > 0$ such that

$$\|(U, U_t, \theta)(t)\|^2_E \le Ct^{-1}(\|U^0\|^2_{2,2} + \|U^1\|^2_{1,2} + \|\theta^0\|^2_{2,2}), \quad t > 0$$

for every solution to (2.75)–(2.78) with initial data (U^0, U^1, θ^0) in $\left(H^2(\Omega) \cap H_0^1(\Omega)\right) \times H_0^1(\Omega) \times \left(H^2(\Omega) \cap H_0^1(\Omega)\right)$.

For the two- and three-dimensional Cauchy problem in the isotropic case, (2.75), (2.76) in $\mathbb{R}^n$ ($n = 2$ or $n = 3$), we shall apply the method used for the one-dimensional case. Without loss of generality, we may assume $\tilde{\delta} = 1$ in (2.76) and rewrite Eqs. (2.75) and (2.76) as follows

$$\left.\begin{aligned} &U_{tt} - \mu\Delta U - (\lambda + \mu)\nabla\nabla' U + \tilde{\gamma}\nabla\theta = 0, \\ &\theta_t - \kappa\Delta\theta + \tilde{\gamma}\nabla' U_t = 0, \end{aligned}\right\} \tag{2.114}$$

where $t \geq 0$, $x \in \mathbb{R}^n$ ($n = 2$ or 3), with prescribed initial data

$$U(t=0) = U^0, \qquad U_t(t=0) = U^1, \qquad \theta(t=0) = \theta^0 \quad \text{in } \mathbb{R}^n. \tag{2.115}$$

With the decomposition of U into its curl-free part U^{po} and its divergence-free part U^{so}, $U = U^{po} + U^{so}$, we obtain a decomposition of (2.114) into the two systems

$$\left.\begin{aligned} &U_{tt}^{po} - \alpha\nabla\nabla' U^{po} + \tilde{\gamma}\nabla\theta = 0, \\ &\theta_t - \kappa\Delta\theta + \tilde{\gamma}\nabla' U_t^{po} = 0, \\ &U^{po}(t=0) = (U^0)^{po}, U_t^{po}(t=0) = (U^1)^{po}, \theta(t=0) = \theta^0 \quad \text{in } \mathbb{R}^n, \end{aligned}\right\} \tag{2.116}$$

where $\alpha = \lambda + 2\mu$, and

$$\left.\begin{aligned} &U_{tt}^{so} - \mu\Delta U^{so} = 0, \\ &U^{so}(t=0) = (U^0)^{so}, \qquad U_t^{so}(t=0) = (U^1)^{so} \quad \text{in } \mathbb{R}^n. \end{aligned}\right\} \tag{2.117}$$

The asymptotic behavior of solutions to the linear wave equation (2.117) is well known, see for example [91, Chapter 2], we have for all $t \geq 0$ the fact that

$$\|(\partial_t U^{so}, \nabla U^{so})(t)\|_q \leq C(1+t)^{-\frac{(n-1)}{2}(1-2/q)} \|(\partial_t U^{so}, \nabla U^{so})(0)\|_{m_p,p}, \tag{2.118}$$

where $2 \leq q \leq \infty$, $1/p + 1/q = 1$ and $m_p \leq 3$ is an integer. (In fact, $m_p > n(1 - 2/q)$ is an arbitrary integer for $2 < q < \infty$ and $m_p = n(1 - 2/q)$ for $q = 2$ or $q = \infty$.) The asymptotic behavior of U^{po} will be described now. Let $V^{po} := (\sqrt{\alpha}\nabla' U^{po}, U_t^{po}, \theta)' \in \mathbb{R}^{n+2}$. Thus we may write the equations (2.116) as a first-order system

$$V_t^{po} + AV^{po} - B\Delta V^{po} = 0, \qquad V^{po}(0) = (V^{po})^0, \tag{2.119}$$

where $(V^{po})^0 := (\sqrt{\alpha}\nabla'(U^0)^{po}, (U^1)^{po}, \theta^0)'$ and

$$A := \begin{pmatrix} 0 & -\sqrt{\alpha}\nabla' & 0 \\ -\sqrt{\alpha}\nabla & 0 & \tilde{\gamma}\nabla \\ 0 & \tilde{\gamma}\nabla' & 0 \end{pmatrix}, \quad B := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \kappa \end{pmatrix}.$$

Taking the Fourier transform on both sides of (2.119), we arrive at

$$\begin{aligned} &\hat{V}_t^{po} + \left\{ \mathrm{i}|\xi| A(\omega) + |\xi|^2 B \right\} \hat{V}^{po} = 0, \\ &\omega := \xi/|\xi| \equiv (\omega_1, \ldots, \omega_n)' \in S^{n-1}, \end{aligned} \tag{2.120}$$

where

$$A(\omega) := \begin{pmatrix} 0 & -\sqrt{\alpha}\omega' & 0 \\ -\sqrt{\alpha}\omega & 0 & \tilde{\gamma}\omega \\ 0 & \tilde{\gamma}\omega' & 0 \end{pmatrix} \quad \text{is real symmetric.}$$

We introduce a real symmetric matrix

$$K(\omega) := \begin{pmatrix} 0 & -\omega' & 0 \\ \omega & 0 & 2\sqrt{\alpha}\tilde{\gamma}^{-1}\omega \\ 0 & -2\sqrt{\alpha}\tilde{\gamma}^{-1}\omega' & 0 \end{pmatrix}, \quad \omega = (\omega_1, \ldots, \omega_n)' \in S^{n-1}.$$

Note that the 2nd to $(n+1)$th components of the solution V^{po} of (2.120) build a curl-free vector, that is, for $V^{po} = (v_1, W, v_{n+2})' \in \mathbb{R}^{n+2}$, one has $\nabla \times W = 0$ (or equivalently $\omega \times \hat{W} = 0$ for $\omega \in S^{n-1}$). Since

$$\operatorname{sym}\,(\beta K(\omega) A(\omega) + B) = \begin{pmatrix} \beta\sqrt{\alpha} & 0 & \dfrac{\beta}{2\tilde{\gamma}}(2\alpha - \tilde{\gamma}^2) \\ 0 & \beta\sqrt{\alpha}\omega\omega' & 0 \\ \dfrac{\beta}{2\tilde{\gamma}}(2\alpha - \tilde{\gamma}^2) & 0 & \kappa - 2\beta\sqrt{\alpha} \end{pmatrix},$$

it is easy to see that for any $\hat{V}^{po} = (\hat{v}_1, \hat{W}, \hat{v}_{n+2})' \in \mathbb{R}^{n+2}$ with $\omega \times \hat{W} = 0$ we have

$$\left\langle \operatorname{sym}[\beta K(\omega) A(\omega) + B] \hat{V}^{po}, \hat{V}^{po} \right\rangle \geq C |\hat{V}^{po}|^2 \tag{2.121}$$

for some suitably small $\beta > 0$, where $\langle \cdot, \cdot \rangle$ also denotes the standard inner product in $\mathbb{C}^{n+2}$. Thus, using (2.120) and (2.121) and following the same procedure as used for (2.58)–(2.61) in Subsection 3.1.2, we get

$$|\hat{V}^{po}(t,\xi)|^2 \leq C \mathrm{e}^{-C_1 \rho(|\xi|) t} |\hat{V}^{po}(0,\xi)|^2, \quad (t,\xi) \in [0,\infty) \times \mathbb{R}^n, \tag{2.122}$$

where $\rho(r) = r^2(1+r^2)^{-1}$, and C_1 is a positive constant.

Recalling $\Delta = \nabla\nabla' - \nabla \times \nabla \times$, we have

$$\|\nabla U^{po}\|^2 = (\Delta U^{po}, U^{po}) = \|\nabla' U^{po}\|^2. \tag{2.123}$$

So, using (2.123), we obtain analogously to (2.62)–(2.64)

LEMMA 2.35. *Let k, l be integers, $0 \le k \le l$, and let $p \in [1,2]$. Assume that $(\nabla'(U)^{po}, (U^1)^{po}, \theta^0)' \in H^l(\mathbb{R}^n) \cap L^p(\mathbb{R}^n)$. Then the solution of (2.116) satisfies*

$$\begin{aligned}&\|\nabla^l(\nabla U^{po}, U_t^{po}, \theta)(t)\|^2 \le C\|\nabla^l V^{po}(t)\|^2\\&\quad\le C\left\{\mathrm{e}^{-C_1 t}\|\nabla^l V^{po}(0)\|^2 + (1+t)^{-(2\lambda+l-k)}\|\nabla^k V^{po}(0)\|_p^2\right\},\end{aligned}$$

where $V^{po} = (\sqrt{\alpha}\nabla' U^{po}, U_t^{po}, \theta)'$, $\lambda = n[1/(2p) - 1/4]$, and C_1 is a positive constant.

Applying (2.122) and interpolation theory, we can obtain the L^p–L^q-estimates for V.

LEMMA 2.36. *Let l be an integer and $q \in [2,\infty]$. Then*

$$\|\nabla^l V^{po}(t)\|_q \le C(1+t)^{-[(n+l)/2-n/q]}\|V^{po}(0)\|_{N_p+l,p}, \quad t \ge 0,$$

where $V^{po} = (\sqrt{\alpha}\nabla' U^{po}, U_t^{po}, \theta)'$, $N_p \in \mathbb{N}$, $N_p \le 4$.

(In fact, $N_p \ge (n+1)(1-2/q)$ is an arbitrary integer where the equality holds only for $q = 2$ and $q = \infty$.) □

Using (2.122), the estimates

$$\begin{aligned}&\int |\xi|^l |\hat{V}^{po}(t,\xi)|\mathrm{d}\xi\\&\quad\le C\left(\int \mathrm{e}^{-C_1\rho(|\xi|)t}|\xi|^{2l}(1+|\xi|)^{-m}\right)^{1/2}\left(\int (1+|\xi|)^m |\hat{V}^{po}(0,\xi)|^2\mathrm{d}\xi\right)^{1/2},\\&\int |\xi|^{2l}|\hat{V}^{po}(t,\xi)|^2\mathrm{d}\xi\\&\quad\le C\sup_\xi(1+|\xi|)^m|\hat{V}^{po}(0,\xi)|^2\int |\xi|^{2l}(1+|\xi|)^{-m}\mathrm{e}^{-C_1\rho(|\xi|)t}\mathrm{d}\xi\end{aligned}$$

for an integer $m \ge 0$, we obtain the following L^2–L^∞- resp. L^1–L^2-estimates.

LEMMA 2.37. *Let l be an integer. Then*

$$\begin{aligned}&\|\nabla^l V^{po}(t)\|_\infty \le C(1+t)^{-(n/4+l/2)}\|V(0)\|_{l+2,2},\\&\|\nabla^l V^{po}(t)\| \le C(1+t)^{-(n/4+l/2)}\|V(0)\|_{l+2,1}, \quad t \ge 0\end{aligned}$$

where $V^{po} = (\sqrt{\alpha}\nabla' U^{po}, U_t^{po}, \theta)'$.

Altogether we get

THEOREM 2.38. *We have for $l \ge 0$ and all $t \ge 0$,*

$$\begin{aligned}&\|\nabla^l V(t)\|_q \le C(1+t)^{-\frac{(n-1)}{2}(1-2/q)}\|V(0)\|_{N_p+l,p}, \qquad (2.124)\\&\|\nabla^l \partial_t^j \theta\|_\infty \le C(1+t)^{-[n/4+(l+j)/2]}\|V(0)\|_{l+2+2j,2}, \quad j = 0,1,\\&\|\nabla^l \partial_t^j \theta\| \le C(1+t)^{-[n/4+(l+j)/2]}\|V(0)\|_{l+2+2j,1}, \quad j = 0,1,\end{aligned}$$

where $V(t) := (\nabla U, U_t, \theta)(t)$, $q \in [2, \infty]$, $p^{-1} + q^{-1} = 1$, *and* $N_p \leq 4$ *is the same as in Lemma* 2.36.

We note that the detailed asymptotics in exterior domains (complements of compact sets) have not yet been treated. First steps with low-frequency asymptotics of resolvents will be contained in a paper [77], essentially written for a system with *second sound*.

The special structure of the differential equations in the isotropic case above is used to obtain the decay rates just proved. This changes if the medium is anisotropic. This phenomenon was already observed for pure elasticity, and similarly for the equations of crystal optics, see [110,111,62,91]. There, it led to different decay rates for solutions depending on the special cubic medium. The complexity of the analysis in the pure elastic case restricted the considerations to spatially two-dimensional models. For thermoelastic problems in $\mathbb{R}^2$, Borkenstein first investigated decay rates [2] (cp. [47]), then Doll [18] extended the results to more general anisotropic situations. Recently Reissig and Wirth [102,121] improved the results with a rather general approach.

Notes: The results for radial symmetry in bounded domains are taken from Jiang, Muñoz Rivera and Racke [45]; see also Jiang [44] for an earlier result in annular domains and, recently, Soufyane [108]. Koch [55] provided the results on the slow decay, motivated by the results of Henry [28]; extended results on the generic asymptotic behavior are given by Lebeau and Zuazua [58]. The presentation of the isotropic Cauchy problem follows Kawashima's thesis [50]. In [50] Kawashima obtained the decay rates of solutions only in the L^2-norm, while here we gave the L^p–L^q decay estimates using the interpolation theory and the Marcinkiewicz multiplier theorem; see also Dassios and Grillakis [16], Gawinecki [24], Muñoz Rivera [73] and Racke [89,91], the approach of the latter is based on the work of Zheng and Shen [126]. For inhomogeneous media or additional damping terms see Carvalho Pereira and Perla Menzala [5,6], and Racke [92], for contact problems see Muñoz Rivera and Racke [75].

2.4. *Nonlinear systems*

Global well-posedness results, in one or more space dimensions, are the subject of this section. Besides theorems on the global existence of smooth solutions for *small* data — in more than one space dimension for appropriate assumptions on the nonlinearities admitted —, we have a blow-up result for large data in one dimension, as well as a blow-up result for small data with certain nonlinearities in three space dimensions. The results mentioned up to now have also been discussed by Hsiao and Jiang [33] as an example of hyperbolic–parabolic systems in Chapter 4 of volume 1 of the *Handbook of Differential Equations*. Therefore we shall make this presentation short but include it for completeness. For details see [33] and in particular [47].

Moreover we shall give the result on the global well-posedness in one dimension for *large* data in the class of weak solutions.

We start with global well-posedness for smooth solutions in one space dimension.

The equations for the displacement u and the temperature difference $\theta = T - T_0$ are those given by (2.1), (2.3), where we shall assume, without loss of generality, that the medium is homogeneous and that the density ρ equals one. Writing f_1 and f_2 for the body force and the heat supply, respectively, we have

$$u_{tt} - \tilde{S}(u_x, \theta)_x = f_1 \quad \text{in } [0, \infty) \times \Omega, \tag{2.125}$$

$$(\theta + T_0)\eta(u_x, \theta)_t + q(u_x, \theta_x, \theta)_x = f_2 \quad \text{in } [0, \infty) \times \Omega. \tag{2.126}$$

Here, the reference configuration is represented by $\Omega := (0, 1)$. Introducing

$$a := \frac{\partial \tilde{S}}{\partial u_x}, \quad b := -\frac{\partial \tilde{S}}{\partial \theta}, \quad c := \frac{\partial \eta}{\partial \theta},$$

$$d := -\frac{1}{(\theta + T_0)}\frac{\partial q}{\partial \theta_x}, \quad g := \frac{f_2}{(\theta + T_0)} \quad (\text{for } |\theta| < T_0), \quad f := f_1,$$

observing $-\partial \tilde{S}/\partial \theta = \partial \eta / \partial u_x$ and assuming $q = q(\theta_x)$ for simplicity, we may rewrite the equations as follows:

$$u_{tt} - a(u_x, \theta)u_{xx} + b(u_x, \theta)\theta_x = f, \tag{2.127}$$

$$c(u_x, \theta)\theta_t + b(u_x, \theta)u_{tx} - d(\theta, \theta_x)\theta_{xx} = g. \tag{2.128}$$

Additionally we consider the following boundary and initial conditions, respectively:

$$u(t, 0) = u(t, 1) = \theta(t, 0) = \theta(t, 1) = 0 \quad \text{in } [0, \infty), \tag{2.129}$$

$$u(0, x) = u^0(x), \qquad u_t(0, x) = u^1(x), \qquad \theta(t = 0) = \theta^0(x) \quad \text{in } \Omega, \tag{2.130}$$

with prescribed data u^0, u^1 and θ^0.

The following assumption will be made.

ASSUMPTION 2.4.A. a, b, c, d are C^2-functions of their arguments. There exist positive constants a_0, c_0, d_0, K, with $K < T_0$, such that if $|u_x| \leq K$, $|\theta| \leq K$, $|\theta_x| \leq K$ we have

$$a(u_x, \theta) \geq a_0, \qquad c(u_x, \theta) \geq c_0, \qquad d(\theta, \theta_x) \geq d_0,$$
$$b(u_x, \theta) \neq 0.$$

f_1 and f_2 satisfy:

$$f_1, f_2 \in C^2([0, \infty), L^2(\Omega)) \cap C^1([0, \infty), H^1(\Omega)).$$

Since we are looking for the solution in a K-neighborhood of the origin, we can assume, without loss of generality, that the functions a, b, c, d and their derivatives are bounded.

Let

$$u^2 := u_{tt}(t = 0), \qquad \theta^1 := \theta_t(t = 0)$$

be given formally through the differential equation, explicitly in terms of the initial data u^0, u^1, θ^0:

$$u^2 = a(u_x^0, \theta^0)u_{xx}^0 + b(u_x^0, \theta^0)\theta_{xx}^0 + f(t = 0),$$
$$\theta^1 = \frac{1}{c(u_x^0, \theta^0)}\{d(\theta^0, \theta_x^0)\theta_{xx}^0 - b(u_x^0, \theta^0)u_x^1 + g(t = 0)\}.$$

ASSUMPTION 2.4.B. Suppose $u^0 \in H^3(\Omega)$, $u^1 \in H^2(\Omega)$, $u^2 \in H^1(\Omega)$, $\theta^0 \in H^3(\Omega)$, $\theta^1 \in H^2(\Omega)$ and $|\theta^0(x)| < T_0$ in $\overline{\Omega}$,

$$u^0 = u^1 = u^2 = 0 \quad \text{on } \partial\Omega,$$
$$\theta^0 = \theta^1 = 0 \quad \text{on } \partial\Omega.$$

Then we have

THEOREM 2.39. *Let*

$$\lambda(t) := \sum_{j=0}^{1} \|D^j(f_1, f_2)(t, \cdot)\|^2 + \|\partial_t^2(f_1, f_2)(t, \cdot)\|^2 + \|\partial_t\partial_x(f_1, f_2)(t, \cdot)\|^2$$

and suppose that Assumptions 2.4.A and 2.4.B *are satisfied. Then there exists a small constant* $\epsilon_0 > 0$ *such that if*

$$\|u^0\|_{2,2}^2 + \|u^1\|_{2,2}^2 + \|u^2\|_{1,2}^2 + \|\theta^0\|_{2,2}^2 + \|\theta^1\|_{2,2}^2 + \sup_{t\geq 0} \lambda(t) \leq \epsilon_0,$$

then the initial boundary value problem (2.125), (2.126), (2.129) and (2.130) *admits a unique global solution*

$$u \in \bigcap_{j=0}^{3} C^j([0, \infty), H^{3-j}(\Omega)), \quad \theta \in \bigcap_{j=0}^{1} C^j([0, \infty), H^{3-j}(\Omega)),$$
$$\theta \in C^2([0, \infty), L^2(\Omega)).$$

Moreover, there are constants $d_1, d_2 > 0$ *such that for* $t \geq 0$*:*

$$\sum_{j=0}^{3} \|D^j u(t)\|^2 + \sum_{j=0}^{2} \|D^j \theta(t)\|^2$$
$$\leq d_1 e^{-d_2 t} \left(\|u^0\|_{2,2}^2 + \|u^1\|_{2,2}^2 + \|u^2\|_{1,2}^2 + \|\theta^0\|_{2,2}^2 + \|\theta^1\|_{2,2}^2 + \int_0^t e^{d_2 r} \lambda(r) dr \right).$$

We remark that here and for subsequent existence theorems, corresponding local existence theorems are needed. This topic is extensively discussed in [47], for domains with boundaries and for the Cauchy problem, in all space dimensions.

Now we turn to the Cauchy problem in one dimension. We introduce the variables $w := u_x$, $v := u_t$ to write the system (2.125), (2.126) as a first-order system:

$$w_t = v_x, \tag{2.131}$$

$$v_t = a(w,\theta)w_x - b(w,\theta)\theta_x \quad (\Leftrightarrow v_t = \sigma(w,\theta)_x), \tag{2.132}$$

$$c(w,\theta)\theta_t = \frac{[\hat{\kappa}(w,\theta)\theta_x]_x}{(\theta + T_0)} - b(w,\theta)v_x \tag{2.133}$$

with the prescribed initial data

$$(w, v, \theta)(t = 0) = (w^0, v^0, \theta^0) \equiv V^0 \quad \text{in } \mathbb{R}, \tag{2.134}$$

where

$$a(w,\theta) := \frac{\partial \tilde{S}(w,\theta)}{\partial w}, \qquad b(w,\theta) := -\frac{\partial \tilde{S}(w,\theta)}{\partial \theta},$$
$$c(w,\theta) := \frac{\partial \eta(w,\theta)}{\partial \theta}, \qquad \sigma(w,\theta) := \tilde{S}(w,\theta),$$

and for simplicity, the Fourier law $q = -\hat{\kappa}(w,\theta)\theta_x$ for the heat flux, and $f_1 = f_2 = 0$ are assumed.

By exploiting some relations which are associated with the second law of thermodynamics and that embody the dissipative effect induced by thermal diffusion, the L^2-energy method can be used to establish the global existence and uniqueness of smooth solutions of (2.131)–(2.134) with smooth and small initial data.

We require (cp. Assumption 2.4.A)

ASSUMPTION 2.4.C. The functions a, b, c resp. $\hat{\kappa}$ are C^2-resp. C^3-functions of their arguments. There exist positive constants $\tilde{\gamma}_0$, $\tilde{\gamma}_1$ and K, with $K < T_0$, such that if $|w| \leq K$, $|\theta| \leq K$ we have

$$\tilde{\gamma}_0 \leq a(w,\theta), \qquad c(w,\theta), \qquad \hat{\kappa}(w,\theta) \leq \tilde{\gamma}_1, \qquad b(w,\theta) \neq 0.$$

Let (w, v, θ) be a solution of (2.131)–(2.134) on $[0, T]$ $(T > 0)$ satisfying

$$\left.\begin{array}{l} w, v \in C^0([0,T], H^2(\mathbb{R})) \cap C^1([0,T], H^1(\mathbb{R})), \\ \theta \in C^0([0,T], H^2(\mathbb{R})) \cap C^1([0,T], L^2(\mathbb{R})), \\ \quad \theta_t, \theta_{xx} \in L^2([0,T], H^1(\mathbb{R})), \end{array}\right\} \tag{2.135}$$

$$\|(w(t), v(t), \theta(t))\|_\infty \leq K, \quad t \in [0, T]. \tag{2.136}$$

Then we have

THEOREM 2.40. *Let $V^0 \in H^2(\mathbb{R})$ and Assumption 2.4.C be satisfied. Then there exists an $\epsilon > 0$ such that if $\|V^0\|_{2,2} \le \epsilon$, the Cauchy problem (2.131)–(2.134) has a unique global solution $V = (w, v, \theta)$ satisfying*

$$\begin{aligned}
&w, v \in C^0([0,\infty), H^2(\mathbb{R})) \cap C^1([0,\infty), H^1(\mathbb{R})),\\
&\theta \in C^0([0,\infty), H^2(\mathbb{R})) \cap C^1([0,\infty), L^2(\mathbb{R})),\\
&\theta_t, \theta_{xx} \in L^2((0,\infty), H^1(\mathbb{R})),\\
&\|V(t)\|_\infty \le K < T_0 \quad \textit{for all } t \ge 0.
\end{aligned}$$

For large data, smooth solutions can blow-up in finite time, as the following result shows. This is discussed in [33,47] as for the previous results. For this purpose we consider special constitutive equations of the form

$$\left.\begin{aligned}
&\tilde{S}(w,\theta) := p(w) + \tilde{\gamma}\theta,\\
&\varepsilon(w,\theta) := P(w) + \tilde{\delta}\theta - \tilde{\gamma} T_0 w,\\
&q(w,\theta,\theta_x) := -\kappa\theta_x,
\end{aligned}\right\} \tag{2.137}$$

where $T_0, \kappa, \tilde{\delta} > 0$, $\tilde{\gamma} \neq 0$ are constants, T_0 is the reference temperature; $p : (-1,\infty) \longrightarrow \mathbb{R}$ is a given function,

$$P(w) := \int_0^w p(\xi)\mathrm{d}\xi, \quad w > -1. \tag{2.138}$$

It is easy to see that these constitutive equations satisfy $\varepsilon_w = \tilde{S} - (\theta + T_0)\tilde{S}_\theta$ and $g\,q(w,\theta,g) \le 0$, and that

$$\psi(w,\theta) = P(w) + \tilde{\gamma} w\theta + \tilde{\delta}(\theta + T_0)\log\frac{T_0}{(\theta + T_0)} + \tilde{\delta}\theta \tag{2.139}$$

is a Helmholtz free energy; moreover, the corresponding entropy is given by

$$\eta(w,\theta) = -\tilde{\gamma} w - \tilde{\delta}\log\frac{T_0}{(\theta + T_0)}.$$

Inserting the relations (2.137)–(2.139) into the thermoelastic system (2.131)–(2.133), we obtain

$$\left.\begin{aligned}
&w_t = v_x,\\
&v_t = p'(w)w_x + \tilde{\gamma}\theta_x,\\
&\tilde{\delta}\theta_t = \kappa\theta_{xx} + \tilde{\gamma}(\theta + T_0)v_x, \quad x \in \mathbb{R}, t > 0.
\end{aligned}\right\} \tag{2.140}$$

We shall consider the Cauchy problem for (2.140) with initial conditions

$$(w(0,x), v(0,x), \theta(0,x)) = (w^0(x), v^0(x), \theta^0(x)) \equiv V^0(x), \quad x \in \mathbb{R}. \tag{2.141}$$

For (2.140) we further assume that

$$p \in C^4(-1,\infty), \qquad p(0)=0, \qquad p'(\xi)>0, \quad \xi>-1. \tag{2.142}$$

Let $V^0 \in H^3(\mathbb{R})$ and $\|w^0\|_\infty + \|\theta^0\|_\infty$ be small, then, under (2.142), (2.140), (2.141) has a unique solution $V=(w,v,\theta)$ on $[0,T]$ $(T>0)$ with $w, v \in C^0([0,T],H^3(\mathbb{R})) \cap C^1([0,T],H^2(\mathbb{R}))$, $\theta \in C^0([0,T],H^3(\mathbb{R})) \cap C^1([0,T],H^1(\mathbb{R}))$. We have the following blow-up theorem.

THEOREM 2.41. *Assume that (2.142) holds and $p''(0)>0$. Let β, L, $J>0$ be given. Then there exist ϵ, $M>0$ (depending on β, L, J) with the following property: For each w^0, v^0, $\theta^0 \in H^3(\mathbb{R})$,*

$$\left.\begin{aligned} &|w^0(x)|, |v^0(x)|, |\theta^0(x)| \le \epsilon, |\partial_x\theta^0(x)| \le J \quad \text{for } x \in \mathbb{R},\\ &\|(w^0,v^0,\theta^0)\|^2 + \|\theta^0\|_1 \le \epsilon^2,\\ &\min_{x\in\mathbb{R}}\left\{\partial_x v^0 + p'(w_0)^{1/2}\partial_x w^0\right\}(x)\\ &\quad + \min_{x\in\mathbb{R}}\left\{\partial_x v^0 - p'(w_0)^{1/2}\partial_x w^0\right\} \ge -J \end{aligned}\right\} \tag{2.143}$$

and

$$\begin{aligned} &\max_{x\in\mathbb{R}}\left\{\partial_x v^0 + p'(w_0)^{1/2}\partial_x w^0\right\}(x)\\ &\quad + \max_{x\in\mathbb{R}}\left\{\partial_x v^0 - p'(w_0)^{1/2}\partial_x w^0\right\}(x) \ge M, \end{aligned} \tag{2.144}$$

the length T_m of the maximal interval of existence of a smooth solution (w,v,θ) of (2.140), (2.141) is less than (or equal to) L; moreover, the local solution satisfies

$$|w(t,x)|, |v(t,x)|, |\theta(x,t)| \le \beta, \quad x \in \mathbb{R}, t \in [0,T_m).$$

This blow-up result proves the non-existence of *smooth* (strong, classical) solutions. *Weak* solutions can also exist globally for large data as we shall see in more detail next.

The nonlinear equations, which can be dealt with here, arise from the differential equations (2.125), (2.126) under the simplifying assumption that $\tilde{S}$ satisfies

$$\tilde{S}(u_x,\theta) = \sigma(u_x) + \theta$$

for a function σ, and that the equation for θ in (2.126) is linear. Using the notation given in (2.12), (2.13) we thus consider the Cauchy problem

$$u_{tt} - \sigma(u_x)_x + \tilde{\gamma}\theta_x = 0, \tag{2.145}$$

$$\tilde{\delta}\theta_t - \kappa\theta_{xx} + \tilde{\gamma}u_{tx} = 0, \tag{2.146}$$

$$u(t=0)=u^0, \qquad u_t(t=0)=u^1, \qquad \theta(t=0)=\theta^0 \tag{2.147}$$

for $(t,x)\in[0,\infty)\times\mathbb{R}^1$.

It was shown in the blow-up results that second derivatives of u may develop singularities in finite time if the data are large. This is the typical hyperbolic effect, and Eq. (2.145) keeps the essential nonlinearity in the term $\sigma(u_x)_x$. Now the existence of weak solutions will be proved yielding u, the first-order derivatives of which are in $L^\infty(((0,\infty)\cap K)\times\mathbb{R}^1)$ for all $K\subset\subset\mathbb{R}^1$.

The first idea is to compute θ in terms of u from the linear equation (2.146),

$$\theta=F(u_{tx},\theta^0),$$

and to insert this representation into (2.145) leading to a (perturbed) nonlinear hyperbolic wave equation for u. For the necessary *a priori* estimates the term

$$\tilde{\gamma}\theta_x=\tilde{\gamma}(F(u_{tx},\theta^0))_x$$

causes difficulties. Introducing a new variable – a new dependent function $\tilde{w}$ instead of $w=u_t$, see below – will overcome this problem. To return from $\tilde{w}$ to w, the time interval $\mathbb{R}$ is divided into sufficiently small pieces.

Introducing the transformation

$$v:=u_x, \qquad w:=u_t,$$

Eqs. (2.145)–(2.147) turn into the following first-order system in t:

$$v_t-w_x=0, \tag{2.148}$$
$$w_t-\sigma(v)_x=-\tilde{\gamma}\theta_x, \tag{2.149}$$
$$\tilde{\delta}\theta_t-\kappa\theta_{xx}=-\tilde{\gamma}w_x, \tag{2.150}$$
$$v(t=0)=u_x^0=:v^0, \qquad w(t=0)=u^1=:w^0, \qquad \theta(t=0)=\theta^0. \tag{2.151}$$

The Cauchy problem (2.148)–(2.151) will be solved by the method of vanishing viscosity and with compensated compactness arguments known from the purely hyperbolic theory. The usual invariant regions needed for L^∞-estimates will have to be replaced by controllable expanding regions.

The assumption on the tensor σ will be that $\sigma\in C^2(\mathbb{R},\mathbb{R})$ and that for all $s\in\mathbb{R}$

$$\begin{aligned}&\sigma'(s)>0, \qquad \sigma''(s)\cdot s>0 \quad (s\neq 0),\\ &\sqrt{\sigma'(s)}\leq c_1\left(1+\left|\int_0^s\sqrt{\sigma'(z)}\mathrm{d}z\right|\right),\end{aligned} \tag{2.152}$$

with a fixed positive constant c_1. The first condition assures the hyperbolicity of the system in (2.149), the second assumption is a restriction but is satisfied e.g. for rubber-like material (cf. [17]). The last assumption as a growth condition is satisfied e.g. for polynomially growing functions. For $n \in \mathbb{N}$

$$\sigma(s) := \frac{1}{2n+1} s^{2n+1} + s$$

defines a function σ that satisfies all assumptions.

In order to define admissible weak solutions, we need the concept of entropy–entropy-flux pairs from the purely hyperbolic theory. A pair (π, p) with $\pi \in C^2(\mathbb{R}^2, \mathbb{R})$, $p \in C^1(\mathbb{R}^2, \mathbb{R})$ is called an "entropy–entropy-flux pair corresponding to (2.148), (2.149)" if

$$\forall v, w \in \mathbb{R} : p_v(v, w) = -\sigma'(v)\pi_w(v, w), \quad p_w(v, w) = -\pi_v(v, u).$$

If, additionally, π is convex, it is called a "convex entropy–entropy-flux pair".

DEFINITION 2.42. *Let $v^0, w^0, \theta^0 \in L^\infty(\mathbb{R}^1)$. Then (v, w, θ) is called a weak solution to (2.148)–(2.151) if*

$$v, w, \theta \in L^\infty((0, \infty) \cap K) \times \mathbb{R}^1$$

for any $K \subset\subset \mathbb{R}$, and if for all $\varphi \in C_0^\infty(\mathbb{R}^2, \mathbb{R})$:

$$-\int_0^\infty \int_{\mathbb{R}} (\varphi_t(t, x)v(t, x) - \varphi_x(t, x)w(t, x))\mathrm{dx}\,\mathrm{dt} = \int_{\mathbb{R}} \varphi(0, x)v^0(x)\mathrm{dx},$$

$$-\int_0^\infty \int_{\mathbb{R}} (\varphi_t(t, x)w(t, x) - \varphi_x(t, x)\sigma(v(t, x)) + \tilde{\gamma}\varphi_x(t, x)\theta(t, x))\mathrm{dx}\,\mathrm{dt} = \int_{\mathbb{R}} \varphi(0, x)w^0(x)\mathrm{dx},$$

$$-\int_0^\infty \int_{\mathbb{R}} (\tilde{\delta}\varphi_t(t, x)\theta(t, x) + \kappa\varphi_{xx}(t, x)\theta(t, x) + \tilde{\gamma}\varphi_x(t, x)w(t, x))\mathrm{dx}\,\mathrm{dt} = \int_{\mathbb{R}} \tilde{\delta}\varphi(0, x)\theta^0(x)\mathrm{dx}.$$

A weak solution (v, w, θ) to (2.148)–(2.151) is called admissible, if θ_x exists in $L^2(K \times \mathbb{R})$ for each compact $K \subset [0, \infty)$, and if for each convex entropy–entropy-flux pair (π, p) to (2.148), (2.149) and for each non-negative $\varphi \in C_0^\infty((0, \infty) \times \mathbb{R})$:

$$-\int_0^\infty \int_{\mathbb{R}} (\varphi_t \pi(v, w) + \varphi_x p(v, w))(t, x)\mathrm{dx}\,\mathrm{dt} \le -\int_0^\infty \int_{\mathbb{R}} \tilde{\gamma}(\varphi\pi_w(v, w)\theta_x)(t, x)\mathrm{dx}\,\mathrm{dt}.$$

The existence theorem – uniqueness is open as for many purely hyperbolic problems – now reads as follows:

THEOREM 2.43. *Let $w^0, \theta^0 \in L^\infty(\mathbb{R}) \cap L^2(\mathbb{R})$, $v^0 \in L^\infty(\mathbb{R})$, $(v^0 - \xi_0) \in L^2(\mathbb{R})$ (for $\sigma(\xi_0) = 0$) and let σ satisfy (2.152). Then there exists an admissible weak solution to (2.148)–(2.151).*

We sketch the

PROOF OF THEOREM 2.43. Step 1. Parabolic approximation.

Let $\tilde{\delta} = 1$, without loss of generality, and let ξ_0 be the zero of σ. There is a family $(v_\epsilon^0, w_\epsilon^0, \theta_\epsilon^0)_{\epsilon>0} \subset C^\infty(\mathbb{R})$ with bounded derivatives such that

$$(v_\epsilon^0 - \xi_0, w_\epsilon^0, \theta_\epsilon^0) \to (v^0 - \xi_0, w^0, \theta^0) \quad \text{in } L^2(\mathbb{R}) \quad \text{as } \epsilon \downarrow 0,$$

$v_\epsilon^0 - \xi_0, w_\epsilon^0, \theta_\epsilon^0$ are bounded in $L^\infty(\mathbb{R})$ and $L^2(\mathbb{R})$ by the corresponding norms of $v^0 - \xi_0$, w^0 and θ^0, respectively, and $\lim_{x\to\pm\infty}(v_\epsilon^0, w_\epsilon^0, \theta_\epsilon^0) = (\xi^0, 0, 0)$.

Let $g^\kappa = g^\kappa(x - y, t - \alpha)$ for $x, y \in \mathbb{R}, t, \alpha \in [0, \infty), t > \alpha$, denote the fundamental solution to

$$z_t - \kappa z_{xx} = 0.$$

For $n \in \mathbb{N}_0, T > 0$ let

$$T_n := (nT, (n+1)T),$$
$$G_n := T_n \times \mathbb{R}, \qquad G_n(t) := (nT, t) \times \mathbb{R}$$

and define

$$K_n : L^\infty(G_n) \longrightarrow L^\infty(G_n), \qquad w \mapsto K_n w$$

by

$$(K_n w)(t, x) := \int_{nT}^t \int_{\mathbb{R}} g^\kappa(x - y, t - s) w(s, y) \mathrm{d}y \, \mathrm{d}s,$$

moreover let

$$K_{n,x} w := K_n w_x.$$

Using well-known estimates for the fundamental solution g^κ (see [57]), we obtain

$$\|K_n w(t)\|_\infty \le C T^{1/4} \|w\|_{L^2(G_n)}, \quad t \in T_n, \tag{2.153}$$

$$\left\| \frac{\partial}{\partial x} K_n w(t) \right\|_\infty \le C T^{1/4} \sup_{\alpha \in [nT, t]} \|w(\alpha)\|_2, \quad t \in T_n, \tag{2.154}$$

where C denotes a constant which is independent of n, T, and t. Moreover, for any $\nu_0 > 0$ there is $T > 0$ such that for all $t \in \overline{T_n}$:

$$\|K_n\|_B < \nu_0, \qquad \left\| \frac{\partial}{\partial x} K_n \right\|_B \le C, \tag{2.155}$$

where B is one of the spaces

$$L^\infty(G_n(t)),\, C^0([nT,t], L^2(\mathbb{R})) \quad \text{or} \quad L^2(G_n(t)),$$

and

$$\|K_n w\|_{H^{(\varrho/2,\varrho)}(G_n)} \leq C\|w\|_{L^\infty(G_n)}, \tag{2.156}$$

where, for $\varrho \in (0,1)$ arbitrary, but fixed, $H^{(\varrho/2,\varrho)}(G_n)$ denotes the Hölder space of continuous, bounded functions on G_n, for which the norm

$$\|w\|_{H^{(\varrho/2,\varrho)}(G_n)} := \|w\|_{L^\infty(G_n)} + \sup_{x;t\neq t'} \frac{|w(t,x)-w(t',x)|}{|t-t'|^{\varrho/2}} + \sup_{t;x\neq x'} \frac{|w(t,x)-w(t,x')|}{|x-x'|^{\varrho}}$$

is bounded.

The initial value problem (2.148)–(2.151) will now be divided into a sequence of initial value problems with time step size $T > 0$, which are approximated by adding so-called viscosity terms and using the smoothed initial data. That is, we consider for $n \in \mathbb{N}_0$ the following sequence of problems (2.157)–(2.160):

$$v_t^{n,\epsilon} - w_x^{n,\epsilon} = \epsilon v_{xx}^{n,\epsilon} \quad \text{in } G_n, \tag{2.157}$$

$$w_t^{n,\epsilon} - \alpha(v^{n,\epsilon})_x = -\tilde{\gamma}\theta_x^{n,\epsilon} + \epsilon w_{xx}^{n,\epsilon} - \epsilon\frac{\partial^2}{\partial x^2}(\tilde{\gamma} K_n(\tilde{\gamma} w^{n,\epsilon})) \quad \text{in } G_n, \tag{2.158}$$

$$\theta_t^{n,\epsilon} - \kappa\theta_{xx}^{n,\epsilon} = \tilde{\gamma} w_x^{n,\epsilon} \quad \text{in } G_n, \tag{2.159}$$

$$\left.\begin{array}{l} v^{n,\epsilon}(t=nT) = v^{n-1,\epsilon}(nT), \qquad w^{n,\epsilon}(t=nT) = w^{n-1,\epsilon}(nT), \\ \theta^{n,\epsilon}(t=nT) = \theta^{n-1,\epsilon}(nT) \quad \text{in } \mathbb{R}, \end{array}\right\} \tag{2.160}$$

with

$$v^{-1,\epsilon}(t,x) := v_\epsilon^0(x), \qquad w^{-1,\epsilon}(t,x) := w_\epsilon^0(x), \qquad \theta^{-1,\epsilon}(t,x) := \theta_\epsilon^0(x).$$

The last term in (2.158) is chosen anticipating the transformation in Step 4 below.

By linearizing in (2.158) and using a classical iteration procedure (cf. [11] and [78]), a classical solution to (2.157)–(2.160) is obtained ($\epsilon > 0$ fixed):

LEMMA 2.44. *Let $n \in \mathbb{N}_0$ and $(v^{n-1,\epsilon}, w^{n-1,\epsilon}, \theta^{n-1,\epsilon}) \in B_{n-1}^{\varrho}$. Then there is a unique solution $(v^{n,\epsilon}, w^{n,\epsilon}, \theta^{n,\epsilon})$ to (2.157)–(2.160).*

Here

$$B_n^\varrho := B_n^\varrho(\xi_0) \times B_n^\varrho(0) \times B_n^\varrho(0),$$

where

$$B_n^\varrho(z) := \{v \in H^{(\varrho/2+1,\varrho+2)}(G_n) \mid v - z \in L^\infty(T_n, L^2(\mathbb{R})),\, v_x \in L^2(G_n)\},$$

and $H^{\varrho/2+1,\varrho+2}(G_n)$ denotes the Hölder space with norm

$$\|w\|_{H^{(\varrho/2+1,\varrho+2)}(G_n)} := \sum_{j=0}^{1} \|\partial_t^j w\|_{L^\infty(G_n)} + \sum_{j=1}^{2} \|\partial_x^j w\|_{L^\infty(G_n)} + \sum_{0<2+\varrho-2r-s<2} \sup_{x;t\neq t'} \frac{|\partial_t^r \partial_x^s w(t,x) - \partial_t^r \partial_x^s w(t',x)|}{|t-t'|^{(2+\varrho-2r-s)/2}} + \sum_{2r+s=2} \sup_{t,x\neq x'} \frac{|\partial_t^r \partial_x^s w(t,x) - \partial_t^r \partial_x^s w(t,x')|}{|x-x'|^{\varrho}}.$$

Step 2. L^2-estimate.
Multiplying (2.158) by $w^{n,\epsilon}$ and (2.159) by $\theta^{n,\epsilon}$, integrating over $(nT,t)\times(-m,m)$ for $m\in\mathbb{N}$, integrating by parts, letting $m\to\infty$, Gronwall's inequality and induction over n yields the following *a priori* estimate, independent of ϵ:

LEMMA 2.45. *The solution $(v^{n,\epsilon}, w^{n,\epsilon}, \theta^{n,\epsilon})$ given in Lemma 2.44 satisfies:*

$$\begin{aligned}
&\exists C = C(n) > 0 \quad \forall \epsilon > 0 \quad \forall t \in T_n: \\
&\int_{\mathbb{R}} \left\{ |w^{n,\epsilon}|^2 + |\theta^{n,\epsilon}|^2 + \int_{\xi_0}^{v^{n,\epsilon}} \sigma(s)\mathrm{d}s \right\}(t,x)\mathrm{d}x \\
&+ \int_{nT}^{t} \int_{\mathbb{R}} \left\{ |\theta_x^{n,\epsilon}|^2 + \epsilon\left(|w_x^{n,\epsilon}|^2 + \sigma'(v^{n,\epsilon})|v_x^{n,\epsilon}|^2 \right) \right\}(s,x)\mathrm{d}x\,\mathrm{d}s \\
&\le C\left(\int_{\mathbb{R}} \left\{ |u_\epsilon^0|^2 + |\theta_\epsilon^0|^2 + \int_{\xi_0}^{v_\epsilon^0} \sigma(s)\mathrm{d}s \right\}(x)\mathrm{d}x + 1 \right).
\end{aligned}$$

Here $T>0$ is still arbitrary but fixed, and will be chosen once and forever in the next step.
Step 3. A transformation, fixing T.
Solving the equation for $\theta^{n,\epsilon}$ in (2.159) in terms of $w_x^{n,\epsilon}$ and the data, would lead to a system (2.157), (2.158) for $(v^{n,\epsilon}, w^{n,\epsilon})$, but with a difficult term $\frac{\partial}{\partial x^2} K_n(\tilde{\gamma} w^{n,\epsilon})$. The transformation

$$w^{n,\epsilon} \mapsto \tilde{w}^{n,\epsilon} := \underbrace{\left(\mathrm{Id} - \frac{\tilde{\gamma}}{\alpha} K_n(\tilde{\gamma}\,\cdot)\right)}_{=:\Phi_n} w^{n,\epsilon} \quad \text{with } \alpha = \frac{\kappa}{\tilde{\delta}}$$

solves this problem since

$$\Phi_n : L^\infty(G_n) \longrightarrow L^\infty(G_n)$$

is a homeomorphism by (2.155), (2.156), choosing $T>0$ (once) sufficiently small, and to solve (2.157)–(2.160) for $(v^{n,\epsilon}, w^{n,\epsilon}, \theta^{n,\epsilon})$ is equivalent to solving the following system for $(v^{n,\epsilon}, \tilde{w}^{n,\epsilon}, \theta^{n,\epsilon})$:

$$v_t^{n,\epsilon} - \tilde{w}_x^{n,\epsilon} = \epsilon v_{xx}^{n,\epsilon} - \frac{\tilde{\gamma}}{\alpha} K_{n,x}(\tilde{\gamma}\psi_n \tilde{w}^{n,\epsilon}) \quad \text{in } G_n, \tag{2.161}$$

$$\tilde{w}_t^{n,\epsilon} - \sigma(v^{n,\epsilon})_x = \epsilon \tilde{w}_{xx}^{n,\epsilon} - \frac{\tilde{\gamma}^2}{\alpha}\psi_n \tilde{w}^{n,\epsilon}$$
$$- \tilde{\gamma}\frac{\partial}{\partial x}\int_{\mathbb{R}} g^\kappa(\cdot - y, \cdot - nT)\theta^{n-1,\epsilon}(nT, y)\mathrm{dy} \quad \text{in } G_n, \tag{2.162}$$
$$\theta^{n,\epsilon} = -K_{n,x}(\tilde{\gamma}\psi_n \tilde{w}^{n,\epsilon})$$
$$+ \int_{\mathbb{R}} g^\kappa(\cdot - y, \cdot - nT)\theta^{n-1,\epsilon}(nT, y)\mathrm{dy} \quad \text{in } G_n, \tag{2.163}$$
$$\left.\begin{array}{ll} v^{n,\epsilon}(t = nT) = v^{n-1,\epsilon}(nT), & \\ \tilde{w}^{n,\epsilon}(t = nT) = \tilde{w}^{n-1,\epsilon}(nT) := \Phi_n w^{n-1,\epsilon}(nT), & \\ \theta^{n,\epsilon}(z = nT) = \theta^{n-1,\epsilon}(nT) & \text{in } \mathbb{R}, \end{array}\right\} \tag{2.164}$$

with

$$\psi_n := \Phi_n^{-1}$$

bounded as mapping

$$\psi_n : B \longrightarrow B,$$

where B can be one of the spaces

$$L^2(G_n(t)), \qquad C^0([nT, t], L^2(\mathbb{R})), \qquad H^{(\varrho/2,\varrho)}(G_n),$$
$$H^{(\varrho/2+1,\varrho+2)}(G_n), \qquad B_n^\varrho(0).$$

The L^2-estimate from Lemma 2.45 carries over to an L^2-estimate for $(v^{n,\epsilon}, \tilde{w}^{n,\epsilon}, \theta^{n,\epsilon})$:

LEMMA 2.46. *The solution* $(v^{n,\epsilon}, \tilde{w}^{n,\epsilon}, \theta^{n,\epsilon}) \in B_n^\varrho$ *to* (2.161)–(2.164) *satisfies:*

$$\exists c = c(n) > 0 \quad \forall \epsilon > 0 \quad \forall t \in T_n :$$
$$\int_{\mathbb{R}} (|v^{n,\epsilon} - \xi_0|^2 + |\tilde{w}^{n,\epsilon}|^2 + |\theta^{n,\epsilon}|^2)(t, x)\mathrm{dx}$$
$$+ \int_{nT}^t \int_{\mathbb{R}} \{|\theta_x^{n,\epsilon}|^2 + \epsilon(|\tilde{w}_x^{n,\epsilon}|^2 + |v_x^{n,\epsilon}|^2)(\alpha, x)\}\mathrm{dx}\,\mathrm{dt} \le c.$$

Step 4. L^∞-estimate.
In order to get an *a priori* estimate for $(v^{n,\epsilon}, \tilde{w}^{n,\epsilon}, \theta^{n,\epsilon})$ in $L^\infty(G_n)$, independent of ϵ, the Riemann invariants

$$r_\pm(v^{n,\epsilon}, \tilde{w}^{n,\epsilon}) := \tilde{w}^{n,\epsilon} \pm \int_0^{v^{n,\epsilon}} \sqrt{\sigma'(s)}\,\mathrm{ds}$$

are used as in [11]. In contrast to the situation there, due to a lack of sign conditions in certain terms arising in the equations for $r_\pm$, the existence of invariant regions cannot be proved. Instead, we obtain a growth in time, which nevertheless can still be controlled.

For $(t_0, x_0) \in G_n$ with $\frac{\partial}{\partial x} r_\pm(v^{n,\epsilon}(t_0, x), \tilde{w}^{n,\epsilon}(t_0, x))_{|x=x_0} = 0$ one has

$$\begin{aligned}
\frac{\partial}{\partial t} r_\pm(v^{n,\epsilon}(t, x_0), \tilde{w}^{n,\epsilon}(t, x_0))_{|t=t_0} &= \Bigg\{ \frac{\partial^2}{\partial x^2} r_\pm(v^{n,\epsilon}, \tilde{w}^{n,\epsilon}) \\
&\mp \epsilon \frac{\sigma''(v^{n,\epsilon})}{2\sqrt{\sigma'(v^{n,\epsilon})}} (v_x^{n,\epsilon})^2 - \tilde{\gamma} \frac{\partial}{\partial x} \int_{\mathbb{R}} g^\kappa(\cdot - y, \cdot - nT) \theta^{n-1,\epsilon}(nT, y) \mathrm{dy} \Bigg\} (t_0, x_o) \\
&\underbrace{- \left(\frac{\tilde{\gamma}^2}{\kappa} \psi_n \tilde{w}^{n,\epsilon} \pm \sqrt{\sigma'(v^{n,\epsilon})} \frac{\tilde{\gamma}}{\alpha} K_{n,x}(\tilde{\gamma} \psi_n \tilde{w}^{n,\epsilon}) \right) (t_0, x_0)}_{=:R_n(t_0, x_0)},
\end{aligned} \tag{2.165}$$

and for $(t, x) \in G_n$

$$|R_n(t, x)| \leq c_2 \left(1 + |\tilde{w}^{n,\epsilon}(t, x)| + \left| \int_0^{v^{n,\epsilon}(t,x)} \sqrt{\sigma'(s)} \mathrm{ds} \right| \right), \tag{2.166}$$

where $c_2 = c_2(n)$ does not depend on ϵ or (t, x).

For the last estimate one has to use (2.153), (2.154), Lemma 2.45 and the Assumption (2.152) on σ.

LEMMA 2.47. *With c_2 given in (2.166), $n \in \mathbb{N}_0$, $t \in \overline{T}_n$ let*

$$\begin{aligned}
z_n(t) := 2\mathrm{e}^{2c_2(t-nT)} \Bigg\{ & 1 + \max \left\{ \| r_\pm(v^{n-1,\epsilon}(nT), \tilde{w}^{n-1,\epsilon}(nT)) \|_\infty \right\} \\
& + |\tilde{\gamma}| \int_{nT}^t \left\| \int_{\mathbb{R}} g_x^\kappa(\cdot - y, s - nT) \theta^{n-1,\epsilon}(nT, y) \right\|_\infty \mathrm{ds} \Bigg\}.
\end{aligned}$$

Then

$$\forall t \in \overline{T}_n : 1 + \max \left\{ \left\| r_\pm(v^{n,\epsilon}(t), \tilde{w}^{n,\epsilon}(t)) \right\|_\infty \right\} \leq z_n(t) \tag{2.167}$$

holds.

This essential estimate for the Riemann invariants can be proved by looking at the largest $t_0 \in T_n$ for which (2.167) holds, and by leading the assumption that $t_0 < (n+1)T$ to a contradiction. To do so, the four regions in x where $1 \pm r_\pm(v^{n,\epsilon}(t_0, x), \tilde{w}^{n,\epsilon}(t_0, x)) = z_n(t_0)$ are considered and (2.165), (2.166) are exploited.

Using

$$\left| \int_0^{v^{n,\epsilon}(t,x)} \sqrt{\sigma'(s)} \mathrm{ds} \right| \geq c |v^{n,\epsilon}(t, x)|$$

the desired *a priori* L^∞-estimate follows:

LEMMA 2.48. $\forall n \exists c = c(n) > 0 \forall \epsilon > 0 : \|v^{n,\epsilon}\|_{L^\infty(G_n)} + \|\tilde{w}^{n,\epsilon}\|_{L^\infty(G_n)} + \|\theta^{n,\epsilon}\|_{L^\infty(G_n)} \leq c$.

Step 5. Limit $\epsilon \downarrow 0$.
By Lemma 2.48 there exists a sequence $(\epsilon_m)_m$, $\epsilon_m \downarrow 0$ as $m \to \infty$ such that

$$(v^{n,\epsilon_m}, \tilde{w}^{n,\epsilon_m}, \theta^{n,\epsilon_m}) \rightharpoonup (v^n, \tilde{w}^n, \theta^n), \quad \text{as } m \to \infty,$$

weak-$*$ in $L^\infty(G_n)$ for some $(v^n, \tilde{w}^n, \theta^n)$. The question arises as to whether $(\sigma(v^{n,\epsilon_m}))_m$ also converges properly, in order to finally conclude that $(v^n, \tilde{w}^n, \theta^n)$ is the weak solution (in G_n) that we are looking for. Here the tools from the theory of compensated compactness (cf. [17,76,113]) are applicable and the pointwise convergence of $(v^{n,\epsilon_m}, \tilde{w}^{n,\epsilon_m})$ (and then of θ^{n,ϵ_m}) almost everywhere, follows from Lemma 2.49 (cf. [17]).

LEMMA 2.49. *Let $n \in \mathbb{N}_0$. Then for all entropy–entropy-flux pairs (η, q) associated to (2.161), (2.162)*

$$\left(\frac{\partial}{\partial t}\eta(v^{n,\epsilon}, \tilde{w}^{n,\epsilon}) + \frac{\partial}{\partial x}q(v^{n,\epsilon}, \tilde{w}^{n,\epsilon})\right)_\epsilon$$

belongs to a compact subset of $H^{-1}_{\text{loc}}(G_n)$ (H^{-1}: dual space to H^1_0).

Using results from [76,113], it is sufficient for the proof of Lemma 2.49 to show

(i) $\left(\frac{\partial}{\partial t}\eta(v^{n,\epsilon}, \tilde{w}^{n,\epsilon}) + \frac{\partial}{\partial x}q(v^{n,\epsilon}, \tilde{w}^{n,\epsilon})\right)_\epsilon$ belongs to a bounded subset of $W^{-1,\infty}(G_{n,m})$, $m \in \mathbb{N}$, and
(ii) $\frac{\partial}{\partial t}\eta(v^{n,\epsilon}, \tilde{w}^{n,\epsilon}) + \frac{\partial}{\partial x}q(v^{n,\epsilon}, \tilde{w}^{n,\epsilon}) = g_1^\epsilon + g_2^\epsilon$, such that $(g_1^\epsilon)_\epsilon$ belongs to a compact subset of $H^{-1}(G_{n,m})$, and $(g_2^\epsilon)_\epsilon$ is bounded in $L^1(G_{n,m})$,

where $G_{n,m} := T_n \times (-m, m)$.

The pointwise convergence (almost everywhere) of $\tilde{w}^{n,\epsilon}$ carries over to that of $w^{n,\epsilon}$ to some w^n.

Putting together the pieces (v^n, w^n, θ^n) given on G_n,

$$v := \sum_{n=0}^{\infty} \chi_n v^n, \qquad w := \sum_{n=0}^{\infty} \chi_n w^n, \qquad \theta := \sum_{n=0}^{\infty} \chi_n \theta^n,$$

where χ_n denotes the characteristic function of T_n, (v, w, θ) is the admissible weak solution to (2.148)–(2.151), the existence of which was claimed in Theorem 2.43. The properties "weak solution" and "admissibility" follow from the weak form of the equations for $(v^{n,\epsilon_m}, w^{n,\epsilon_m}, \theta^{n,\epsilon_m})$ and the limit $\epsilon_m \downarrow 0$. □

In [20], the basis for the discussion above, more general Cauchy problems – coefficients depending on x, nonlinear right-hand sides – are considered as well as a class of initial boundary value problems.

We restrict ourselves to three space dimensions. The relevant equations – first for a general situation, while we shall turn to the initially isotropic case below – are (2.168), (2.169), i.e., the differential equations (2.1), (2.4) with $\rho = \text{Id}$ in the form:

$$\partial_t^2 U - \nabla' \tilde{S}(\nabla U, \theta) = f, \tag{2.168}$$

$$(\theta + T_0)\left\{a(\nabla U, \theta)\frac{\partial\theta}{\partial t} - \mathrm{tr}\left[\left(\frac{\partial \tilde{S}(\nabla U, \theta)}{\partial\theta}\right)' \cdot (\partial_t \nabla U)\right]\right\}$$
$$+ \nabla' q(\nabla U, \theta, \nabla\theta) = g, \tag{2.169}$$

where $U = (U_1, U_2, U_3)'$ denotes the displacement vector again, and θ represents the temperature difference. We assume throughout, the following assumptions:

ASSUMPTION 2.4.D. 1. $\psi \in C^{s+1}(\mathbb{R}^{3\times 3} \times \mathbb{R})$, $q \in C^s(\mathbb{R}^{3\times 3} \times \mathbb{R} \times \mathbb{R}^3)$.

2.

$$\frac{\partial q_i(P, \mu, \nu)}{\partial \nu_j} = \frac{\partial q_j(P, \mu, \nu)}{\partial \nu_i},$$
$$P = \{P_{ij}\} \in \mathbb{R}^{3\times 3}, \nu = \{\nu_i\} \in \mathbb{R}^3, \mu \in \mathbb{R}.$$

3. There are constants $\kappa_0, \kappa_1 > 0$ such that

$$C_{ijkl}(P, \mu)\xi_j\xi_l\eta_i\eta_k \geq \kappa_0|\xi|^2|\eta|^2, \qquad -\frac{\partial q_i(P, \mu, \nu)}{\partial \nu_j}\xi_i\xi_j \geq \kappa_0|\xi|^2;$$
$$\kappa_0 \leq a(P, \mu) \leq \kappa_1, \qquad P \in \mathbb{R}^{3\times 3}, \quad \xi = \{\xi_i\},$$
$$\eta = \{\eta_i\}, \quad \nu \in \mathbb{R}^3, \mu \in \mathbb{R},$$

where

$$C_{ijkl}(P, \mu) = \frac{\partial \tilde{S}^{ij}(P, \mu)}{\partial P_{kl}} = \frac{\partial^2 \psi(P, \mu)}{\partial P_{kl}\partial P_{ij}}, \qquad \tilde{S} = (\tilde{S}^{ij}).$$

4.

$$\partial_t^m f, \partial_t^m g \in C^0([0, T], H^{s-2-m}(\Omega)), \quad m = 0, 1, \ldots, s-2;$$
$$\partial_t^{s-1} f, \partial_t^{s-1} g \in L^2([0, T], L^2(\Omega)),$$

where $s \geq [3/2] + 3 = 4$ is an arbitrary but fixed integer. These hypotheses imply that (2.168), (2.169) form a hyperbolic–parabolic coupled system. The equations of nonlinear thermoelasticity for a homogeneous medium with unit reference density in $\Omega \subset \mathbb{R}^3$ are now obtained from (2.168), (2.169) as

$$\frac{\partial^2 U_i}{\partial t^2} = C_{ijkl}(\nabla U, \theta)\frac{\partial^2 U_k}{\partial x_j \partial x_l} + \tilde{C}_{ij}(\nabla U, \theta)\frac{\partial\theta}{\partial x_j}, \quad i = 1, 2, 3, \tag{2.170}$$

$$a(\nabla U, \theta)\theta_t = -\frac{1}{\chi(\theta)}\nabla' q(\nabla U, \theta, \nabla\theta) + \tilde{C}_{ij}(\nabla U, \theta)\frac{\partial^2 U_i}{\partial x_j \partial t}, \tag{2.171}$$

where

$$C_{ijkl} = \frac{\partial \tilde{S}^{ij}}{\partial(\partial_l U_k)}, \qquad \tilde{C}_{ij} = \frac{\partial \tilde{S}^{ij}}{\partial\theta}, \tag{2.172}$$

$$\tilde{S}^{ij} = \frac{\partial \psi}{\partial(\partial_j U_i)}, \qquad a = -\frac{\partial^2 \psi}{\partial \theta^2} \geq a_0 > 0 \tag{2.173}$$

for some positive constant a_0. Here $U = (U_1, U_2, U_3)'$, $q = (q_1, q_2, q_3)'$, χ is a C^∞-function such that $\chi(\theta) = \theta + T_0$ for $|\theta| \leq T_0/2$ and $0 < \chi_1 \leq \chi(\theta) \leq \chi_2 < \infty$, χ_1, χ_2 are constants, $-\infty < \theta < \infty$, and $T_0 > 0$ is the reference temperature. (2.170), (2.171) are derived for small values of $|\theta|$, i.e., for $|\theta| \leq T_0/2$, which is *a posteriori* justified by the smallness of the solutions obtained later on.

We consider again the initial conditions

$$U(t=0) = U^0, \quad U_t(t=0) = U^1, \quad \theta(t=0) = \theta^0 \tag{2.174}$$

and the Dirichlet boundary conditions

$$U_{|\partial\Omega} = 0, \quad \theta_{|\partial\Omega} = 0. \tag{2.175}$$

We assume that the medium is *initially isotropic*, i.e.,

$$\begin{aligned} &C_{ijkl}(0,0) = \lambda\tilde{\delta}_{ij}\tilde{\delta}_{kl} + \mu(\tilde{\delta}_{ik}\tilde{\delta}_{jl} + \tilde{\delta}_{jk}\tilde{\delta}_{il}), \quad \tilde{C}_{ij}(0,0) = -\tilde{\gamma}\tilde{\delta}_{ij}, \\ &\frac{\partial q_i(0,0,0)}{\partial(\partial\theta/\partial x_k)} = -\kappa T_0 \tilde{\delta}_{ik}, \quad a(0,0) = \tilde{\delta}, \end{aligned} \tag{2.176}$$

and that

$$\frac{\partial q_i(0,0,0)}{\partial \theta} = 0, \frac{\partial q_i(0,0,0)}{\partial(\partial U_j/\partial x_l)} = 0, \quad 1 \leq i,k,j,l \leq 3. \tag{2.177}$$

Using (2.176), we can write (2.170), (2.171) in the form:

$$U_{tt} - \mu\Delta U - (\mu+\lambda)\nabla\nabla' U + \tilde{\gamma}\nabla\theta = f^1(\nabla U, \nabla^2 U, \theta, \nabla\theta), \tag{2.178}$$

$$\tilde{\delta}\theta_t - \kappa\Delta\theta + \tilde{\gamma}\nabla' U_t = f^2(\nabla U, \nabla^2 U, \nabla U_t, \theta, \nabla^2\theta), \tag{2.179}$$

where $f^1 = (f_1, f_2, f_3)'$, and

$$\begin{aligned} f_i^1 := &\left(C_{ijkl}(\nabla U, \theta) - C_{ijkl}(0,0)\right)\frac{\partial^2 U_k}{\partial x_j \partial x_l} \\ &+ \left(\tilde{C}_{ij}(\nabla U, \theta) - \tilde{C}_{ij}(0,0)\right)\frac{\partial\theta}{\partial x_j}, \quad i = 1,2,3, \end{aligned} \tag{2.180}$$

$$\begin{aligned} f^2 := &\tilde{\delta}\left(\frac{1}{a(0,0)\chi(0)}\frac{\partial q_i(0,0,0)}{\partial(\partial\theta/\partial x_k)} - \frac{1}{a(\nabla U,\theta)\chi(\theta)}\frac{\partial q_i(\nabla U,\theta,\nabla\theta)}{\partial(\partial\theta/\partial x_k)}\right)\frac{\partial^2\theta}{\partial x_i \partial x_k} \\ &+ \tilde{\delta}\left(\frac{\tilde{C}_{ij}(\nabla U,\theta)}{a(\nabla U,\theta)} - \frac{\tilde{C}_{ij}(0,0)}{a(0,0)}\right)\frac{\partial^2 U_i}{\partial x_j \partial t} \end{aligned}$$

$$
\begin{aligned}
&- \frac{\tilde{\delta}}{a(\nabla U, \theta)\chi(\theta)} \frac{\partial q_i(\nabla U, \theta, \nabla\theta)}{\partial \theta} \frac{\partial \theta}{\partial x_i} \\
&- \frac{\tilde{\delta}}{a(\nabla U, \theta)\chi(\theta)} \frac{\partial q_i(\nabla U, \theta, \nabla\theta)}{\partial(\partial U_j/\partial x_l)} \frac{\partial^2 U_j}{\partial x_i \partial x_l}.
\end{aligned} \tag{2.181}
$$

Let U^j $(j = 2, 3, 4)$, θ^j $(j = 1, 2, 3)$ be defined through (2.178), (2.179) by

$$
U^j := \partial_t^j U|_{t=0}, \quad j = 2, 3, 4; \quad \theta^j := \partial_t^j \theta|_{t=0}, \quad j = 1, 2, 3.
$$

In fact, U^j, θ^j are obtained successively from U^0, U^1, and θ^0 by differentiating (2.178), (2.179) with respect to t at $t = 0$.

To get the global existence for small initial data we assume,in addition, that for $(\tilde{U}, \tilde{\theta})$ with $\operatorname{rot} \tilde{U} = 0$

$$
C_{ijkl}(\nabla\tilde{U}, \tilde{\theta}) \frac{\partial^2 \tilde{U}_k}{\partial x_j \partial x_l} \equiv A_{ik}(\nabla\tilde{U}, \tilde{\theta}) \Delta \tilde{U}_k, \quad i = 1, 2, 3 \tag{2.182}
$$

and that Assumption 2.4.d with $f = 0$ and $g = 0$ holds for the arguments in a neighborhood of zero. Then the main result for bounded domains is

THEOREM 2.50. *Assume that* $U^j \in H^{4-j}(\Omega)$ $(j = 0, \ldots, 4)$, $\theta^j \in H^{4-j}(\Omega)$ $(j = 0, 1, 2)$, $\theta^3 \in L^2(\Omega)$, *and that the initial data are compatible with boundary conditions* (2.175). *Let* (2.182) *hold, and let* $\operatorname{rot} U = 0$ *for* $(t, x) \in (0, T] \times \Omega$. *Then there is a constant* $\epsilon > 0$ *such that if*

$$
\sum_{j=0}^{4} \|U^j\|_{4-j,2}^2 + \sum_{j=0}^{2} \|\theta^j\|_{4-j,2}^2 + \|\theta^3\|^2 \le \epsilon^2, \tag{2.183}
$$

then there exists a unique solution (U, θ) *of* (2.178), (2.179), (2.174) and (2.175) *on* $(0, \infty)$ *satisfying*

$$
\left.
\begin{aligned}
&U \in \bigcap_{j=0}^{4} C^j([0, \infty), H^{4-j}(\Omega)), \theta \in \bigcap_{j=0}^{2} C^j([0, \infty), H^{4-j}(\Omega)), \\
&\theta_{ttt} \in C^0([0, \infty), L^2(\Omega)) \cap L^2([0, \infty), H^1(\Omega)), \\
&\forall (t, x) \in [0, \infty) \times \bar{\Omega} : \\
&\quad |\nabla U(t, x)|, |\theta(t, x)|, |\nabla\theta(t, x)| < \min\{1, T_0/2\}.
\end{aligned}
\right\} \tag{2.184}
$$

Moreover, $\|U(t)\|_{4,2}$, $\|\theta(t)\|_{4,2}$ *decay to zero exponentially as* $t \to \infty$.

Now we consider the Eqs. (2.170) and (2.171), together with the initial conditions (2.174), for the Cauchy problem in three space dimensions, i.e., for

$$
\Omega = \mathbb{R}^3,
$$

and for a homogeneous, initially isotropic medium. This leads to the equations

$$U_{tt} - \mathcal{D}'S\mathcal{D}U + \tilde{\gamma}\nabla\theta = f^1\left(\nabla U, \nabla^2 U, \theta, \nabla\theta\right), \tag{2.185}$$

$$\theta_t - \kappa\Delta\theta + \tilde{\gamma}\nabla' U_t = f^2\left(\nabla U, \nabla U_t, \nabla^2 U, \theta, \nabla\theta, \nabla^2\theta\right), \tag{2.186}$$

where we have assumed, without loss of generality, that $\tilde{\delta} = 1$. It is known (cf. [91] and the references therein) that in the case of pure elasticity there are global, small solutions if the nonlinearity degenerates up to order two, i.e., if the nonlinearity f^1 is cubic (near zero values of its arguments). In the "genuinely nonlinear" case, a blow-up in finite time has to be expected; this was proved for plane waves and for radial solutions, cf. [48,49,91]. On the other hand, quadratic nonlinearities in $\mathbb{R}^3$ still lead to global, small solutions of the heat equation, cf. [91]. The question remains, whether the dissipative impact through heat conduction is strong enough to prevent solutions from blowing up, at least for small data.

The answer to this question will be positive if one excludes purely quadratic nonlinearities in the displacement. This perfectly corresponds to the fact that for these nonlinearities one has to expect a blow-up, as we shall see below. Thus, we admit all possible cubic nonlinearities (in the final setting) or those quadratic terms which involve θ, which guarantees that a damping effect (dissipation) is present in each equation.

The assumption on the nonlinearity will be

> There are no purely quadratic terms only involving ∇U, ∇U_t, $\nabla^2 U$ and additionally one of the following two cases is given:
> Case I: Only quadratic terms appear.
> Case II: At least only cubic terms appear and one quadratic term of the type $\theta\Delta\theta$. (2.187)

The specific quadratic nonlinearity of the type $\theta\Delta\theta$ always appears, and is due to the special function $\chi(\theta)$. This quadratic term cannot be assumed to vanish by any assumption of the general nonlinearities. The terms excluded are typically $\nabla U \nabla^2 U$ and $\nabla U \nabla U_t$. As mentioned above, the appearance of purely quadratic terms in ∇U, ∇U_t, $\nabla^2 U$ may lead to the development of singularities. The global existence theorem then reads

THEOREM 2.51. *Let the nonlinearity satisfy (2.187). Then there exist an integer s_0 and a $\tilde{\delta} > 0$ such that the following holds:*

If $(\nabla U^0, U^1, \theta^0) \in H^s(\mathbb{R}^3) \cap W^{s,1}(\mathbb{R}^3)$ with $s \geq s_0$ and

$$\left\|\left(\nabla U^0, U^1, \theta^0\right)\right\|_{s,2} + \left\|\left(\nabla U^0, U^1, \theta^0\right)\right\|_{s,1} < \tilde{\delta},$$

then there is a unique solution (U, θ) of the initial value problem to the nonlinear equations of thermoelasticity (2.170), (2.171), (2.174) with

$$(\nabla U, U_t) \in C^0\left([0,\infty), H^s(\mathbb{R}^3)\right) \cap C^1\left([0,\infty), H^{s-1}(\mathbb{R}^3)\right),$$

$$\theta \in C^0\left([0,\infty), H^s(\mathbb{R}^3)\right) \cap C^1\left([0,\infty), H^{s-2}(\mathbb{R}^3)\right).$$

Moreover, the asymptotic behavior can be described as follows:
Case I (no cubic terms): There exist integers $l < k' < k \leq s$ *such that for* $\epsilon < 1/8$ *we have*

$$\begin{aligned}
&\|(\nabla U, U_t, \theta)(t)\|_{s,2} = \mathcal{O}(1), \\
&\|\nabla^3\theta(t)\|_{k,2} + \|(\nabla U, U_t)(t)\|_{l,\infty} = \mathcal{O}\left(t^{-3/4+\epsilon}\right), \\
&\|\nabla^3\theta(t)\|_{k',2} = \mathcal{O}\left(t^{-3/2+2\epsilon}\right), \\
&\|\nabla\theta(t)\|_{1,2} + \|\theta_t(t)\|_{\infty} = \mathcal{O}\left(t^{-5/4}\right), \\
&\|\theta(t)\|_{2,2} = \mathcal{O}\left(t^{-3/4}\right), \\
&\|\theta(t)\|_{2,\infty} = \mathcal{O}\left(t^{-3/2+\epsilon}\right) \quad \text{as } t \to \infty.
\end{aligned}$$

Case II (no quadratic terms but $\theta\Delta\theta$*): There exist integers* $s_1, s_3, s_5 \leq s$ *with*

$$\begin{aligned}
&\|(\nabla U, U_t, \theta)(t)\|_{s,2} = \mathcal{O}(1), \\
&\|(\nabla U, U_t)(t)\|_{s_1,9/2} = \mathcal{O}\left(t^{-5/9}\right), \\
&\|\theta(t)\|_{s_3,13/2} + \|\nabla^2\theta(t)\|_{s_5,26/11} + \|\nabla^2\theta(t)\|_{s_5,18/7} = \mathcal{O}\left(t^{-27/26}\right) \\
&\quad \text{as } t \to \infty.
\end{aligned}$$

For the proof of the last theorem and the complementing following theorem on the possible blow-up of solutions for quadratic nonlinearities see [33] or [47].

THEOREM 2.52. *There exist quadratic nonlinearities such that for compactly supported non-vanishing smooth data* $(U^{0,\sigma}, U^{1,\sigma})$ *which are sufficiently small, i.e.,*

$$\sup_{x_1\in\mathbb{R}} \left|\partial_1\left(\partial_1 U_2^0, \partial_1 U_3^0, U_2^1, U_3^1\right)(x_1)\right|$$

$$\left\{\text{resp. } \sup_{x_1\in\mathbb{R}} \left|\partial_1^2\left(\partial_1 U_2^0, \partial_1 U_3^0, U_2^1, U_3^1\right)(x_1)\right|\right\}$$

is sufficiently small, a plane-wave solution of the nonlinear equations of thermoelasticity (2.170), (2.171), (2.174) in $\mathbb{R}^3$ *cannot be of class* C^2 $\{$*resp.* C^3 $\}$ *for all positive* t.

Notes: The existence results for bounded domains are taken from the paper of Racke, Shibata, and Zheng [96], the results on the Cauchy problem mainly from the papers of Hrusa and Tarabek [32] and Kawashima [51]. The semi-axes problem was discussed by Jiang [40], stationary forces by Kawashima and Shibata [53]. The blow-up result in one dimension was obtained by Hrusa and Messaoudi [31], while the existence of weak solutions is due to Durek [20].

Further results for one-dimensional systems, both decay results and global existence theorems for various boundary conditions, are contained in the papers by Slemrod [107], Kawashima [50], Zheng [123,124], Kawashima and Okada [52], Shen and Zheng [104], Zheng and Shen [126], Racke and Shibata [95], Shibata [105], Jiang [41–43], see also [88].

The development of singularities of solutions to the Cauchy problem for large data was first shown by Dafermos and Hsiao [14]. I. Hansen [25] obtained a similar blow-up result for initial boundary value problems. Periodic small solutions are also studied by Feireisl [21].

Weak solutions for systems in one space dimension, that do not conduct heat, have been studied by Chen and Dafermos [9].

The radially symmetric case was discussed by Jiang [44] and more generally by Jiang, Muñoz Rivera and Racke [45], the latter being the basis for the presentation here. Other boundary conditions were studied by Rieger [103]. The results on global existence and blow-up for the Cauchy problem are based on the original work by Racke [89,90] and Ponce and Racke [80]. For local well-posedness results see also Chrzęszczyk [10], Jiang and Racke [46], Dan [15], and Mukoyama [71].

3. Thermoelasticity with Second Sound

The classical thermoelastic system as a hyperbolic–parabolic coupling inhibits the paradox of infinite speed of propagation of signals. This is due to the classical heat equation based on Fourier's law for the dependence of the heat flux on the temperature (gradient). Taking the simplest one-dimensional model from Section 2, given in the Eqs. (2.12) and (2.13), we may rewrite these for zero right-hand sides as

$$u_{tt} - \alpha u_{xx} + \tilde{\gamma}\theta_x = 0, \tag{3.1}$$

$$\tilde{\delta}\theta_t + q_x + \tilde{\gamma} u_{tx} = 0, \tag{3.2}$$

$$q + \kappa\theta_x = 0, \tag{3.3}$$

where q denotes the heat flux. Here (3.3) is the Fourier law representing an instantaneous response to changes in the gradient of the temperature visible in the heat flux. This causes the effect of infinite propagation of signals in this model. For some applications like working with very short laser pulses in laser cleaning of computer chips, see the references in [93], it is worthwhile thinking of another model removing this paradox, but still keeping the essentials of a heat conduction process. One such model – for a survey compare Chandrasekharaiah [8], for general Cattaneo models cf. Öncü and Moodie [79] – is given by the simplest Cattaneo law replacing Fourier's law (3.3),

$$\tau q_t + q + \kappa\theta_x = 0, \tag{3.4}$$

now regarding the heat flux vector as another function to be determined through the differential equation, initial conditions, and, if applicable, boundary conditions. The positive parameter τ is the relaxation time describing the time lag in the response of the heat flux to a gradient in the temperature.

Neglecting the elastic effects for a moment ($u = 0, \gamma = 0$) and just looking at heat conduction models, we obtain for the temperature the equation

$$\tau\theta_{tt} + \theta_t - \tilde{\beta}\theta_{xx} = 0, \qquad (\tilde{\beta} = \kappa/\tilde{\delta}). \tag{3.5}$$

For $\tau = 0$ obtained under Fourier's law, we have the classical parabolic heat equation, while for $\tau > 0$, obtained under Cattaneo's law, we see the classical damped wave equation. For both, under proper initial and boundary conditions, we have the exponential stability in bounded domains in $\mathbb{R}^n$, and for both we have the same decay rate of the L^∞-norm $(n/2)$ for the Cauchy problem in $\mathbb{R}^n$. Hence there is an expectation that the characteristic behavior should be the same also in connection with elastic systems.

Indeed, we shall demonstrate this for various situations in what follows. Actually, for one-dimensional models we shall see that for a series of real materials the quantitative behavior is similar. But it is important to notice that this is not true, for example, for Timoshenko-type thermoelastic systems, where a system can be or remain exponentially stable under Fourier's law, while it loses this property under Cattaneo's law, see [23]. Therefore, a careful analysis of systems with second sound is necessary. Technical difficulties arise from the fact that there the heat flux no longer has the same regularity as the temperature gradient. Recent investigations of Sprenger [109] for contact problems in thermoelasticity with second sound underline this.

The general differential equations are formally the same as (2.1) and (2.2) (for right-hand sides being zero in the sequel), i.e.,

$$\rho U_{tt} - \nabla' \tilde{S} = 0, \tag{3.6}$$

$$\varepsilon_t - \mathrm{tr}\{\tilde{S} F_t\} + \nabla' q = 0. \tag{3.7}$$

The difference is now two-fold: First, the Helmholtz free energy additionally depends on the heat flux,

$$\psi = \psi(\nabla U, \theta, q) \tag{3.8}$$

turning (3.7) into

$$\begin{aligned}(\theta + T_0)\left\{\tilde{a}(\nabla U, \theta, q)\theta_t - \mathrm{tr}\left[(S_\theta(\nabla U, \theta, q))' \nabla U_t\right]\right\}\\ + \operatorname{div} q = \tilde{b}(\nabla U, \theta, q) q_t\end{aligned} \tag{3.9}$$

with

$$\tilde{a} := -\psi_{\theta\theta}, \quad \tilde{b}(\nabla U, \theta, q) := (\theta + T_0)\psi_{\theta q}(\nabla U, \theta, q) - \psi_q(\nabla U, \theta, q),$$

and, second, Cattaneo's law is assumed,

$$\tau(\nabla U, \theta) q_t + q + k(\nabla U, \theta)\nabla\theta = 0, \tag{3.10}$$

where τ is the tensor of relaxation times.

We shall present a survey of results for spatially one- and two- or three-dimensional models, covering the time asymptotics, the well-posedness of nonlinear systems, and the propagation of singularities. Remarks on possible further (hyperbolic) models will conclude the considerations.

3.1. *Linearized systems*

For the linearized model, for which the well-posedness can be proved with semigroup theory again, we consider, exemplarily, the system

$$u_{tt} - \alpha u_{xx} + \beta\theta_x = 0, \tag{3.11}$$

$$\theta_t + \gamma q_x + \delta u_{tx} = 0, \tag{3.12}$$

$$\tau_0 q_t + q + \kappa\theta_x = 0, \tag{3.13}$$

where $x \in (0, L)$, and where $\alpha, \beta, \gamma, \delta, \tau_0, \kappa$ are positive constants, appropriate for an underlying homogeneous medium. In particular, τ_0 is the relaxation time, a parameter being small in comparison to the others.

If we rewrite the equations as a first-order system for $W := (u_x, u_t, \theta, q)$ we see that W satisfies $W_t = AW_x + B(W)$, where $B(W) = (0, 0, 0 - \frac{1}{\tau_0}q)'$.$A$ has four real, distinct eigenvalues, hence it is a strictly hyperbolic system in the main part $W_t = AW_x$ with a damping term $B(W)$.

The differential equations are made complete with initial conditions

$$u(0,\cdot) = u_0, \quad u_t(0,\cdot) = u_1, \quad \theta(0,\cdot) = \theta_0, \quad q(0,\cdot) = q_0 \tag{3.14}$$

and boundary conditions

$$u(t,0) = u(t,L) = \theta(t,0) = \theta(t,L) = 0, \quad t \geq 0. \tag{3.15}$$

Then we have the following result on exponential stability.

THEOREM 3.1. *Let (u, θ, q) be the solution to (3.11)–(3.15). Then the associated energy of first and second order,*

$$E(t) = \frac{1}{2}\sum_{j=1}^{2}\int_0^L \Big\{\kappa\delta(\partial_t^{j-1}u_t)^2 + \kappa\delta\alpha(\partial_t^{j-1}u_x)^2 + \kappa\beta\left(\partial_t^{j-1}\theta\right)^2$$
$$+ \gamma\beta\tau_0(\partial_t^{j-1}q)^2\Big\}(t,x)\mathrm{dx}$$

decays exponentially, i.e.,

$$\exists d_0, C_0 > 0 \quad \forall t \geq 0: \quad E(t) \leq C_0 \mathrm{e}^{-d_0 t}E(0).$$

The proof works with multiplicative techniques, as for the classical case in Section 3.2, but has to overcome the technical difficulties arising from the fact that the heat flux has a different regularity to the temperature gradient. We shall consider a few details in the three-dimensional situation below.

For different real materials the optimal decay rate d_{tss} was computed for the boundary conditions $u = q = 0$ (assuming $\int_0^L \theta_0 \mathrm{dx} = 0$) as the spectral bound of the associated first-order operator, i.e., as the minimum of the real parts of the eigenvalues. This was possible since these boundary conditions allow a Fourier series expansion, e.g.

$$u(t,x) = \sum_{j=1}^{\infty} a_j(t) s_j(x),$$

where

$$s_j(x) := \sqrt{\frac{L}{2}} \sin(\sqrt{\lambda_j} x), \qquad c_j(x) := \sqrt{\frac{L}{2}} \cos(\sqrt{\lambda_j} x), \quad \lambda_j := \frac{j^2 \pi^2}{L^2}.$$

The characteristic polynomial is

$$\tau_0 \rho^4 - \rho^3 + (\tau_0 \alpha + \tau_0 \beta \delta + \gamma \kappa) \lambda_j \rho^2 - (\alpha + \beta \delta) \lambda_j \rho + \alpha \gamma \kappa \lambda_j^2 = 0, \qquad (3.16)$$

with roots $\rho_k(j)$, $k = 1, 2, 3, 4$. Then the optimal decay rate is

$$d_{tss} = \inf_{j \in \mathbb{N}, k=1,2,3,4} \mathrm{Re} \rho_k(j).$$

For $\tau_0 = 0$ we have the situation in classical thermoelasticity. It turns out that the optimal decay rates are the *same* for classical thermoelasticity (d_{ct}) and for thermoelasticity with second sound (d_{tss}), within the computed range given in the following table, listing these decay rates for different materials. The domain length L was chosen to be $L = 6.25 \times 10^{-4}$[m] as it is typical for the example from laser cleaning mentioned above; the value for the relaxation is also a typical one of order 10^{-12} [s]. For a comparison we also list the corresponding values d_h for pure classical heat conduction ($u = 0$, $\gamma = 0$, $\tau_0 = 0$).

Material	$d_{ct} = d_{tss}$	d_h
silicon	$7.462 \cdot 10^{-1}$	$2.240 \cdot 10^3$
aluminum alloy	$1.819 \cdot 10^1$	$1.504 \cdot 10^3$
steel	$2.721 \cdot 10^{-1}$	$1.191 \cdot 10^2$
germanium	$9.930 \cdot 10^{-1}$	$8.881 \cdot 10^2$
gallium arsenid	$7.838 \cdot 10^{-1}$	$7.921 \cdot 10^2$
indium arsenid	$2.667 \cdot 10^{-1}$	$4.803 \cdot 10^2$
copper	$2.117 \cdot 10^1$	$3.066 \cdot 10^3$
diamond	$2.440 \cdot 10^0$	$1.793 \cdot 10^4$

Of course there are differences in the behavior for different parts of the solution, in particular for high-frequency (j large) parts, but the mean behavior as time tends to infinity is very much the same. We remark that it is also possible to have second sound systems that decay faster than the corresponding classical ones. Below we shall get more information on the closeness of the two thermoelastic systems for specific materials.

For the Cauchy problem in one space dimension, a detailed analysis of the Fourier transformed system and the different regions of low, middle and high frequency parts, gives the following description of polynomial decay rates

THEOREM 3.2. *Let* (u, θ, q) *be the solution to the Cauchy problem corresponding to* (3.11)–(3.14). *Then it satisfies*

$$\|(u_t, u_x, \theta, q)(t, \cdot)\|_{L^q(\mathbb{R})} \leq C(1+t)^{-\frac{1}{2}(1-\frac{2}{q})}\|(u_1, u_0, \theta_0, q_0)\|_{W^{2,p}(\mathbb{R})},$$

where $2 \leq q \leq \infty$, $1/p + 1/q = 1$, *and* C *is a positive constant, neither depending on* t *nor on the data.*

Now let $\Omega \subset \mathbb{R}^n$, $n = 2, 3$, be a bounded reference configuration with smooth boundary $\partial\Omega$ (C^2 is sufficient). Then the differential equations for a homogeneous isotropic medium are

$$U_{tt} - \mu\Delta U - (\mu + \lambda)\nabla\nabla' U + \beta\nabla\theta = 0, \tag{3.17}$$

$$\theta_t + \gamma\nabla' q + \delta\nabla' U_t = 0, \tag{3.18}$$

$$\tau_0 q_t + q + \kappa\nabla\theta = 0. \tag{3.19}$$

The parameters $\mu, \beta, \gamma, \delta, \tau_0$ and κ are again positive constants with $\lambda + 2\mu > 0$. These equations are completed by initial and boundary conditions.

$$U(0, \cdot) = U_0, \qquad U_t(0, \cdot) = U_1, \qquad \theta(0, \cdot) = \theta_0, \qquad q(0, \cdot) = q_0, \tag{3.20}$$

$$U(t, \cdot) = 0, \qquad \theta(t, \cdot) = 0 \quad \text{on } \partial\Omega, t \geq 0. \tag{3.21}$$

Having in mind the results for classical thermoelasticity from Section 2, we assume

$$\operatorname{rot} U = \operatorname{rot} q = 0 \quad \text{in } [0, \infty) \times \Omega, \tag{3.22}$$

$$\nu \times q = 0 \quad \text{in } [0, \infty) \times \partial\Omega. \tag{3.23}$$

This is, in particular, satisfied for radially symmetric situations. We then have

$$\nabla\nabla' U = \Delta U \quad \text{and} \quad \|\nabla U\| = \|\nabla' U\| \tag{3.24}$$

and

$$\mu\Delta + (\mu + \lambda)\nabla\nabla' U = \alpha\Delta U,$$

where

$$\alpha := 2\mu + \lambda.$$

Let the energy terms of first and second order be defined as

$$E_1(t) := \frac{1}{2}\int_\Omega \left\{\kappa\delta|U_t|^2 + \kappa\delta\alpha|\nabla U|^2 + \kappa\beta|\theta|^2 + \tau\beta\gamma|q|^2\right\}(t, x)\mathrm{dx}$$
$$\equiv E(t; U, \theta, q),$$

$$E_2(t) := E(t; U_t, \theta_t, q_t).$$

THEOREM 3.3. *Let* (U, θ, q) *be the solution to* (3.17)–(3.21) *satisfying* (3.22), (3.23). *Then the associated energy*

$$E(t) := E_1(t) + E_2(t)$$

decays exponentially, i.e.,

$$\exists d_0, C_0 > 0 \quad \forall t \geq 0 : \quad E(t) \leq C_0 \mathrm{e}^{-d_0 t} E(0).$$

PROOF.

$$\frac{\mathrm{d}}{\mathrm{dt}} E_1(t) = -\beta\gamma \|q\|^2, \qquad \frac{\mathrm{d}}{\mathrm{dt}} E_2(t) = -\beta\gamma \|q_t\|^2. \tag{3.25}$$

The differential equation (3.19) yields

$$\|\nabla\theta\|^2 \leq \frac{2\tau_0^2}{\kappa^2}\|q_t\|^2 + \frac{2}{\kappa^2}\|q\|^2. \tag{3.26}$$

Multiplying (3.17) by $\frac{1}{\alpha}\Delta U$ and observing (3.24) we get

$$\|\Delta U\|^2 \leq -\frac{1}{\alpha}\frac{\mathrm{d}}{\mathrm{dt}}\langle \nabla U_t, \nabla U\rangle + \frac{1}{\alpha}\|\nabla U_t\|^2 + \frac{3}{4}\frac{\beta^2}{\alpha^2}\|\nabla\theta\|^2 + \frac{1}{3}\|\Delta U\|^2$$

implying

$$\frac{2}{3}\|\Delta U\|^2 + \frac{1}{\alpha}\frac{\mathrm{d}}{\mathrm{dt}}\langle \nabla U_t, \nabla U\rangle \leq \frac{1}{\alpha}\|\nabla U_t\|^2 + \frac{3}{4}\frac{\beta^2}{\alpha^2}\|\nabla\theta\|^2. \tag{3.27}$$

Multiplying (3.18) by $\frac{3}{\alpha\delta}\nabla' U_t$ we obtain

$$\frac{3}{\alpha}\|\nabla' U_t\|^2 = \frac{3\gamma}{\alpha\delta}\langle q, \nabla\nabla' U_t\rangle - \frac{3\gamma}{\alpha\delta}\langle \nu q, \nabla' U_t\rangle_{\partial\Omega} + \frac{3}{\alpha\delta}\langle \nabla\theta_t, U_t\rangle,$$

where $\langle\cdot,\cdot\rangle_{\partial\Omega}$ denotes the $L^2(\partial\Omega)$-inner product with norm $\|\cdot\|_{\partial\Omega}$. Thus, using the Eqs. (3.17) and (3.19) again,

$$\begin{aligned}\frac{3}{\alpha}\|\nabla' U_t\|^2 \leq{}& \frac{3\gamma}{\alpha^2\delta}\frac{\mathrm{d}}{\mathrm{dt}}\langle q, U_{tt}\rangle - \frac{3\beta\gamma\tau_0}{\alpha^2\delta\kappa}\frac{\mathrm{d}}{\mathrm{dt}}\langle q, q_t\rangle - \frac{3\beta\gamma}{\alpha^2\delta\kappa}\frac{\mathrm{d}}{\mathrm{dt}}\|q\|^2 \\ &+ \frac{27\gamma^2}{\alpha^2\delta^2}\|q_t\|^2 + \frac{1}{12}\|\Delta U\|^2 - \frac{3\tau_0}{\alpha\delta\kappa}\frac{\mathrm{d}}{\mathrm{dt}}\langle q_t, U_t\rangle - \frac{3}{\alpha\delta\kappa}\frac{\mathrm{d}}{\mathrm{dt}}\langle q, U_t\rangle \\ &+ \frac{27}{\delta^2}\|\nabla\theta\|^2 + \frac{1}{12}\|\Delta U\|^2 + \frac{3\beta}{\alpha\delta}\|\nabla\theta\|^2 - \frac{3\gamma}{\alpha\delta}\langle \nu q, \nabla' U_t\rangle_{\partial\Omega}.\end{aligned} \tag{3.28}$$

Combining (3.27) and (3.28) we obtain, observing (3.24),

$$\frac{2}{\alpha}\|\nabla U_t\|^2 + \frac{1}{2}\|\Delta U\|^2 + \frac{\mathrm{d}}{\mathrm{dt}}G_1(t) \le \left(\frac{3}{4}\frac{\beta^2}{\alpha^2} + \frac{27}{\delta^2} + \frac{3\beta}{\alpha\delta}\right)\|\nabla\theta\|^2 + \frac{27\gamma^2}{\alpha^2\delta^2}\|q_t\|^2 - \frac{3\gamma}{\alpha\delta}\langle \nu q, \nabla' U_t\rangle_{\partial\Omega}, \quad (3.29)$$

where

$$G_1(t) := \frac{1}{\alpha}\langle\nabla U_t, \nabla U\rangle - \frac{3\gamma}{\alpha^2\delta}\langle q, U_{tt}\rangle + \frac{3\beta\gamma\tau_0}{\alpha^2\delta\kappa}\langle q, q_t\rangle + \frac{3\beta\gamma}{\alpha^2\delta\kappa}\|q\|^2 + \frac{3\tau_0}{\alpha\delta\kappa}\langle q_t U_t\rangle + \frac{3}{\alpha\delta\kappa}\langle q, U_t\rangle. \quad (3.30)$$

Using the first Poincaré estimate for U_t and θ, as well as (3.17) for U_{tt}, we obtain

$$\|U_{tt}\|^2 + \|U_t\|^2 + \|\theta\|^2 \le c(\|\Delta U\|^2 + \|\nabla\theta\|^2 + \|\nabla U_t\|^2), \quad (3.31)$$

where $c > 0$ will again denote a constant that may take different values at different places, $c = c(\alpha, \beta, \gamma, \delta, \tau_0, \Omega)$.

Multiplying (3.17) by U we get

$$\frac{\alpha}{2}\|\nabla U\|^2 \le c(\|U_{tt}\|^2 + \|\nabla\theta\|^2). \quad (3.32)$$

A multiplication of (3.18) by θ_t yields

$$\|\theta_t\|^2 = \frac{\mathrm{d}}{\mathrm{dt}}\langle q, \nabla\theta\rangle - \gamma\langle q_t, \nabla\theta\rangle + \frac{\delta^2}{2}\|\nabla' U_t\|^2 + \frac{1}{2}\|\theta_t\|^2$$

whence, for arbitrary $\tilde{\mu} > 0$,

$$\tilde{\mu}\|\theta_t\|^2 - 2\tilde{\mu}\frac{\mathrm{d}}{\mathrm{dt}}\langle q, \nabla\theta\rangle \le \tilde{\mu}\gamma\|q_t\|^2 + \tilde{\mu}\gamma\|\nabla\theta\|^2 + \delta^2\tilde{\mu}\|\nabla' U_t\|^2 \quad (3.33)$$

follows, and $\tilde{\mu}$ will be determined later.

The boundary term appearing in (3.29) will be treated next, similar to Section 2.

$$\left|\frac{3\gamma}{\alpha\delta}\langle \nu q, \nabla' U_t\rangle_{\partial\Omega}\right| \le \frac{c_1}{\hat{\varepsilon}}\|\nu q\|^2_{\partial\Omega} + \hat{\varepsilon}\left\|\nabla' U_t\right\|^2_{\partial\Omega}, \quad (3.34)$$

where $1 > \hat{\varepsilon} > 0$ is still arbitrary and will be determined later, and c_1 (similarly $c_2, c_3, \ldots$) denotes a fixed constant.

Let $\sigma \in (C^1(\overline{\Omega}))^3$ be such that $\sigma = (\sigma_i)_{i=1,\cdot,n}$ with $\sigma_k = \nu_k$ on $\partial\Omega$, and let $\partial_k := \frac{\partial}{\partial x_k}$, $k = 1, \ldots, n$.

Multiplying (3.18) by $-\frac{\beta}{\delta}\sigma_k\partial_k\theta_t$ (summation convention) we obtain

$$0 = -\frac{\beta}{\delta}\|\theta_t\|^2_{\partial\Omega} + \frac{\beta}{2\delta}\langle(\nabla'\sigma)\theta_t, \theta_t\rangle - \frac{\beta\gamma}{\delta}\frac{\mathrm{d}}{\mathrm{dt}}\langle\nabla' q, \sigma_k\partial_k\theta\rangle$$

$$- \frac{\beta\gamma}{\tau_0\delta}\langle\nabla' q, \sigma_k\partial_k\theta\rangle - \frac{\beta\gamma\kappa}{\tau_0\delta}\langle\Delta\theta, \sigma_k\partial_k\theta\rangle - \beta\langle\nabla' U_t, \sigma_k\partial_k\theta_t\rangle. \tag{3.35}$$

Since

$$\begin{aligned}\langle\Delta\theta, \sigma_k\partial_k\theta\rangle &= \langle\nu\nabla\theta, \nu_k\partial_k\theta\rangle_{\partial\Omega} - \langle\nabla\theta, (\nabla\sigma_k)\partial_k\theta\rangle - \langle\nabla\theta, \sigma_k\partial_k\nabla\theta\rangle \\ &= \frac{1}{2}\left\|\frac{\partial\theta}{\partial\nu}\right\|^2_{\partial\Omega} - \langle\nabla\theta, (\nabla\sigma_k)\partial_k\theta\rangle + \frac{1}{2}\langle(\nabla'\sigma)\nabla\theta, \nabla\theta\rangle,\end{aligned}$$

where we used the boundary condition $\theta_{|\partial\Omega} = 0$ to conclude that $\nabla\theta = \frac{\partial\theta}{\partial\nu}\cdot\nu$, we get from (3.35),

$$\begin{aligned}0 = {}& \frac{2\beta}{\delta}\|\theta_t\|^2_{\partial\Omega} - \frac{\beta}{\delta}\langle(\nabla'\sigma)\theta_t, \theta_t\rangle + \frac{2\beta\gamma}{\delta}\frac{\mathrm{d}}{\mathrm{dt}}\langle\nabla' q, \sigma_k\partial_k\theta\rangle \\ &+ \frac{2\beta\gamma}{\tau_0\delta}\langle\nabla' q, \sigma_k\partial_k\theta\rangle + \frac{\beta\gamma\kappa}{\tau_0\delta}\left\|\frac{\partial\theta}{\partial\nu}\right\|^2_{\partial\Omega} - \frac{2\beta\gamma\kappa}{\tau_0\delta}\langle\nabla\theta, (\nabla\sigma_k)\partial_k\theta\rangle \\ &- \frac{\beta\gamma\kappa}{\tau_0\delta}\langle(\nabla'\sigma)\nabla\theta, \nabla\theta\rangle + 2\beta\langle\nabla' U_t, \sigma_k\partial_k\theta_t\rangle.\end{aligned} \tag{3.36}$$

On the other hand, we obtain in the same way, differentiating (3.17) with respect to t and multiplying by $\sigma_k\partial_k U_t$, cf. Lemma 2.25, that

$$\begin{aligned}0 = {}& \left\|\frac{\partial U_t}{\partial\nu}\right\|^2_{\partial\Omega} + (\mu+\lambda)\left\|\nabla' U_t\right\|^2_{\partial\Omega} - 2\frac{\mathrm{d}}{\mathrm{dt}}\langle U_{tt}, \sigma_k\partial_k U_t\rangle \\ &- \langle(\nabla'\sigma)U_{tt}, U_{tt}\rangle - 2\mu\langle\partial_j U_t^i, (\partial_j\sigma_k)\partial_k U_t^i\rangle \\ &+ \mu\langle(\nabla'\sigma)\nabla U_t, \nabla U_t\rangle - 2(\mu+\lambda)\langle\nabla' U_t, (\nabla\sigma_k)\partial_k U_t\rangle \\ &+ (\mu+\lambda)\langle(\nabla'\sigma)\nabla' U_t,\ \nabla' U_t\rangle - 2\beta\langle\nabla\theta_t, \sigma_k\partial_k U_t\rangle.\end{aligned} \tag{3.37}$$

Moreover, we have

$$\begin{aligned}\mathrm{Re}\langle\nabla' U_t, \sigma_k\partial_k\theta_k\rangle = {}& \mathrm{Re}\langle\nabla\theta_t, \sigma_k\partial_k U_t\rangle + \mathrm{Re}\langle\theta_t, (\nabla\sigma_k)\partial_k U_t \\ &- (\nabla'\sigma)\nabla' U_t\rangle.\end{aligned} \tag{3.38}$$

Adding (3.36) and (3.37), taking real parts and observing (3.38), we obtain

$$\begin{aligned}&\frac{2\beta}{\delta}\|\theta_t\|^2_{\partial\Omega} + \frac{\beta\gamma\kappa}{\tau_0\delta}\left\|\frac{\partial\theta}{\partial\nu}\right\|^2_{\partial\Omega} + \left\|\frac{\partial U_t}{\partial\nu}\right\|^2_{\partial\Omega} + (\mu+\lambda)\left\|\nabla' U_t\right\|^2_{\partial\Omega} \\ &\quad + \frac{\mathrm{d}}{\mathrm{dt}}\mathrm{Re}\left(\frac{2\beta\gamma}{\delta}\langle\nabla' q, \sigma_k\partial_k\theta\rangle - 2\langle U_{tt}, \sigma_k\partial_k U_t\rangle\right) \\ &\quad\quad \le c_2(\|\theta_t\|^2 + \|\nabla\theta\|^2 + \|\nabla U_t\|^2 + \|\Delta U\|^2).\end{aligned} \tag{3.39}$$

Until now we only used $\mathrm{rot}\, U = 0$ from the assumptions (3.22), (3.23). In the next step we also exploit $\mathrm{rot}\, q = 0$ and $\nu\times q = 0$ on $\partial\Omega$.

Multiplying (3.18) by σq yields

$$\tau_0\langle\theta_t, \sigma q\rangle + \gamma\langle\nabla' q, \sigma q\rangle + \delta\langle\nabla' U_t, \sigma q\rangle = 0 \tag{3.40}$$

with

$$|\tau_0\langle\theta_t, \sigma q\rangle| \le \tilde{\varepsilon}\|\theta_t\|^2 + \frac{c_3}{\tilde{\varepsilon}}\|q\|^2 \tag{3.41}$$

for some $\tilde{\varepsilon} > 0$ to be determined later. Then

$$|\delta\langle\nabla' U_t, \sigma q\rangle| \le \tilde{\varepsilon}\|\nabla' U_t\|^2 + \frac{c_4}{\tilde{\varepsilon}}\|q\|^2, \tag{3.42}$$

$$\langle\nabla' q, \sigma q\rangle = \frac{1}{2}\|\nu q\|^2 - \langle q_j, (\partial_j\sigma_k)q_k\rangle + \frac{1}{2}\langle q, (\nabla'\sigma)q\rangle. \tag{3.43}$$

Combining (3.40)–(3.43) we conclude ($\tilde{\varepsilon}$ small)

$$\|\nu q\|^2 \le \tilde{\varepsilon}(\|\theta_t\|^2 + \|\nabla' U_t\|^2) + \frac{c_5}{\tilde{\varepsilon}}\|q\|^2. \tag{3.44}$$

From (3.29), (3.33), (3.34) and (3.39) (multiplied by $\frac{\hat{\varepsilon}}{\mu+\lambda}$), (3.44) follows for sufficiently small $\tilde{\mu}, \hat{\varepsilon}, \tilde{\varepsilon}$:

$$\frac{1}{\alpha}\|\nabla U_t\|^2 + \frac{1}{4}\|\Delta U\|^2 + c_6\|\theta_t\|^2 + \frac{\mathrm{d}}{\mathrm{dt}}H(t) \le c_7(\|q_t\|^2 + \|q\|^2), \tag{3.45}$$

where (cp. (3.30) for $G_1(t)$)

$$H(t) := G_1(t) + \frac{\hat{\varepsilon}}{\mu+\lambda}\mathrm{Re}\left(\frac{2\beta\gamma}{\delta}\langle\nabla' q, \sigma_k\partial_k\theta\rangle - 2\langle U_{tt}, \sigma_k\partial_k U_t\rangle\right) \\ - 2\tilde{\mu}\langle q, \nabla\theta\rangle. \tag{3.46}$$

A suitable Lyapunov function F is defined by

$$F(t) := \frac{1}{\varepsilon}(E_1(t) + E_2(t)) + H(t),$$

where $\varepsilon > 0$.

Combining (3.25), (3.31), (3.32) and (3.45) we see that F satisfies, for sufficiently small ε,

$$\frac{\mathrm{d}}{\mathrm{dt}}F(t) \le -d_1(E_1(t) + E_2(t)) \tag{3.47}$$

for some $d_1 > 0$. On the other hand (ε small enough)

$$\exists C_1, C_2 > 0 \quad \forall t \ge 0 : C_1 E(t) \le F(t) \le C_2 E(t),$$

with $E(t) = E_1(t) + E_2(t)$. This yields the assertion with

$$d_0 := \frac{d_1}{C_2}, \qquad C_0 := \frac{C_2}{C_1}. \qquad \square$$

As in Section 2, we have the application to the radially symmetric situation.

The Cauchy problem

$$\left.\begin{array}{l} U_{tt} - \mu\Delta U - (\lambda+\mu)\nabla\nabla' U + \beta\nabla\theta = 0, \\ \theta_t + \gamma\nabla' q + \delta\nabla' U_t = 0, \\ \tau_0 q_t + q + \kappa\nabla\theta = 0, \end{array}\right\} \tag{3.48}$$

where $t \geq 0, x \in \mathbb{R}^3$, with prescribed initial data, can be treated as in the classical case, by decomposing U into its curl-free part U^{po} and its divergence-free part U^{so}, $U = U^{po} + U^{so}$. Then U^{so} again satisfies a wave equation as in (2.117), and (U^{po}, θ, q) satisfies ($\alpha := \lambda + 2\mu$)

$$\left.\begin{array}{l} U_{tt}^{po} - \alpha\nabla\nabla' U^{po} + \beta\nabla\theta = 0, \\ \tau_0\theta_t + \gamma\nabla' q + \delta\nabla' U_t^{po} = 0, \\ q_t + q + \kappa\nabla\theta = 0. \end{array}\right\} \tag{3.49}$$

Then, with methods as in the one-dimensional case: diagonalizing in Fourier space the principal terms corresponding to low, high, and middle frequencies, and obtaining the asymptotic behavior of the characteristic roots, the parabolic-like decay for the latter system can be obtained, yielding, for example (cf. Section 2.3).

THEOREM 3.4. *The solution (U^{po}, θ, q) satisfies*

$$\begin{aligned} &\|(\partial_t U^{po}, \nabla U^{po}, \theta, q)(t)\|_q \\ &\quad \leq C(1+t)^{-\frac{(3)}{2}(1-2/q)} \|(\partial_t U^{po}, \nabla U^{po}, \theta, q)(0)\|_{m_p,p}, \end{aligned}$$

where $C > 0$ is independent of t and of the initial data, and $2 \leq q \leq \infty$, $1/p + 1/q = 1$, $m_p \geq 3(1-2/q)$.

We note that the detailed asymptotics in exterior domains (complements of compact sets) has not yet been treated. First steps with low-frequency asymptotics of resolvents will be given in [77]; there $\tau_0 = 0$, i.e., the classical case is also considered.

A detailed comparison of the linear system of second sound to that of classical thermoelasticity has been given in [36]. Let Ω be any domain in $\mathbb{R}^n$, $n = 1, 2, 3$, smoothly bounded. We consider the initial-boundary value problem (3.17)–(3.21) again and denote for $\tau_0 > 0$ the solution by $(U^\tau, \theta^\tau, q^\tau)$, writing in the rest of this section $\tau := \tau_0$ for simplicity. By $(\tilde{U}, \tilde{\theta}, \tilde{q})$ we denote the solution for $\tau = 0$, i.e., for classical thermoelasticity. Of course, $\tilde{q} = -\kappa\nabla\tilde{\theta}$, and initial compatibility $\tilde{q}_0 = -\kappa\tilde{\theta}_0$ is assumed.

The energy terms of order $j \in \mathbb{N}$ are defined for $\tau \geq 0$ by

$$E_\tau^j(t) := E_\tau[\partial_t^{j-1} U, \partial_t^{j-1}\theta, \partial_t^{j-1} q](t),$$

where

$$E_\tau[U,\theta,q](t) := \frac{1}{2}\int_\Omega \left(|U_t|^2 + \mu|\nabla U|^2 + (\lambda+\mu)|\nabla' U|^2 + \frac{\beta}{\delta}|\theta|^2 \right.$$
$$\left. + \frac{\beta\tau\gamma}{\delta\kappa}|q|^2 \right)(t,x)\mathrm{dx}.$$

If

$$\varepsilon_0^j(t) := \begin{cases} 2t\sqrt{E_0^j(0)E_0^{j+1}(0)} - t^2 E_0^{j+1}(0), & \text{for } 0 \le t \le \sqrt{\dfrac{E_0^j(0)}{E_0^{j+1}(0)}}, \\ E_0^j(0), & \text{otherwise,} \end{cases}$$

and

$$\mathrm{d}E_\tau^j(t) := E_\tau[\partial_t^{j-1}(U^\tau - \tilde{U}), \partial_t^{j-1}(\theta^\tau - \tilde{\theta}), \partial_t^{j-1}(q^\tau - \tilde{q})](t)$$

then we have the following general comparison of the systems with $\tau > 0$ and $\tau = 0$, respectively, in terms of the energy of the difference of the solutions, $\mathrm{d}E_\tau(t)$. This can be obtained by differentiating $E_\tau^j(t)$ and using the monotonicity.

THEOREM 3.5. *For all $j \in \mathbb{N}$ and $t \ge 0$ we have*

$$\mathrm{d}E_\tau^j(t) \le \frac{\tau^2}{4}\varepsilon_0^{j+1}(t) + \mathrm{d}E_\tau^j(0).$$

In particular for $j = 1, 2$:

$$\mathrm{d}E_\tau^1(t) \le \frac{\tau^2}{4}\varepsilon_0^2(t), \qquad \mathrm{d}E_\tau^2(t) \le \frac{\tau^2}{4}\varepsilon_0^3(t) + \sqrt{E_0^2(0)E_0^3(0)}.$$

Moreover, a time-independent estimate is given by

$$\mathrm{d}E_\tau^j(t) \le \frac{\tau^2}{4}E_0^{j+1}(0) + \mathrm{d}E_\tau^j(0).$$

It is easy to get a time-independent estimate of order τ^2 for exponentially stable situations as in bounded intervals in one dimension or in the bounded radially symmetric case, cf. [93,94]. But there the exponential stability is used, and the constants are not given explicitly. We shall finally use the estimates to get results for real materials below. Moreover, the estimates here carry over to anisotropic situations and can be put into a more general framework then applying, for example, to thermoelastic plates.

For the application, we consider now the systems where a right-hand side is given in (3.19)

$$\tau q_t + q + \kappa\theta_x = f, \tag{3.50}$$

where the support of f is supposed to be in $[t_I, t_F] \times \Omega$ together with zero initial values,

$$U(0,\cdot) = 0, \qquad U_t(0,\cdot) = 0, \qquad \theta(0,\cdot) = 0, \qquad q(0,\cdot) = 0. \tag{3.51}$$

THEOREM 3.6. *For all $j \in \mathbb{N}$ and $t \geq 0$ we have*

$$\mathrm{d}E_\tau^j(t) \leq \frac{\tau^2}{4} \int_{t_I}^{t_F} \int_\Omega \frac{\beta\gamma}{\delta\kappa} |\partial_t^j f|^2(s,x)\mathrm{d}x\mathrm{d}s.$$

As an application we consider the model arising in laser cleaning of silicon wafers. Here $\Omega = (0, L) \subset \mathbb{R}^1$, L denoting the length in metres. We have the following parameter values.

α	β	γ	δ	κ	τ	L
$9.62 \cdot 10^7$	392	$5.98 \cdot 10^{-7}$	164	148	10^{-12}	$6.25 \cdot 10^{-4}$

The laser pulse is modeled by

$$f(t,x) := I_0 \cdot \mathrm{e}^{-A_0 x} \cdot j_T(t - \tau_1 T),$$

where

$$j_T(t) := \exp\left(-\frac{-T^2}{1 - \frac{t^2}{2\tau_1^2 T^2}} + T^2 \right)$$

for $t \in (-\sqrt{2}\tau_1 T, \sqrt{2}\tau_1 T)$, and zero otherwise, and

I_0	A_0	τ_1	T
$2 \cdot 10^{11}$	$1.04 \cdot 10^6$	10^{-8}	10

Then the following estimates show for this example the interesting result that the differences of the systems for $\tau > 0$ and $\tau = 0$, respectively, are experimentally in part not relevant (values for the displacement in metres, for the temperature in degrees Kelvin).

COROLLARY 3.7. *For any $t \geq 0$ we have*

$$\|(U^\tau - \tilde{U})(t,\cdot)\|_\infty \leq 1.16 \cdot 10^{-10}, \qquad \|(U_x^\tau - \tilde{U}_x)(t,\cdot)\|_\infty \leq 1.43 \cdot 10^{-5},$$
$$\|(U_t^\tau - \tilde{U}_t)(t,\cdot)\|_\infty \leq 1.43 \cdot 10^{-2},$$

and for $t \geq t_F$ (following the laser pulse),

$$\frac{1}{L}\int_0^L |\theta^\tau - \tilde{\theta}(t,\cdot)|\mathrm{d}x \leq 2.36 \cdot 10^{-3}, \qquad \|(\theta^\tau - \tilde{\theta})(t,\cdot)\|_\infty \leq 1.74 \cdot 10^0.$$

The behavior of discontinuous solutions has also been investigated, as well as the behavior of discontinuities as the relaxation parameter τ tends to zero. We summarize the results in the three-dimensional case. In studying expansions with respect to the relaxation parameter of the jumps of the potential part of the system on the evolving characteristic surfaces, we obtain that the jump of the temperature goes to zero, while the jumps of the heat flux and the gradient of the potential part of the elastic wave are propagated along the characteristic curves of the elastic fields when the relaxation parameter goes to zero. An interesting phenomenon is that, when time goes to infinity, the behavior will depend on the mean curvature of the initial surface of discontinuity. These jumps decay exponentially when time goes to infinity, more rapidly for smaller heat conductive coefficients.

We consider the homogeneous isotropic case again. Since the solenoidal part of the displacement vector, and of the heat flux, satisfy the classical wave equation and a simple ordinary differential equation, respectively, we directly look at the coupled potential part, i.e.,

$$U_{tt}^{po} - \alpha^2 \Delta U^{po} p + \beta \nabla \theta = 0, \tag{3.52}$$

$$\theta_t + \gamma \nabla' q^{po} + \delta \nabla' U_t^{po} = 0, \tag{3.53}$$

$$\tau q_t^{po} + q^{po} + \kappa \nabla \theta = 0, \tag{3.54}$$

$$\begin{aligned} &U^{po}(0,\cdot) = U_0^{po}, \qquad U_t^{po}(0,\cdot) = U_1^{po}, \\ &\theta(0,\cdot) = \theta^0, \qquad q^{po}(0,\cdot) = q_0^{po}, \end{aligned} \tag{3.55}$$

where $\alpha := \sqrt{2\mu + \lambda}$. The data $(\nabla U^0, U^1, \theta^0, q^0)$ are assumed to be smooth away from, and possibly having jumps on, a given smooth surface

$$\sigma = \{x \in \Omega_0 \subset \mathbb{R}^3 \mid \Phi^0(x) = 0\}$$

described by a C^2-function $\Phi^0 : \Omega_0 \subset \mathbb{R}^3 \to \mathbb{R}$, with Ω_0 open, satisfying

$$|\nabla \Phi^0(x)| = 1 \quad \text{on } \sigma. \tag{3.56}$$

Assumption (3.56) is important for the application of the following techniques, but is made without loss of generality, since for a surface σ, one can choose the distance function $\Phi^0(x) := \operatorname{dist}(x, \sigma)$. Let

$$\tilde{U} = (\tilde{U}_1, \tilde{U}_2, \tilde{U}_3, \tilde{U}_4)' := (\nabla' U^{po}, U_t^{po}, \theta, q^{po})'.$$

Then (3.52)–(3.55) turns into the following system for $\tilde{U}$:

$$\partial_t \tilde{U} + \sum_{j=1}^{3} A_j \partial_j \tilde{U} + A_0 \tilde{U} = 0, \qquad \tilde{U}(0,\cdot) = \tilde{U}^0, \tag{3.57}$$

where $\partial_j = \frac{\partial}{\partial x_j}$, $j = 1, 2, 3$,

$$\sum_{j\equiv 1}^{3} A_j \partial_j \equiv \begin{pmatrix} 0 & -\nabla' & 0 & 0 \\ -\alpha^2\nabla & 0 & \beta\nabla & 0 \\ 0 & \delta\nabla' & 0 & \gamma\nabla' \\ 0 & 0 & \frac{\kappa}{\tau}\nabla & 0 \end{pmatrix}, \qquad A_0 := \begin{pmatrix} 0_{5\times 5} & 0_{5\times 3} \\ 0_{3\times 5} & \frac{1}{\tau}\mathrm{Id}_{\mathbb{R}^3} \end{pmatrix}. \tag{3.58}$$

The characteristic polynomial

$$\det\left(\lambda \mathrm{Id}_{\mathbb{R}^8} - \sum_{j=1}^{3} \xi_j A_j\right)$$
$$= \lambda^4 \left(\lambda^4 - \lambda^2 \left(\alpha^2 + \beta\delta + \frac{\kappa\gamma}{\tau}\right) |\xi|^2 + \frac{\kappa\gamma\alpha^2}{\tau} |\xi|^4\right) \tag{3.59}$$

has real roots λ_k, $1 \le k \le 8$, and, under the assumption (3.56), we have that the eigenvalues of

$$B := \sum_{j=1}^{3} \partial_j \Phi^0 A_j \tag{3.60}$$

are

$$\lambda_k = 0, \quad 1 \le k \le 4, \tag{3.61}$$

$$\lambda_k = \mu_k = \pm \left\{ \frac{1}{2}\left(\alpha^2 + \beta\delta + \frac{\kappa\gamma}{\tau}\right) \pm \frac{1}{2}\sqrt{\left(\alpha^2 + \beta\delta + \frac{\kappa\gamma}{\tau}\right)^2 - \frac{4\kappa\gamma\alpha^2}{\tau}} \right\}^{\frac{1}{2}}, \quad 5 \le k \le 8 \tag{3.62}$$

taking

$$\mu_8 < \mu_6 < 0 < \mu_7 < \mu_5. \tag{3.63}$$

The characteristic surfaces $\Sigma_k = \{(t, x) \mid \Phi_k(t, x) = 0\}$, $1 \le k \le 8$, evolving from the initial surface $\sigma = \{(0, x) \mid \Phi^0(x) = 0\}$, are determined by

$$\partial_t \Phi_k + \mu_k |\nabla \Phi_k| = 0, \quad \Phi_k(0, \cdot) = \Phi^0 \tag{3.64}$$

hence

$$\Sigma_k = \{(t, x) \mid -\mu_k t + \Phi^0(x) = 0\}. \tag{3.65}$$

Let $\nabla_{\tau_1}\Phi^0$ and $\nabla_{\tau_2}\Phi^0$ denote two orthogonal vectors on the tangential plane to the initial surface σ. The associated right (column) eigenvectors r_k and left (row) eigenvectors l_k of B, with the normalization condition

$$l_j r_k = \begin{cases} 1, & j = k, \\ 0, & j \neq k, \end{cases}$$

can be computed explicitly, as well as expansions of these and the eigenvalues in powers of τ. For $k = 5, 6$ we have the expansion, as $\tau \to 0$,

$$\begin{aligned} \mu_k^2 = \lambda_k^2 &= \frac{1}{2}\left(\alpha^2 + \beta\delta + \frac{\kappa\gamma}{\tau} + \sqrt{\left(\alpha^2 + \beta\delta + \frac{\kappa\gamma}{\tau}\right)^2 - \frac{4\kappa\gamma\alpha^2}{\tau}}\right) \\ &= \frac{\kappa\gamma}{\tau} + \beta\delta + \frac{\alpha^2\beta\delta}{\kappa\gamma}\tau + \mathcal{O}(\tau^2), \end{aligned} \tag{3.66}$$

and for $k = 7, 8$

$$\begin{aligned} \mu_k^2 = \lambda_k^2 &= \frac{1}{2}\left(\alpha^2 + \beta\delta + \frac{\kappa\gamma}{\tau} - \sqrt{\left(\alpha^2 + \beta\delta + \frac{\kappa\gamma}{\tau}\right)^2 - \frac{4\kappa\gamma\alpha^2}{\tau}}\right) \\ &= \alpha^2 - \frac{\alpha^2\beta\delta}{\kappa\gamma}\tau + \mathcal{O}(\tau^2). \end{aligned} \tag{3.67}$$

Let the matrices L and R be given by

$$L := \begin{pmatrix} l_1 \\ \vdots \\ l_8 \end{pmatrix}, \qquad R := (r_1, \dots, r_8).$$

Then $V := L\tilde{U}$, with $\tilde{U}$ satisfying (3.57), satisfies

$$\partial_t V + \sum_{j=1}^{3}(LA_jR)\partial_j V + \widetilde{A}_0 V = 0, \qquad V(t=0) = V^0 := L\tilde{U}^0, \tag{3.68}$$

where

$$\widetilde{A}_0 := LA_0R + \sum_{j=1}^{3} LA_j\partial_j R. \tag{3.69}$$

Let $[H]_{\Sigma_k}$ denote the jump of H along Σ_k. Then $(V_1, V_2, V_3.V_4)$ are continuous at $\cup_{k=5}^{8}\Sigma_k$, and V_j, $j = 5, \dots, 8$, does not have any jump on Σ_k, $k = 1, \dots, 8$ for $k \neq j$. Moreover, $[V_k]_{\Sigma_0} = 0$ for $1 \le k \le 4$.

Now we turn to the interesting evolutional equations of $[V_k]_{\Sigma_k}$ for $5 \le k \le 8$. Denote by D_{mk} the mth row and kth column element of a (8×8)-matrix D. Obviously, the kth equation given in (3.68) can be explicitly written as

$$\partial_t V_k + \sum_{m=1}^{8}\sum_{j=1}^{3}(LA_jR)_{km}\partial_j V_m = -\sum_{m=1}^{8}(\widetilde{A}_0)_{km}V_m + \widetilde{F}_k \tag{3.70}$$

for $5 \le k \le 8$. Since

$$\left(\lambda_k \mathrm{Id}_{\mathbb{R}^8} - \sum_{j=1}^{3} L\partial_j\Phi^0 A_j R\right)_{kk} = (\lambda_k \mathrm{Id}_{\mathbb{R}^8} - \widetilde{\Lambda})_{kk} = 0 \tag{3.71}$$

with $\widetilde{\Lambda} = \mathrm{diag}(\lambda_1, \dots, \lambda_8)$, we know that the operator $\partial_t + \sum_{j=1}^{3}(LA_jR)_{kk}\partial_j$ is tangential to $\Sigma_k = \{\lambda_k t + \Phi^0(x) = 0\}$. Similarly, since all entries in line k of $\lambda_k \mathrm{Id}_{\mathbb{R}^8} - \sum_{j=1}^{3} L\partial_j\Phi^0 A_j R = \lambda_k \mathrm{Id}_{\mathbb{R}^8} - \widetilde{\Lambda}$ vanish, we deduce that for $m \neq k$, $(0, (LA_1R)_{km}, (LA_2R)_{km}, (LA_3R)_{km})$ is orthogonal to the normal direction $(-\lambda_k, (\nabla\Phi^0)')$ of Σ_k, which implies that $\sum_{j=1}^{3}(LA_jR)_{km}\partial_j$ is tangential to Σ_k when $m \neq k$.

Therefore we obtain that for each $5 \le k \le 8$, $[V_k]_{\Sigma_k}$ satisfies the following transport equation:

$$\left(\partial_t + \sum_{j=1}^{3}(LA_jR)_{kk}\partial_{x_j} + (\widetilde{A}_0)_{kk}\right)[V_k]_{\Sigma_k} = [\tilde{F}_k]_{\Sigma_k} \tag{3.72}$$

with initial conditions

$$[V_k]_{\Sigma_k}(t=0) = [V_k^0]_\sigma. \tag{3.73}$$

In order to determine the behavior of $[V_k]_{\Sigma_k}$ from (3.72), (3.73), it is essential to study $(\widetilde{A}_0)_{kk}$, where, by (3.69), $\widetilde{A}_0$ is given by

$$\widetilde{A}_0 = LA_0R + \sum_{j=1}^{3} LA_j\partial_j R. \tag{3.74}$$

Denoting the row l_k and the column r_k by $l_k = (l_{k1}, \dots, l_{k8})$ resp. $r_k = (r_{1k}, \dots, r_{8k})'$ we have, using (3.58), for $k = 5, \dots, 8$

$$(LA_0R)_{kk} = \frac{1}{\tau}\sum_{j=6}^{8} l_{kj}r_{jk}$$

hence

$$(LA_0R)_{kk} = \frac{c_k\kappa\gamma}{\tau^2\lambda_k^2}.$$

This implies for $k = 5, 6$, using (3.66),

$$\begin{aligned}(LA_0R)_{kk} &= \left(\frac{1}{2} + \mathcal{O}(\tau^2)\right)\frac{\kappa\gamma}{\tau^2(\frac{k\gamma}{\tau} + \beta\delta + \frac{\alpha^2\beta\delta}{\kappa\gamma}\tau + \mathcal{O}(\tau^2))}\\ &= \frac{1}{2\tau} - \frac{\beta\delta}{2\kappa\gamma} + \mathcal{O}(\tau),\end{aligned} \tag{3.75}$$

and for $k = 7, 8$ using (3.67),

$$\begin{aligned}(LA_0R)_{kk} &= \left(\frac{\alpha^2\beta\delta}{2\kappa^2\gamma^2}\tau^2 + \mathcal{O}(\tau^3)\right)\frac{\kappa\gamma}{\tau^2(\alpha^2 - \frac{\alpha^2\beta\delta}{\kappa\gamma}\tau + \mathcal{O}(\tau^2))}\\ &= \frac{\beta\delta}{2\kappa\gamma} + \mathcal{O}(\tau).\end{aligned} \tag{3.76}$$

Now we compute $(\sum_{j=1}^{3} LA_j\partial_j R)_{kk}$. Let $\vec{e}_j$ be the jth standard unit vector having only zero components besides a 1 in the jth component. Then

$$A_1 = \begin{pmatrix} 0 & -\vec{e}_1{}' & 0 & 0\\ -\alpha^2\vec{e}_1 & 0 & \beta\vec{e}_1 & 0\\ 0 & \delta\vec{e}_1{}' & 0 & \gamma\vec{e}_1{}'\\ 0 & 0 & \frac{\kappa}{\tau}\vec{e}_1 & 0\end{pmatrix}$$

and A_2, A_3 arise from A_1 by replacing $\vec{e}_1$ by $\vec{e}_2$ and $\vec{e}_3$, respectively.

Denote by $LA_j\partial_j R = (a_{ik}^j)_{8\times 8}$ for any $j = 1, 2, 3$. Then, by a direct computation, we deduce that for $k = 5, \ldots, 8$,

$$\begin{aligned}a_{kk}^1 &= -l_{k1}\partial_1 r_{2k} + l_{k2}(\beta\partial_1 r_{5k} - \alpha^2\partial_1 r_{1k}) + l_{k5}(\delta\partial_1 r_{2k} + \gamma\partial_1 r_{6k})\\ &\quad + l_{k6}\frac{\kappa}{\tau}\partial_1 r_{5k}\\ a_{kk}^2 &= -l_{k1}\partial_2 r_{3k} + l_{k3}(\beta\partial_2 r_{5k} - \alpha^2\partial_2 r_{1k}) + l_{k5}(\delta\partial_2 r_{3k} + \gamma\partial_2 r_{7k})\\ &\quad + l_{k7}\frac{\kappa}{\tau}\partial_2 r_{5k}\\ a_{kk}^3 &= -l_{k1}\partial_3 r_{4k} + l_{k4}(\beta\partial_3 r_{5k} - \alpha^2\partial_3 r_{1k}) + l_{k5}(\delta\partial_3 r_{4k} + \gamma\partial_3 r_{8k})\\ &\quad + l_{k8}\frac{\kappa}{\tau}\partial_3 r_{5k}\end{aligned}$$

which gives

$$\left(\sum_{j=1}^{3} LA_j\partial_j R\right)_{kk} = -l_{k1}(\partial_1 r_{2k} + \partial_2 r_{3k} + \partial_3 r_{4k})$$
$$+ l_{k5}\left(\delta(\partial_1 r_{2k} + \partial_2 r_{3k} + \partial_3 r_{4k}) + \gamma(\partial_1 r_{6k} + \partial_2 r_{7k} + \partial_3 r_{8k})\right).$$

We conclude

$$\left(\sum_{j=1}^{3} LA_j\partial_j R\right)_{kk} = \left((-\delta l_{k5} + l_{k1})\frac{\beta\lambda_k}{\alpha^2 - \lambda_k^2} + l_{k5}\frac{\kappa\gamma}{\tau\lambda_k}\right)\Delta\Phi^0, \tag{3.77}$$

where Δ denotes the Laplace operator in $\mathbb{R}^3$. Since

$$\delta l_{k5} - l_{k1} = c_k\left(\delta - \frac{\alpha^2\delta}{\alpha^2 - \lambda_k^2}\right) = -c_k\frac{\delta\lambda_k^2}{\alpha^2 - \lambda_k^2}$$

we get from (3.77)

$$\left(\sum_{j=1}^{3} LA_j\partial_j R\right)_{kk} = c_k\left(\frac{\beta\delta\lambda_k^3}{(\alpha^2 - \lambda_k^2)^2} + \frac{\kappa\gamma}{\tau\lambda_k}\right)\Delta\Phi^0. \tag{3.78}$$

Expanding we obtain for $k = 5, 6$

$$c_k\left(\frac{\beta\delta\lambda_k^3}{(\alpha^2 - \lambda_k^2)^2} + \frac{\kappa\gamma}{\tau\lambda_k}\right) = \left(\frac{1}{2} + \mathcal{O}(\tau^2)\right)\frac{1}{\lambda_k}\left(\frac{\kappa\gamma}{\tau} + \beta\delta + \mathcal{O}(\tau)\right),$$
$$\frac{1}{\lambda_k} = \pm\sqrt{\frac{\tau}{\kappa\gamma}}(1 + \mathcal{O}(\tau))$$

implying

$$\left(\sum_{j=1}^{3} LA_j\partial_j R\right)_{kk} = \pm\frac{1}{2}\sqrt{\frac{\kappa\gamma}{\tau}}\,(1 + \mathcal{O}(\tau))\,\Delta\Phi^0, \quad k = 5, 6. \tag{3.79}$$

Similarly we obtain

$$\left(\sum_{j=1}^{3} LA_j\partial_j R\right)_{kk} = \pm\left(\frac{\alpha}{2} + \mathcal{O}(\tau)\right)\Delta\Phi^0, \quad k = 7, 8. \tag{3.80}$$

Combining (3.74)–(3.76) and (3.79) with (3.80), we obtain

$$(\widetilde{A}_0)_{kk} = \begin{cases} \dfrac{1}{2\tau} \pm \dfrac{1}{2}\sqrt{\dfrac{\kappa\gamma}{\tau}}\Delta\Phi^0 + \mathcal{O}(1), & k = 5, 6, \\ \dfrac{\beta\delta}{2\kappa\gamma} \pm \dfrac{\alpha}{2}\Delta\Phi^0 + \mathcal{O}(\tau), & k = 7, 8. \end{cases} \tag{3.81}$$

In (3.81) we already notice the phenomenon that the mean curvature H of the initial surface σ,

$$H = \frac{\Delta\Phi^0}{2}$$

will play an essential role in the behavior of the jumps as $t \to \infty$ or as $\tau \to 0$.

For any fixed $x^0 \in \sigma$, denote by

$$t \to (t, x_1(t; 0, x^0), x_2(t; 0, x^0), x_3(t; 0, x^0))$$

the characteristic line of the operator $\partial_t + \sum_{j=1}^{3}(LA_jR)_{kk}\partial_{x_j}$ passing through $(0, x^0)$, i.e., $x_j(t; 0, x^0)$ satisfies

$$\begin{cases} \dfrac{\partial x_j(t; 0, x^0)}{\partial t} = (LA_jR)_{kk}(t, x_1(t; 0, x^0), x_2(t; 0, x^0), x_3(t; 0, x^0)) \\ x_j(0; 0, x^0) = x_j^0, \quad 1 \le j \le 3 \end{cases}$$

and

$$[V_k]_{\Sigma_k(t)} = [V_k]_{\Sigma_k}(t, x_1(t; 0, x^0), x_2(t; 0, x^0), x_3(t; 0, x^0)). \tag{3.82}$$

Then, from (3.72) and (3.81) we conclude:

$$\begin{aligned} [V_k]_{\Sigma_k(t)} &= [V_k^0]_\sigma \mathrm{e}^{-\int_0^t (\widetilde{A}_0)_{kk}(x(s;0,x^0))\mathrm{d}s} \\ &= \begin{cases} \mathrm{e}^{-\frac{t}{2\tau} \mp \frac{1}{2}\sqrt{\frac{\kappa\gamma}{\tau}}\int_0^t (\Delta\Phi^0(x(s;0,x^0)) + O(\sqrt{\tau}))\mathrm{d}s}[V_k^0]_\sigma, & k = 5, 6, \\ \mathrm{e}^{-\frac{\beta\delta}{2\kappa\gamma}t \mp \frac{\alpha}{2}\int_0^t (\Delta\Phi^0(x(s;0,x^0)) + \mathcal{O}(\tau))\mathrm{d}s}[V_k^0]_\sigma, & k = 7, 8. \end{cases} \end{aligned} \tag{3.83}$$

For $\tau \to 0$ the dominating term for $k = 5, 6$ is $\mathrm{e}^{-\frac{t}{2\tau}}$, i.e., we have exponential decay of the jumps of V_k on Σ_k as $\tau \to 0$ or $t \to \infty$ for a fixed small $\tau > 0$. If $k = 7, 8$ the dominating term, for $\tau \to 0$, is $\exp(-\int_0^t (\frac{\beta\delta}{2\kappa\gamma} \pm \frac{\alpha}{2}\Delta\Phi^0(x(s; 0, x^0)))\mathrm{d}s)$, whether the jumps of V_k on Σ_k decay exponentially depends on the size of the mean curvature $(= \Delta\Phi^0/2)$.

EXAMPLE. Let σ be the sphere of radius r:

$$\sigma = \{x \in \mathbb{R}^3 \mid |x| = r\} = \{x \mid \Phi^0(x) \equiv r - |x| = 0\}.$$

Then, we have

$$\Sigma_k = \{(t,x) \mid \mu_k t = r - |x|\} = \{(t,x) \mid |x| = r - \mu_k t\}.$$

We recall the convention on the signs of μ_k given in (3.63). Spreading surfaces, as $t \to \infty$, are Σ_6, Σ_8, and

$$\Delta\Phi^0(x_0) = \frac{2}{|x_0|} = \frac{2}{r} > 0.$$

Thus, as $t \to +\infty$, $[V_6]_{\Sigma_6}$ is decaying exponentially, while $[V_8]_{\Sigma_8}$ decays (grows resp.) exponentially if

$$\frac{\beta\delta}{\alpha\kappa\gamma} > \Delta\Phi^0 = \frac{2}{r} \qquad \left(\frac{\beta\delta}{\alpha\kappa\gamma} < \Delta\Phi^0 = \frac{2}{r} \text{ resp.}\right), \tag{3.84}$$

that is, depending on the size of the mean curvature $H = \frac{1}{r}$.

To complete the discussion on the linear problem, we have to compute $U = RV$ from V,

$$U = (\nabla' U^{po}, U_t^{po}, \theta, q^p)' = RV = \sum_{k=1}^{8} V_k r_k.$$

We obtain

$$[\nabla' U^{po}]_{\Sigma_k} = \begin{cases} \mathcal{O}(\tau)\mathrm{e}^{-\frac{t}{2\tau}(1+\mathcal{O}(\sqrt{\tau}))}, & k=5,6,\\ \frac{1}{2\alpha}\left(\alpha[\nabla' U_0^{po}]_\sigma \mp \nabla\Phi^0 \cdot [U_1^{po}]_\sigma + \mathcal{O}(\tau)\right) & \\ \quad \cdot\, \mathrm{e}^{-\int_0^t (\frac{\beta\delta}{2\kappa\gamma} \pm \frac{\alpha}{2}\Delta\Phi^0(x(s;0,x^0)) + \mathcal{O}(\tau))\mathrm{d}s}, & k=7,8, \end{cases} \tag{3.85}$$

$$[U_t^{po}]_{\Sigma_k} = \begin{cases} \mathcal{O}(\sqrt{\tau})\mathrm{e}^{-\frac{t}{2\tau}(1+\mathcal{O}(\sqrt{\tau}))}, & k=5,6,\\ \mp\frac{1}{2}\left(\alpha[\nabla' U_0^{po}]_\sigma \mp \nabla\Phi^0 \cdot [U_1^{po}]_\sigma + \mathcal{O}(\tau)\right) & \\ \quad \cdot\, \mathrm{e}^{-\int_0^t (\frac{\beta\delta}{2\kappa\gamma} \pm \frac{\alpha}{2}\Delta\Phi^0(x(s;0,x^0)) + \mathcal{O}(\tau))\mathrm{d}s}, & k=7,8, \end{cases} \tag{3.86}$$

$$[\theta]_{\Sigma_k} = \begin{cases} \frac{1}{2}([\theta_0]_\sigma + \mathcal{O}(\sqrt{\tau}))\mathrm{e}^{-\frac{t}{2\tau}(1+\mathcal{O}(\sqrt{\tau}))}, & k=5,6,\\ \mathcal{O}(\tau)\mathrm{e}^{-\int_0^t (\frac{\beta\delta}{2\kappa\gamma} \pm \frac{\alpha}{2}\Delta\Phi^0(x(s;0,x^0)) + \mathcal{O}(\tau))\mathrm{d}s}, & k=7,8, \end{cases} \tag{3.87}$$

$$[q^p]_{\Sigma_k} = \begin{cases} \left(\pm\frac{1}{2}\sqrt{\frac{\kappa}{\gamma\tau}}[\theta_0]_\sigma \nabla\Phi^0 + \frac{1}{2}\nabla\Phi^0([q^{0,p} + \frac{\delta}{\gamma}U_1^{po}]_\sigma \right. & \\ \quad \left. \cdot\, \nabla\Phi^0) + \mathcal{O}(\sqrt{\tau})\right)\mathrm{e}^{-\frac{t}{2\tau}(1+\mathcal{O}(\sqrt{\tau}))}, & k=5,6,\\ \pm\frac{\delta}{2\gamma}\left((\alpha[\nabla' U_0^{po}]_\sigma \mp \nabla\Phi^0 \cdot [U_1^{po}]_\sigma)\nabla\Phi^0 + \mathcal{O}(\tau)\right) & \\ \quad \cdot - \mathrm{e}^{-\int_0^t \left(\frac{\beta\delta}{2\kappa\gamma} \pm \frac{\alpha}{2}\Delta\Phi^0(x(s;0,x^0)) + \mathcal{O}(\tau)\right)\mathrm{d}s}, & k=7,8. \end{cases} \tag{3.88}$$

Altogether, from the above discussion we conclude

THEOREM 3.8. *Suppose that the initial data $\nabla' U^{po}, \partial_t U^{po}, \theta$ and q^{po} may have jumps on $\sigma = \{\Phi^0(x) = 0\}$ with $|\nabla\Phi^0(x)| = 1$, then the propagation of strong singularities of solutions to the linearized problem (3.52)–(3.55) is described by (3.85)–(3.88). In particular, we have*

(1) *The jumps of $\nabla' U^{po}, \partial_t U^{po}, \theta, q^{po}$ on Σ_5 and Σ_6 decay exponentially, both when $\tau \to 0$ for a fixed $t > 0$, and when $t \to +\infty$ for a fixed $\tau > 0$.*
(2) *The jumps of $\nabla' U^{po}, \partial_t U^{po}, q^{po}$ on Σ_7 (Σ_8 resp.) are propagated, and when $t \to +\infty$ they will decay exponentially when $\frac{\beta\delta}{\kappa\gamma} + \alpha\Delta\Phi^0$ ($\frac{\beta\delta}{\kappa\gamma} - \alpha\Delta\Phi^0$ resp.) are positive, more rapidly for the smaller heat conductive coefficient $\kappa\gamma$, while the jump of the temperature θ on Σ_7 and Σ_8 of order $O(\tau)$ vanishes when $\tau \to 0$, which shows a smoothing effect in the system (3.52)–(3.54) when the thermoelastic model with second sound converges to the hyperbolic–parabolic type of classical thermoelasticity.*

Semilinear problems have also been dealt with [100].

Notes: For a survey on hyperbolic heat conduction models see Chandrasekharaiah [8], for general Cattaneo models cf. Öncü and Moodie [79], see also Lord and Shulman [66]. The results for the linear theory on exponential stability in one dimension are taken from Irmscher and Racke [35,36,38,93]. The decay result for the Cauchy problem is from Yang and Wang [122], see also Wang and Wang [118]. For the three-dimensional case we have the exponential stability result from our paper [94], while the discussion of the Cauchy problem is given in [119]. The general comparison results and the application are due to Irmscher [36], and the results on the propagation of singularities is taken from our paper with Wang [100], the one-dimensional case had been treated in [98]. Linear systems with time-dependent coefficients have been treated by Weinmann [120]. A low frequency expansion in exterior domains is given by Naito, Shibata and Racke [77].

3.2. *Nonlinear systems*

Nonlinear systems, as introduced at the beginning of Section 3, have been treated for one-dimensional bounded reference configurations as well as for the Cauchy problem. Two- or three-dimensional situations have been studied for the radially symmetric case, more general situations like the fully nonlinear Cauchy problem is still under investigation.

In one space dimension the general nonlinear differential equations given in (3.6), (3.9) and (3.10) turn into the following,

$$u_{tt} - a(u_x, \theta, q)u_{xx} + b(u_x, \theta, q)\theta_x = r_1(u_x, \theta, q)q_x, \tag{3.89}$$

$$\theta_t + g(u_x, \theta, q)q_x + d(u_x, \theta, q)u_{tx} = r_2(u_x, \theta, q)q_t, \tag{3.90}$$

$$\tau(u_x, \theta)q_t + q + k(u_x, \theta)\theta_x = 0, \tag{3.91}$$

where

$$a := \tilde{S}_{u_x}, \quad b := -\tilde{S}_\theta, \quad r_1 := \tilde{S}_q,$$

$$g := \frac{1}{(\theta + T_0)\tilde{a}}, \qquad d := \frac{-\tilde{S}_\theta}{(\theta + T_0)\tilde{a}}, \qquad r_2 := \frac{\tilde{b}}{(\theta + T_0)\tilde{a}}, \quad (\text{for } |\theta| < T_0).$$

The assumptions on $a, b, r_1, \tilde{a}$ (resp. g), d, r_2, τ, k are such that if they are smooth functions and that there exist positive constants c_1, c_2, c_3, c_4 and $K < T_0$ such that if $|u_x|, |\theta|, |q| \leq K$, then

$$a(u_x, \theta, q) \geq c_1, \qquad \tau(u_x, \theta) \geq c_2, \qquad k(u_x, \theta) \geq c_3, \qquad \tilde{a}(u_x, \theta, q) \geq c_4$$

as well as

$$b(u_x, \theta, q) \neq 0 \neq d(u_x, \theta, q).$$

Moreover, we shall assume

$$r_1 \equiv 0, \qquad r_2 \equiv 0.$$

We notice that this assumption could be removed by Messaoudi and Said-Houari in [67]. Observe that always $r_1(0, 0, 0) = r_2(0, 0, 0) = 0$ in accordance with the linearized equations, compare the previous sections. The coefficients $\alpha, \beta, \gamma, \delta, \tau_0, \kappa$ in the linearized equations are, of course, given by

$$\alpha = a(0, 0, 0), \qquad \beta = b(0, 0, 0), \qquad \gamma = g(0, 0, 0),$$

$$\delta = d(0, 0, 0), \qquad \tau_0 = \tau(0, 0), \qquad \kappa = k(0, 0).$$

In addition to the differential equations, we have initial and boundary conditions

$$u(0, \cdot) = u_0, \quad u_t(0, \cdot) = u_1, \qquad \theta(0, \cdot) = \theta_0, \quad q(0, \cdot) = q_0, \tag{3.92}$$

$$u(t, 0) = u(t, 1) = \theta(t, 0) = \theta(t, 1) = 0, \quad t \geq 0. \tag{3.93}$$

With energy methods again, treating the arising nonlinear terms as perturbations of the energy terms, and observing that the perturbations are of cubic order rather than the quadratic energy terms and hence comparatively small, we obtain

THEOREM 3.9. *There exists a constant* $\varepsilon_0 > 0$ *such that if*

$$\Lambda_0 := \|u_0\|_{H^3}^2 + \|u_1\|_{H^2}^2 + \|\theta_0\|_{H^2}^2 + \|q_0\|_{H^2}^2 < \varepsilon_0$$

then (3.89)–(3.93) *has a unique global solution*

$$u \in \bigcap_{m=0}^{3} C^m([0, \infty), H^{3-m}(\Omega)), \qquad \theta, q \in \bigcap_{m=0}^{2} C^m([0, \infty), H^{2-m}(\Omega));$$

moreover, the system is exponentially stable, that is, there exist constants $d_1, d_2 > 0$ *such that for* $t \geq 0$

$$\Lambda(t) := \sum_{j=0}^{3} \|(\partial_t, \partial_x)^j u(t,\cdot)\|^2 + \sum_{j=0}^{2} \|(\partial_t, \partial_x)^j (\theta, q)(t,\cdot)\|^2 \leq d_1 e^{-d_2} \Lambda_0.$$

The one-dimensional Cauchy problem had already been treated by Tarabek [112], and meanwhile, was augmented by a description of decay rates. We shall assume for the rest of the section that τ and k are constant functions,

$$\tau \equiv \tau_0, \qquad k \equiv \kappa \tag{3.94}$$

that is, Cattaneo's law is the linear one. The reason is that the proofs of the L^2-decay results in classical nonlinear thermoelasticity (Fourier's law) repeatedly used the divergence forms of the differential equations arising from the balance law for the momentum (3.6) and for the energy (3.7). The same structure is not present in the general Cattaneo law. We shall also exploit that the free energy ψ and the internal energy have the form

$$\psi(u_x, \theta, q) = \psi_0(u_x, \theta) + \psi_1(u_x, \theta) q^2, \tag{3.95}$$

$$e(u_x, \theta, q) = e_0(u_x, \theta) + e_1(u_x, \theta) q^2, \tag{3.96}$$

with suitable ψ_0, ψ_1, e_0, e_1, see [79]. We rewrite the system (3.89)–(3.91) first as follows,

$$u_{tt} - \alpha u_{xx} + \beta \theta_x = h_1, \tag{3.97}$$

$$\theta_t + \gamma q_x + \delta u_{tx} = h_2, \tag{3.98}$$

$$\tau_0 q_t + q + \kappa \theta_x = 0, \tag{3.99}$$

where

$$\alpha = a(0,0,0), \quad \beta := b(0,0,), \quad \gamma := g(0,0,0), \quad \delta := d(0,0,0), \tag{3.100}$$

$$h_1 := (\tilde{S}(u_x, \theta, q) - \alpha u_x + \beta \theta)_x, \tag{3.101}$$

$$h_2 := (\gamma - g) q_x + (\delta - d) u_{tx} + \frac{\kappa}{\tau} \theta_x + \frac{r_2}{\tau_0} q, \tag{3.102}$$

$$h_3 := \tau_0 \left\{ \left(\frac{1}{\tau_0} - \frac{1}{\tau} \right) q + \left(\frac{\kappa}{\tau_0} - \frac{k}{\tau} \right) \theta_x \right\}. \tag{3.103}$$

Defining

$$V := (\sqrt{\alpha \kappa \delta} u_x, u_t, \theta, q)'$$

we obtain

$$A^0 V_t + A^1 V_x + B V = \widetilde{F}(V, V_x),$$

$$V(0) = V_0 := (\sqrt{\alpha\kappa\delta}u_{0,x}, u_1, \theta_0, q_0)', \tag{3.104}$$

where

$$A^0 := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \kappa\delta & 0 & 0 \\ 0 & 0 & \beta\kappa & 0 \\ 0 & 0 & 0 & \beta\gamma\tau_0 \end{pmatrix}, \qquad A^1 := \begin{pmatrix} 0 & -\sqrt{\alpha\kappa\delta} & 0 & 0 \\ -\sqrt{\alpha\kappa\delta} & 0 & \beta\kappa\delta & 0 \\ 0 & \beta\kappa\delta & 0 & \beta\kappa\gamma \\ 0 & 0 & \beta\kappa\gamma & 0 \end{pmatrix},$$

$$B := \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta\gamma \end{pmatrix}, \qquad \widetilde{F} := \begin{pmatrix} 0 \\ \kappa\delta h_1 \\ \beta\kappa h_2 \\ 0 \end{pmatrix}.$$

We have from (3.100)–(3.102)

$$\widetilde{F}(0,0) = 0, \qquad \widetilde{F}_{(V,V_x)}(0,0) = 0. \tag{3.105}$$

The linearized system, i.e., for $\widetilde{F} = 0$, is solved by

$$V(t) = \mathrm{e}^{tR} V_0,$$

where

$$R := -(A^0)^{-1}(A^1\partial_x + B)$$

generates a C_0-semigroup on $D(R) := (W^{1,2}(\mathbb{R}))^4 \subset (L^2(\mathbb{R}))^4$. Then the solution to (3.104) is represented in general by

$$V(t) = \mathrm{e}^{tR} V_0 + \int_0^t \mathrm{e}^{(t-r)R} F(V, V_x)(r)\mathrm{d}r, \tag{3.106}$$

where

$$F := (A^0)^{-1}\widetilde{F}. \tag{3.107}$$

The asymptotic behavior of the L^2-norm of the solution to the linearized problem is described by

LEMMA 3.10. *Let $F = 0$. Then there are constants $c, c_1 > 0$ such that for all $t \geq 0$:*

(i)

$$\|\partial_x^l V(t)\|_2 \leq c\left\{\mathrm{e}^{-c_1 t}\|\partial_x^l V_0\|_2 + (1+t)^{-\frac{2l+1}{4}}\|V_0\|_1\right\} \quad (l = 0, 1).$$

(ii)

$$\|\partial_t V(t)\|_2 \le c\left\{e^{-c_1 t}\|V_0\|_{1,2} + (1+t)^{-\frac{3}{4}}\|V_0\|_1\right\}.$$

(iii)

$$\|V_4(t)\|_2 \le c\left\{e^{-c_1 t}\|V_0\|_{1,2} + (1+t)^{-\frac{3}{4}}\|V_0\|_1\right\}.$$

This lemma follows from the results of Yang and Wang [122], in particular from the expansions of the characteristic values; cp. similar results for classical thermoelasticity in [47, Thm.3.6], going back to Kawashima [50].

Now we can state the theorem on the L^2-decay rates. The global well-posedness is given, cf. [112,99].

THEOREM 3.11. *Let $V_0 \in W^{3,2}(\mathbb{R}) \cap L^1(\mathbb{R})$, and $m_0 := \|V_0\|_{3,2} + \|V_0\|_1$. For sufficiently small m_0, there is $c > 0$ such that for $t \ge 0$, the solution V to (3.104) satisfies*

$$\|V(t)\|_{1,2} \le c(1+t)^{-\frac{1}{4}} \cdot m_0.$$

PROOF. With

$$V = (\sqrt{\alpha\kappa\delta}u_x, u_t, \theta, q)$$

we have

$$V_t - RV = F(V, V_x), \qquad F(V, V_x) = (0, h_1, h_2, 0).$$

The fact that F vanishes quadratically in zero, cp. the next lemma, is not sufficient, despite the linear decay behavior being like that of a heat equation, because terms like "$V \cdot V_x$" appear, cf. Zheng and Chen [125]. Therefore, we shall rewrite the nonlinearities in divergence form in order to be able to exploit the better decay of derivatives in the later estimates for $\|V(t)\|_2$, a modification of the corresponding proof in classical thermoelasticity, cp. [47].

LEMMA 3.12. (i)

$$h_1(V, V_x) = (S(u_x, \theta, q) - \alpha u_x + \beta\theta)_x \equiv (h_{11}(V))_x\,.$$

(ii)

$$h_2(V, V_x) = \left(\left[\theta + \delta u_x - \gamma\left\{\varepsilon - \varepsilon(0,0,0) + \frac{u_t^2}{2}\right\}\right]\right)_t + \gamma(Su_t)_x$$
$$\equiv (h_{21}(V))_t + (h_{22}(V))_x.$$

(iii)

$$|h_{11}(V)|, |h_{21}(V)|, |h_{22}(V)| \le c|V|^2.$$

(iv)

$$|h_1(V, V_x)|, |h_2(V, V_x)| \le c(|V|^2 + |V_x|^2) \quad (near\ V = 0).$$

PROOF: For (i) compare (3.101). For (ii) we first have

$$\left(\varepsilon + \frac{u_t^2}{2}\right)_t = (Su_t)_x - q_x. \tag{3.108}$$

It follows

$$\theta_t + \delta u_{xt} + \gamma q_x = \left(\theta + \delta u_x - \gamma\left\{\varepsilon - \varepsilon(0,0,0) + \frac{u_t^2}{2}\right\}\right)_t + \gamma(Su_t)_x$$

which yields (ii) (cp. (3.98)).

$$|h_{11}(V)|, \qquad |h_{22}(V)| \le c|V|^2$$

follows immediately from the definition of h_{11}, h_{22} and from (3.100) and $S(0,0,0) = 0$, which is assumed without loss of generality.

The estimates (iv) also are immediate consequences of (3.100)–(3.102), so it remains to investigate $h_{21}(V)$. Here we have

$$\begin{aligned} h_{21}(V) = {} & \theta + \delta u_x - \gamma\left\{\varepsilon_{u_x}(0,0,0)u_x + \varepsilon_\theta(0,0,0)\theta + \varepsilon_q(0,0,0)q\right\} \\ & + \mathcal{O}(u_x^2 + \theta^2 + q^2 + u_t^2). \end{aligned}$$

Observing

$$\varepsilon_{u_x} = S + (\theta + T_0)(-S_\theta), \quad \varepsilon_\theta = -(\theta + T_0)\psi_{\theta\theta}$$

as well as (3.96) and (3.100) we obtain

$$\gamma\varepsilon_{u_x}(0,0,0) = \delta, \qquad \gamma\varepsilon_\theta(0,0,0) = 1, \qquad \varepsilon_q(0,0,0) = 0,$$

hence

$$h_{21}(V) = \mathcal{O}(u_x^2 + \theta^2 + q^2 + u_t^2) = \mathcal{O}(|V|^2).$$

This proves Lemma 3.12. □

Continuing the proof of Theorem 3.11, we notice that, as a consequence of Lemma 3.12, we may rewrite F as

$$\begin{aligned} F(V, V_x) &= (0, h_1(V, V_x), h_2(V, V_x), 0)' \\ &= \big(0, (h_{11}(V))_x, (h_{21}(V))_t + (h_{22}(V))_x, O\big)'. \end{aligned}$$

Since $W : t \mapsto \mathrm{e}^{(t-r)R}\,(0, 0, (h_{21}(V(r)))\,, 0)'$ solves

$$W_t - RW = 0, \qquad W(r) = (0, 0, h_{21}(V(r)), 0)'$$

we conclude

$$\begin{aligned}
&\mathrm{e}^{(t-r)R}\,(0, 0, (h_{21}(V(r)))_r\,, 0)'\\
&\quad= \partial_r \left(\mathrm{e}^{(t-r)R}\,(0, 0, (h_{21}(V(r)))\,, 0)'\right) + \partial_t \left(\mathrm{e}^{(t-r)R}\,(0, 0, (h_{21}(V(r)))\,, 0)'\right)\\
&\quad= \partial_r \left(\mathrm{e}^{(t-r)R}\,(0, 0, (h_{21}(V(r)))\,, 0)'\right) + R \left(\mathrm{e}^{(t-r)R}\,(0, 0, (h_{21}(V(r)))\,, 0)'\right)\\
&\quad= \partial_r \left(\mathrm{e}^{(t-r)R}\,(0, 0, (h_{21}(V(r)))\,, 0)'\right)\\
&\qquad - (A^0)^{-1} A^1 \partial_x \mathrm{e}^{(t-r)R}\,(0, 0, (h_{21}(V(r)))\,, 0)'\\
&\qquad - (A^0)^{-1} B \mathrm{e}^{(t-r)R}\,(0, 0, (h_{21}(V(r)))\,, 0)'\,.
\end{aligned} \tag{3.109}$$

By the representation (3.106) we get, using (3.109),

$$\begin{aligned}
V(t) = \mathrm{e}^{tR} V_0 + \int_0^t &\Big\{\partial_x \mathrm{e}^{(t-r)R}(0, h_{11}, h_{22}, 0)'\\
&- (A^0)^{-1} A^1 \partial_x \mathrm{e}^{(t-R)}(0, 0, h_{21}, 0)' - (A^0)^{-1} B \mathrm{e}^{(t-r)R}(0, 0, h_{21}, 0)'\Big\}\,\mathrm{dr}\\
&+ \Big[\mathrm{e}^{(t-r)R}(0, 0, h_{21}(V(r)), 0)'\Big]_{r=0}^{r=t}.
\end{aligned} \tag{3.110}$$

This representation will be used to estimate $\|V(t)\|_2$, while the simpler one given in (3.106) will be sufficient to estimate $\|\partial_x V(t)\|_2$.

Let

$$M(t) := \sup_{0 \le r \le t} (1 + r)^{\frac{1}{4}} \|V(r)\|_{1,2}. \tag{3.111}$$

By (3.110) and Lemma 3.10 we obtain

$$\begin{aligned}
\|V(t)\|_2 \le c(1 + t)^{-\frac{1}{4}} (\|V_0\|_2 + \|V_0\|_1) + c \int_0^t &\Big\{\mathrm{e}^{-c_1(t-r)} \|(h_{11}, h_{22}, h_{21})\|_{1,2}\\
&+ (1 + t - r)^{-\frac{3}{4}} \|(h_{11}, h_{22}, h_{21})\|_1\Big\}\,\mathrm{dr} + c\|h_{21}(V)\|_2,
\end{aligned} \tag{3.112}$$

$c, c_1, \ldots$ denoting positive constants neither depending on t nor on V_0.
Since

$$\begin{aligned}
\|(h_{11}, h_{22}, h_{21})\|_{1,2}(r) &\le c\|V\|_{1,2}\|V\|_\infty(r)\\
&\le c(1 + r)^{-\frac{1}{4}} M(r)\|V_0\|_{2,2},
\end{aligned} \tag{3.113}$$

the latter by the fact that $\|V(t)\|_{2,2} \leq c\|V_0\|_{2,2}$ (cf. [112]), and since

$$\|(h_{11}, h_{22}, h_{21})\|_1 \leq c\|V\|_2^2 \leq c(1+r)^{-\frac{1}{2}}M^2(r), \tag{3.114}$$

we get

$$\begin{aligned}
\|V(t)\|_2 &\leq c(1+t)^{-\frac{1}{4}}m_0 + c\int_0^t \Big\{ e^{-c_1(t-r)}(1+r)^{-\frac{1}{4}}M(r)m_0 \\
&\quad + (1+t-r)^{-\frac{3}{4}}(1+r)^{-\frac{1}{2}}M^2(r)\Big\}\,\mathrm{dr} + c(1+t)^{-\frac{1}{4}}M(t)m_0 \\
&\leq c(1+t)^{-\frac{1}{4}}m_0\,(1+M(t)) + c(1+t)^{-\frac{1}{4}}\Big\{m_0M(t) + M^2(t)\Big\}\cdot J,
\end{aligned}$$

where

$$\begin{aligned}
J = J(t) &= \int_0^t e^{-c_1(t-r)}(1+r)^{-\frac{1}{4}}(1+t)^{\frac{1}{4}}\,\mathrm{dr} \\
&\quad + \int_0^t (1+t-r)^{-\frac{3}{4}}(1+r)^{-\frac{1}{2}}(1+t)^{\frac{1}{4}}\,\mathrm{dr}
\end{aligned}$$

satisfies

$$\sup_{t\geq 0} J(t) < \infty$$

(see Lemma 7.4 in [91]), hence

$$\|V(t)\|_2 \leq c(1+t)^{-\frac{1}{4}}\Big\{m_0(1+M(t)) + M^2(t)\Big\}. \tag{3.115}$$

To estimate $\|\partial_x V(t)\|_2$ we use (3.106) directly and the Lemmas 3.10 and 3.12 to similarly conclude

$$\begin{aligned}
\|\partial_x V(t)\|_2 &\leq c(1+t)^{-\frac{3}{4}}(\|V_0\|_{1,2} + \|V_0\|_1) + c\int_0^t \Big\{ e^{-c_1(t-r)}\|(h_1, h_2)\|_{1,2} \\
&\quad + (1+t-r)^{-\frac{3}{4}}\|(h_1, h_2)\|_1\Big\}\,\mathrm{dr} \\
&\leq c(1+t)^{-\frac{3}{4}}m_0 + c\int_0^t \Big\{ e^{-c_1(t-r)}(1+r)^{-\frac{1}{4}}M(r)m_0 \\
&\quad + (1+t-r)^{-\frac{3}{4}}(1+r)^{-\frac{1}{2}}M^2(r)\Big\}\,\mathrm{dr} \\
&\leq c(1+t)^{-\frac{1}{4}}\Big\{m_0(1+M(t)) + M^2(t)\Big\},
\end{aligned} \tag{3.116}$$

where the term "$\|V_{xx}\|_\infty$" in the estimate for $\|(h_1, h_2)\|_{1,2}$ produces $\|V_0\|_{3,2}$.

Combining (3.115), (3.116) we get

$$M(t) \leq c_2 \left\{ m_0(1 + M(t)) + M^2(t) \right\}.$$

Choosing $m_0 \leq \frac{c_2}{2}$ we obtain

$$M(t) \leq 2c_2 m_0 + c_2 M^2(t).$$

Since

$$M(0) \leq c m_0,$$

it is a standard argument (cp. e.g. [91]) to conclude that – for sufficiently small m_0 – we have $M(t) \leq c_3$ (c_3 being the smallest zero of $f(x) = c_2 x^2 - x + 2c_2 m_0$) and then

$$\|V(t)\|_{1,2} \leq c_4 (1+t)^{-\frac{1}{4}} m_0,$$

which proves Theorem 3.11. □

Finally we report on the recent results on the global well-posedness of the nonlinear initial-boundary value problem in radially symmetric bounded situations in two or three space dimensions. The system is written as

$$U_{tt} - A(\nabla U, \theta, q)\Delta U + B(\nabla U, \theta, q)\nabla\theta = 0, \tag{3.117}$$

$$c(\nabla U, \theta, q)\theta_t + g(\nabla U, \theta, q)\nabla' q + B_{ij}(\nabla U, \theta, q)\partial_t \partial_j U_i = 0, \tag{3.118}$$

$$\tau(\nabla U, \theta) q_t + q + K(\nabla U, \theta)\nabla\theta = 0, \tag{3.119}$$

for functions $(U, \theta, q) = (U, \theta, q)(t, x)$, $t \geq 0$, $x \in \Omega$, Ω being a bounded radially symmetric domain in R^3. We again consider the initial conditions

$$U(t=0) = U^0, \qquad U_t(t=0) = U^1, \qquad \theta(t=0) = \theta^0 \tag{3.120}$$

and the Dirichlet boundary conditions

$$U_{|\partial\Omega} = 0, \qquad \theta_{|\partial\Omega} = 0. \tag{3.121}$$

To get the global existence for small initial data we have assumed, as in Section 2 in (2.182), that for $(\tilde{U}, \tilde{\theta})$ with $\operatorname{rot} \tilde{U} = 0$,

$$C_{ijkl}(\nabla\tilde{U}, \tilde{\theta}, q)\frac{\partial^2 \tilde{U}_k}{\partial x_j \partial x_l} \equiv A_{ik}(\nabla\tilde{U}, \tilde{\theta})\Delta\tilde{U}_k, \quad i = 1, 2, 3, \tag{3.122}$$

and anticipated that we shall prove a result for U with vanishing rotation. For the tensors A, τ, K and the functions c, g we have similar assumptions as in the classical case, i.e., that they are uniformly positive for bounded values of their arguments, and that B is definite there. Again the system shall be initially isotropic, i.e., for $n = 1, 2$,

$$A(0,0,0) = \alpha \mathrm{Id}_{\mathbb{R}^n}, \quad c(0,0,0) = \zeta, \quad \tau(0,0) = \tau_0,$$
$$B(0,0,0) = \beta \mathrm{Id}_{\mathbb{R}^n}, \quad g(0,0,0) = \gamma, \quad K(0,0) = \kappa \mathrm{Id}_{\mathbb{R}^n}.$$

Then we can rewrite (3.117)–(3.119) as

$$U_{tt} - \alpha \Delta U + \beta \nabla \theta = F, \tag{3.123}$$
$$\zeta \theta_t + \gamma \nabla' q + \beta \nabla' U_t = G, \tag{3.124}$$
$$\tau_0 q_t + q + \kappa \nabla \theta = H, \tag{3.125}$$

where

$$F := (A(\nabla U, \theta, q) - \alpha)\Delta U - (B(\nabla U, \theta, q) - \beta)\nabla \theta, \tag{3.126}$$
$$G := -(c(\nabla U, \theta, q) - \zeta)\theta_t - (g(\nabla U, \theta, q) - \gamma)\nabla' q \\ - \operatorname{tr}\left((B(\nabla U, \theta, q) - \beta)\nabla U_t\right), \tag{3.127}$$
$$H := -(\tau(\nabla U, \theta) - \tau_0) q_t - (K(\nabla U, \theta) - \kappa)\nabla \theta. \tag{3.128}$$

Of course we assume that the coefficients are compatible with radial symmetry, cp. Section 2. Furthermore we assume

$$KT = TK \tag{3.129}$$

which is, for example, no restriction in the standard case $T = \tau_0 \mathrm{Id}_{\mathbb{R}^n}$.

THEOREM 3.13. *There is $\epsilon > 0$ such that if*

$$\|U^0\|_{4,2} + \|U^1\|_{3,2} + \|\theta^0\|_{3,2} + \|q^0\|_{3,2} \le \epsilon, \tag{3.130}$$

then there exists a unique solution (U, θ, q) of (3.117)–(3.121) satisfying

$$U \in \bigcap_{j=0}^{4} C^j([0,\infty), H^{4-j}(\Omega)), \theta, q \in \bigcap_{j=0}^{3} C^j([0,\infty), H^{4-j}(\Omega)).$$

Moreover, the system is exponentially stable, i.e., there are constants $C, d > 0$, being independent of the data and of t such that for all $t \ge 0$ we have

$$\Lambda(t) := \sum_{j=0}^{4} \|(\partial_t, \nabla)^j U(t,\cdot)\|^2 + \sum_{j=0}^{3} \|(\partial_t, \nabla)^j (\theta, q)(t,\cdot)\|^2 \le C e^{-dt} \Lambda(0).$$

The subtle proof uses the energy with nonlinear terms,

$$E_1(t) := E_1[U, \theta, q](t) := \frac{1}{2}\int_\Omega (|U_t|^2 + \partial_k U_i A_{km} \partial_i U_m + q_i g K_i^{-1} j T_{jm} q_m)\mathrm{dx},$$
$$E_2(t) := E_1[U_t, \theta_t, q_t](t),$$

and then combines techniques from [93,94]. As in the classical case, the radially symmetric situation is an application of a more general statement in an arbitrary smoothly bounded domain Ω (cf. the linear situation in (3.22)), which just requires

$$\operatorname{rot} U = \operatorname{rot} q = 0 \quad \text{in } [0, \infty) \times \Omega,$$
$$\nu \times q = 0 \quad \text{in } [0, \infty) \times \partial\Omega,$$

where ν denotes the exterior normal vector on $\partial\Omega$.

Notes: The result on exponential stability in one dimension is given in [93]. Extensions were provided by Messaoudi and Said-Houari [67]. Tarabek [112] proved a global well-posedness result for the one-dimensional Cauchy problem and the decay to an equilibrium. Details for the local existence, and decay rates were given by Racke and Wang [99]. The global well-posedness and the exponential stability in three dimensions was proved by Irmscher [36,37]. The propagation of singularities for semilinear systems is included in the paper by Racke and Wang [100]. A semilinar transmission problem has been investigated by Fernández Sare, Muñoz Rivera and Racke [22]. The fully nonlinear Cauchy problem and exterior domains have not yet been treated.

3.3. *Remarks on dual-phase lag systems*

The Cattaneo model is sometimes illustrated by starting with a delay relation between the heat flux and the temperature gradient expressing that the heat flux should notice a gradient with a certain delay, after a relaxation time τ has passed,

$$q(t + \tau, \cdot) = -\kappa\theta(t, \cdot). \tag{3.131}$$

The Fourier law then formally arises from a Taylor expansion of order zero with respect to τ, while the Cattaneo law formally is obtained by an expansion of order one, cf. [8, 116]. Moreover, there exist different mechanisms causing a dual-phase-lag expressed in the delay equation

$$q(t + \tau_q, \cdot) = -\kappa\theta(t + \tau_\theta, \cdot) \tag{3.132}$$

with two relaxation parameters $\tau_q, \tau_\theta > 0$. Conditions under which first- or second-order approximations on the right-hand and/or on the left-hand side lead to exponentially stable systems – of rather hyperbolic and also of rather parabolic type – have been investigated, see, for example, the papers of Quintanilla and Racke [82–85] and the references therein.

More generally, one might look at Taylor expansions

$$q(t,\cdot)+\cdots+\frac{\tau_q^j}{j!}\frac{\partial^j}{\partial_t^j}q(t,\cdot)=-k\nabla\theta(t,\cdot)-\cdots-k\frac{\tau_\theta^m}{m!}\frac{\partial^m}{\partial_t^m}\nabla\theta(t,\cdot) \tag{3.133}$$

and hope for "better" results because of a "better" Taylor approximation. We point out that this is not true but that we have instability if $j-m\geq 2$ (observe that for Fourier's law we have $j-m=0$ and for Cattaneo's law $j-m=1$). This instability can be proved looking at the roots of the associated characteristic polynomial and detecting roots with a sign of the real part leading to exponentially increasing solutions not allowing stability, cf. [19]. This instability for $m=0$ and $j\geq 2$ fits the fact that the delay equation (3.131) leads to an unstable system already for the simple pure heat equation with delay, cf. the book of Prüß[81].

Acknowledgements

For helpful comments during the preparation of this work, I would like to thank Jürgen Saal, Olaf Weinmann and Michael Pokojovy (University of Konstanz).

The DFG-project "Hyperbolic Thermoelasticity" (RA 504/3-1–3) made the compilation of this work as a whole possible and, in particular, supported parts of the author's research described in Section 3.

References

[1] R.A. Adams, *Sobolev Spaces*, Academic Press, New York, 1975.

[2] J. Borkenstein, *L^p–L^q-Abschätzungen der linearen Thermoelastizitätsgleichungen im* $\mathbb{R}^2$, Diploma Thesis, University of Bonn, 1993 .

[3] J.A. Burns, Z. Liu, S. Zheng, *On the energy decay of a linear thermoelastic bar*, J. Math. Anal. Appl. **179** (1993) 574–591.

[4] D.E. Carlson, *Linear thermoelasticity*, Handbuch der Physik **VIa/2** (1972) 297–346.

[5] D. Carvalho Pereira, G. Perla Menzala, *Exponential decay of solutions to a coupled system of equations of linear thermoelasticity*, Mat. Appl. Comp. **8** (1989) 193–204.

[6] D. Carvalho Pereira, G. Perla Menzala, *Exponential stability in linear thermoelasticity: The inhomogeneous case*, Appl. Anal. **44** (1992) 21–35.

[7] C. Cattaneo, *Sulla conduzione del calore*, Atti Sem. Mat. Fis. Univ. Modena **3** (1948) 83–101.

[8] D.S. Chandrasekharaiah, *Hyperbolic thermoelasticity: A review of recent literature*, Appl. Mech. Rev. **51** (1998) 705–729.

[9] G.-Q. Chen, C.M. Dafermos, *The vanishing viscosity method in one-dimensional thermoelasticity*, Trans. Amer. Math. Soc. **347** (1995) 531–541.

[10] A. Chrzęszczyk, *Some existence results in dynamical thermoelasticity. Part I. Nonlinear case*, Arch. Mech. **39** (1987) 605–617.

[11] K.N. Chueh, C.C. Conley, J.A. Smoller, *Positively invariant regions for systems of nonlinear diffusion equations*, Indiana Univ. Math. J. **26** (1977) 373–392.

[12] C.M. Dafermos, *On the existence and the asymptotic stability of solutions to the equations of linear thermoelasticity*, Arch. Rational Mech. Anal. **29** (1968) 241–271.

[13] C.M. Dafermos, *Contraction semigroups and trend to equilibrium in continuum mechanics*, Lecture Notes in Math. **503** (1976) 295–306.

[14] C.M. Dafermos, L. Hsiao, *Development of singularities in solutions of the equations of nonlinear thermoelasticity*, Quart. Appl. Math. **44** (1986) 463–474.

[15] W. Dan, *On a local in time solvability of the Neumann problem of quasilinear hyperbolic parabolic coupled systems*, Math. Methods Appl. Sci. **18** (1995) 1053–1082.

[16] G. Dassios, M. Grillakis, *Dissipation rates and partition of energy in thermoelasticity*, Arch. Rational Mech. Anal. **87** (1984) 49–91.

[17] R.J. DiPerna, *Convergence of approximate solutions to conservation laws*, Arch. Rational Mech. Anal. **82** (1983) 27–70.

[18] M.S. Doll, *Zur Dynamik (magneto-) thermoelastischer Systeme im* $\mathbb{R}^2$, Thesis, University of Konstanz, 2004 .

[19] M. Dreher, R. Quintanilla, R. Racke, *Ill-posed problems in thermomechanics*, Konstanzer Schriften Math. Inf. **248** (2008).

[20] U. Durek, *Ein Existenzsatz für nichtlineare Thermoelastizitätsgleichungen zu großen Daten*, Thesis, University of Konstanz, 1997 .

[21] E. Feireisl, *Forced vibrations in one-dimensional nonlinear thermoelasticity as a local coercive-like problem*, Comment. Math. Univ. Carolinae **31** (1990) 243–255.

[22] H.D. Fernández Sare, J.E. Muñoz Rivera, R. Racke, *Stability for a transmission problem in thermoelasticity*, J. Thermal Stresses **31** (2008) 1170–1189.

[23] H.D. Fernández Sare, R. Racke, *On the stability of damped Timoshenko systems — Cattaneo versus Fourier law*. Arch. Rational Mech. Anal. (in press).

[24] J. Gawinecki, L^p–L^q*-time decay estimate for solution of the Cauchy problem for hyperbolic partial differential equations of linear thermoelasticity*, Ann. Pol. Math. **2** (1991) 135–145.

[25] I. Hansen, *Lebensdauer von klassischen Lösungen nichtlinearer Thermoelastizitätsgleichungen,* Diploma Thesis, University of Bonn, 1994 .

[26] S.W. Hansen, *Exponential decay in a linear thermoelastic rod*, J. Math. Anal. Appl. **167** (1992) 429–442.

[27] A. Haraux, *Semi-groupes linéaires et équations d'évolution linéaires périodiques, Analyse Numérique,* Univ. Pierre et Marie Curie, Paris (1978) Preprint 78011 .

[28] D.B. Henry, *Topics in analysis*, Publ. Secc. Mat., Univ. Autòn. Barc. **31** (1987) 29–84.

[29] D.B. Henry, A Perissinitto Jr., O. Lopes, *On the essential spectrum of a semigroup of thermoelasticity*, Nonlinear Anal., TMA **21** (1993) 65–75.

[30] L. Hörmander, *The Analysis of Linear Partial Differential Operators III*, Springer-Verlag, Heidelberg, 1985.

[31] W.J. Hrusa, S.A. Messaoudi, *On formation of singularities in one-dimensional nonlinear thermoelasticity*, Arch. Rational Mech. Anal. **111** (1990) 135–151.

[32] W.J. Hrusa, M.A. Tarabek, *On smooth solutions of the Cauchy problem in one-dimensional nonlinear thermoelasticity*, Quart. Appl. Math. **47** (1989) 631–644.

[33] L. Hsiao, S. Jiang, *Nonlinear hyperbolic–parabolic coupled systems*, in: C.M. Dafermos, E. Feireisl (Eds.), in: Handbook of Differential Equations, Evolutionary Equations, vol. 1, Elsevier B.V., 2004, pp. 287–384.

[34] D. Huet, *Décomposition Spectrale et Opérateurs*, Presses Universitaires de France, Nancy, 1976.

[35] T. Irmscher, *Rate of stability in hyperbolic thermoelasticity*, Konstanzer Schriften Math. Inf. **214** (2006).

[36] T. Irmscher, *Aspekte hyperbolischer Thermoelastizität*, Thesis, University of Konstanz, 2007 .

[37] T. Irmscher, *Global existence and nonlinear hyperbolic thermoelasticity with radial symmetry*, Konstanzer Schriften Math. Inf. **238** (2007).

[38] T. Irmscher, R. Racke, *Sharp decay rates in parabolic and hyperbolic thermoelasticity*, IMA J. Appl. Math. **71** (2006) 459–478.

[39] S. Jiang, *Global existence and asymptotic behavior of smooth solutions in one-dimensional nonlinear thermoelasticity*, Bonner Math. Schriften **192** (1989) 1–60.

[40] S. Jiang, *Global existence of smooth solutions in one-dimensional nonlinear thermoelasticity*, Proc. R. Soc. Edinburgh **115A** (1990) 257–274.

[41] S. Jiang, *Global solutions of the Dirichlet problem in one-dimensional nonlinear thermoelasticity*, 138, University of Bonn, 1990, SFB 256 Preprint .

[42] S. Jiang, *Global solutions of the Neumann problem in one-dimensional nonlinear thermoelasticity*, Nonlinear Anal., TMA **19** (1992) 107–121.

[43] S. Jiang, *On global smooth solutions to the one-dimensional equations of nonlinear inhomogeneous thermoelasticity*, Nonlinear Anal., TMA **20** (1993) 1245–1256.

[44] S. Jiang, *Exponential decay and global existence of spherically symmetric solutions in thermoelasticity*, Chin. Ann. Math. **19A** (1998) 629–640 (in Chinese).
[45] S. Jiang, J.E. Muñoz Rivera, R. Racke, *Asymptotic stability and global existence in thermoelasticity with symmetry*, Quart. Appl. Math. **56** (1998) 259–275.
[46] S. Jiang, R. Racke, *On some quasilinear hyperbolic–parabolic initial boundary value problems*, Math. Methods Appl. Sci. **12** (1990) 315–339.
[47] S. Jiang, R. Racke, *Evolution Equations in Thermoelasticity*, in: π Monographs Surveys Pure Appl. Math., vol. 112, Chapman & Hall/CRC, Boca Raton, 2000.
[48] F. John, *Formation of singularities in one-dimensional nonlinear wave propagation*, Comm. Pure Appl. Math. **27** (1974) 377–405.
[49] F. John, *Formation of singularities in elastic waves*, in: P.G. Ciarlet, M. Roseau (Eds.), Trends and Applications of Pure Mathematics to Mechanics. Proceedings Palaiseau (1983), in: Lec. Notes Phys., vol. 195, 1984, pp. 194–210.
[50] S. Kawashima, *Systems of a hyperbolic–parabolic composite type, with applications to the equations of magnetohydrodynamics*, Thesis, Kyoto University, 1983 .
[51] S. Kawashima, *Large-time behaviour of solutions to hyperbolic–parabolic systems of conservation laws and applications*, Proc. R. Soc. Edinburgh **106A** (1987) 169–194.
[52] S. Kawashima, M. Okada, *Smooth global solutions for the one-dimensional equations in magnetohydrodynamics*, Proc. Japan Acad., Ser. A **53** (1982) 384–387.
[53] S. Kawashima, Y. Shibata, *On the Neumann problem of one-dimensional nonlinear thermoelasticity with time-independent forces*, Czech. Math. J. **45** (1995) 39–67.
[54] J.U. Kim, *On the energy decay of a linear thermoelastic bar and plate*, SIAM J. Math. Anal. **23** (1992) 889–899.
[55] H. Koch, *Slow decay in linear thermo-elasticity*, Quart. Appl. Math. **58** (2000) 601–612.
[56] V.D. Kupradze, *Three-Dimensional Problems of the Mathematical Theory of Elasticity and Thermoelasticity*, North-Holland, Amsterdam, 1979.
[57] O.A. Ladyzhenskaya, V.A. Solonnikov, N.N. Ural'tseva, *Linear and Quasilinear Equations of Parabolic Type*, Amer Math. Soc., Providence, Rhode Island, 1968.
[58] G. Lebeau, E. Zuazua, *Decay rates for the three-dimensional linear system of thermoelasticity*, Arch. Rational Mech. Anal. **148** (1999) 179–231.
[59] R. Leis, *Außenraumaufgaben in der linearen Elastizitätstheorie*, Math. Methods Appl. Sci. **2** (1980) 379–396.
[60] R. Leis, *Über das asymptotische Verhalten thermoelastischer Wellen im* $\mathbb{R}^3$, Math. Methods Appl. Sci. **3** (1981) 312–317.
[61] R. Leis, *Initial Boundary Value Problems in Mathematical Physics*, B.G Teubner-Verlag, John Wiley & Sons, Stuttgart, Chichester, 1986.
[62] O. Liess, *Decay estimates for the solutions of the system of crystal optics*, Asymptotic Anal. **4** (1991) 61–95.
[63] J.-L. Lions, E. Magenes, *Problèmes aux Limites Non Homogènes et Applications Vol. 1*, Dunod, Paris, 1968.
[64] Z. Liu, S. Zheng, *Exponential stability of semigroup associated with thermoelastic system*, Quart. Appl. Math. **51** (1993) 535–545.
[65] Z. Liu, S. Zheng, *Semigroups Associated with Dissipative Systems*, in: Res. Notes Math., vol. 398, Chapman & Hall/CRC, 1999.
[66] H.W. Lord, Y. Shulman, *A generalized dynamical theory of thermoelasticity*, J. Mech. Phys. Solids **15** (1967) 299–309.
[67] S.A. Messaoudi, B. Said-Houari, *Exponential stability in one-dimensional non-linear thermoelasticity with second sound*, Math. Methods Appl. Sci. **28** (2005) 205–232.
[68] A.M. Micheletti, *Perturbazione dello spettro dell'operatore di Laplace, in relazione ad una variazione del campo*, Ann. Scuola Norm. Sup. Pisa **26** (1972) 151–169.
[69] A.M. Micheletti, *Metrica per famiglie di domini limitati e proprietà generiche degli autovalori*, Ann. Scuola Norm. Sup. Pisa **26** (1972) 683–694.
[70] A.M. Micheletti, *Perturbazione dello spettro di un operatore ellittico di tipo variazionale, in relazione ad una variazione del campo*, Ann. Mat. Pura Appl., Ser. 4 **97** (1973) 267–281.
[71] T. Mukoyama, *On the Cauchy problem for quasilinear hyperbolic–parabolic coupled systems in higher dimensional spaces*, Tsukuba J. Math. **13** (1989) 363–386.

[72] J.E. Muñoz Rivera, *Energy decay rates in linear thermoelasticity*, Funkcial. Ekvac. **35** (1992) 19–30.
[73] J.E. Muñoz Rivera, *Decomposition of the displacement vector field and decay rates in linear thermoelasticity*, SIAM J. Math. Anal. **24** (1993) 1–17.
[74] J.E. Muñoz Rivera, *Asymptotic behaviour in inhomogeneous linear thermoelasticity*, Appl. Anal. **53** (1994) 55–65.
[75] J.E. Muñoz Rivera, R. Racke, *Multidimensional contact problems in thermoelasticity*, SIAM J. Appl. Math. **58** (1998) 1307–1337.
[76] F. Murat, *L'injection du cône positif de H^{-1} dans $W^{-1,q}$ est compacte pour tout $q < 2$*, J. Math. Pures Appl. **60** (1981) 309–322.
[77] Y. Naito, R. Racke, Y. Shibata, *Low frequency expansion in thermoelasticity with second sound in three dimensions*, Konstanzer Schriften Math. Inf. **239** (2008).
[78] J.A. Nohel, R.C. Rogers, A.E. Tzavaras, *Weak solutions for a nonlinear system in viscoelasticity*, Comm. Partial Differential Equations **13** (1988) 97–127.
[79] T.S. Öncü, T.B. Moodie, *On the constitutive relations for second sound in elastic solids*, Arch. Rational Mech. Anal. **121** (1992) 87–99.
[80] G. Ponce, R. Racke, *Global existence of small solutions to the initial value problem for nonlinear thermoelasticity*, J. Differential Equations **87** (1990) 70–83.
[81] J. Prüß, *Evolutionary Integral Equations and Applications*, in: Monographs Math., vol. 87, Birkhäuser, Basel, 1993.
[82] R. Quintanilla, R. Racke, *Qualitative aspects in dual-phase-lag thermoelasticity*, SIAM J. Appl. Math. **66** (2006) 977–1001.
[83] R. Quintanilla, R. Racke, *Qualitative aspects in dual-phase-lag heat conduction*, Proc. R. Soc. A **463** (2007) 659–674.
[84] R. Quintanilla, R. Racke, *A note on stability in dual-phase-lag heat conduction*, Int. J. Heat Mass Transfer **49** (2006) 1209–1213.
[85] R. Quintanilla, R. Racke, *A note on stability in three-phase-lag heat conduction*, Int. J. Heat Mass Transfer **51** (2008) 24–29.
[86] R. Racke, *Eigenfunction expansions in thermoelasticity*, J. Math. Anal. Appl. **120** (1986) 596–609.
[87] R. Racke, *On the time-asymptotic behaviour of solutions in thermoelasticity*, Proc. R. Soc. Edinburgh **107A** (1987) 289–298.
[88] R. Racke, *Initial boundary value problems in one-dimensional non-linear thermoelasticity*, Math. Methods Appl. Sci. **10** (1988) 517–529.
[89] R. Racke, *On the Cauchy problem in nonlinear 3-d-thermoelasticity*, Math. Z. **203** (1990) 649–682.
[90] R. Racke, *Blow-up in nonlinear three-dimensional thermoelasticity*, Math. Methods Appl. Sci. **12** (1990) 267–273.
[91] R. Racke, *Lectures on Nonlinear Evolution Equations. Initial Value Problems*, in: Aspects of Mathematics, vol. E19, Friedr Vieweg & Sohn, Braunschweig, Wiesbaden, 1992.
[92] R. Racke, *Exponential decay for a class of initial boundary value problems in thermoelasticity*, Comp. Appl. Math. **12** (1993) 67–80.
[93] R. Racke, *Thermoelasticity with second sound — Exponential stability in linear and nonlinear 1-d*, Math. Methods Appl. Sci. **25** (2002) 409–441.
[94] R. Racke, *Asymptotic behavior of solutions in linear 2- or 3-d thermoelasticity with second sound*, Quart. Appl. Math. **61** (2003) 315–328.
[95] R. Racke, Y. Shibata, *Global smooth solutions and asymptotic stability in one-dimensional nonlinear thermoelasticity*, Arch. Rational Mech. Anal. **116** (1991) 1–34.
[96] R. Racke, Y. Shibata, S. Zheng, *Global solvability and exponential stability in one-dimensional nonlinear thermoelasticity*, Quart. Appl. Math. **51** (1993) 751–763.
[97] R. Racke, Y.G. Wang, *Propagation of singularities in one-dimensional thermoelasticity*, J. Math. Anal. Appl. **223** (1998) 216–247.
[98] R. Racke, Y.G. Wang, *Asymptotic behavior of discontinuous solutions to thermoelastic systems with second sound*, J. Anal. Appl. **24** (2005) 117–135.
[99] R. Racke, Y.G. Wang, *Nonlinear well-posedness and rates of decay in thermoelasticity with second sound*, J. Hyperbolic Differential Equations **5** (2008) 25–43.
[100] R. Racke, Y.G. Wang, *Asymptotic behavior of discontinuous solutions in 3-d thermoelasticity with second sound*, Quart. Appl. Math. **66** (2008) 707–724.

[101] M. Reissig, Y.G. Wang, *Propagation of mild singularities in higher dimensional thermoelasticity*, J. Math. Anal. Appl. **240** (1999) 398–415.
[102] M. Reissig, J. Wirth, *Anisotropic thermoelasticity in 2d. Part I: A unified treatment*, Asymptotic Anal. **57** (2008) 1–27.
[103] M.O. Rieger, *Exponential stability and global existence in thermoelasticity with radial symmetry*, Quart. Appl. Math. **62** (2004) 1–25.
[104] W. Shen, S. Zheng, *Global smooth solutions to the system of one-dimensional thermoelasticity with dissipation boundary conditions*, Chin. Ann. Math. **7B** (1986) 303–317.
[105] Y. Shibata, *Neumann problem for one-dimensional nonlinear thermoelasticity*, Banach Center Publ. **27** (1992) 457–480.
[106] Y. Shibata, *On a decay rate of solutions to one dimensional thermoelastic equations on a half line; linear part*, in: S. Kawashima, T. Yanagisawa (Eds.), Advances in Nonlinear Partial Differential Equations and Stochastics, in: Ser. Advances Math. Appl. Sci., vol. 48, 1998, pp. 198–291.
[107] M. Slemrod, *Global existence, uniqueness, and asymptotic stability of classical smooth solutions in one-dimensional non-linear thermoelasticity*, Arch. Rational Mech. Anal. **76** (1981) 97–133.
[108] A. Soufyane, *Stabilisation dynamique et approximation numérique de problèmes de contrôle*, Thesis. University of Franche-Comté, Besançon, 1999 .
[109] J. Sprenger, *On a contact problem in thermoelasticity with second sound*, Quart. Appl. Math., (in press) .
[110] M. Stoth, *Globale klassische Lösungen der quasilinearen Elastizitätsgleichungen für kubisch elastische Medien im* $\mathbb{R}^2$, 157, Universität Bonn, 1991, SFB 256 Preprint .
[111] M. Stoth, *Decay estimates for solutions of linear elasticity for anisotropic media*, Math. Methods Appl. Sci. **19** (1996) 15–31.
[112] M.A. Tarabek, *On the existence of smooth solutions in one-dimensional nonlinear thermoelasticity with second sound*, Quart. Appl. Math. **50** (1992) 727–742.
[113] L. Tartar, *Compensated compactness and applications to partial differential equations*, in: R.J. Knops (Ed.), Nonlinear Analysis and Mechanics: Heriott–Watt Symposium, vol. IV, in: Research Notes Math., vol. 39, Pitman Press, 1979, pp. 136–212.
[114] M. Taylor, *Pseudodifferential Operators*, Princeton Univ Press, Princeton, New Jersey, 1981.
[115] M. Taylor, *Partial Differential Equations II*, Springer-Verlag, Heidelberg, 1996.
[116] D.Y. Tzou, *The generalized lagging response in small-scale and high-rate heating*, Int. J. Heat Mass Transfer **38** (1995) 3231–3240.
[117] K. Uhlenbeck, *Eigenfunctions of Laplace operators*, Bull. Amer. Math. Soc. **78** (1973) 1073–1076.
[118] W. Wang, Z. Wang, *The pointwise estimates to solutions for 1-dimensional linear thermoelastic systems with second sound*, J. Math. Anal. Appl. **326** (2007) 1061–1075.
[119] Y.G. Wang, L. Yang, L^p–L^q *decay estimates for Cauchy problems of linear thermoelastic systems with second sound in 3-d*, Proc. R. Soc. Edinburgh, Sec. A **136** (2006) 189–207.
[120] O. Weinmann, *The equations of thermoelasticity with time-dependent coefficients*, Konstanzer Schriften Math. Inf. **221** (2006).
[121] J. Wirth, *Anisotropic thermoelasticity in 2d. Part II: Applications*, Asymptotic Anal. **57** (2008) 29–40.
[122] L. Yang, Y.G. Wang, L^p–L^q *decay estimates for the Cauchy problem of linear thermoelastic systems with second sound in one space variable*, Quart. Appl. Math. **64** (2006) 1–15.
[123] S. Zheng, *Global solutions and applications to a class of quasilinear hyperbolic–parabolic coupled systems*, Sci. Sinica, Ser. A **27** (1984) 1274–1286.
[124] S. Zheng, *Global smooth solutions to the Cauchy problem of nonlinear thermoelastic equations with dissipation*, Chin. Ann. Math. **8B** (1987) 142–155.
[125] S. Zheng, Y. Chen, *Global existence for nonlinear parabolic equations*, Chin. Ann. Math. **7B** (1986) 57–73.
[126] S. Zheng, W. Shen, *Global solutions to the Cauchy problem of quasilinear hyperbolic parabolic coupled systems*, Sci. Sinica, Ser. A **30** (1987) 1133–1149.

CHAPTER 5

Unique Continuation Properties and Quantitative Estimates of Unique Continuation for Parabolic Equations

Sergio Vessella

DiMaD, Università di Firenze, via Lombroso 6/17, 50134 Firenze, Italy
E-mail: sergio.vessella@dmd.unifi.it

Contents

Abstract

We review the main results concerning unique continuation properties and their quantitative versions for solutions to second order parabolic equations. We deal with the backward uniqueness and the spacelike unique continuation properties for such equations. We give detailed proofs of such unique continuation results.

Keywords: Unique continuation properties, parabolic equations

---

HANDBOOK OF DIFFERENTIAL EQUATIONS
Evolutionary Equations, Volume 5
Edited by C.M. Dafermos and M. Pokorný

DOI: 10.1016/S1874-5717(08)00212-0

1. Introduction

In this chapter we will review the main results concerning the unique continuation properties for solutions to second order parabolic equations with real coefficients.

Roughly speaking, we shall say that a linear partial differential operator P, or the equation $Pu = 0$, enjoys a unique continuation property (UCP) in an open set Ω in $\mathbb{R}^N$ if the following type of result holds true, [88]. Let A be a subset of Ω then

$$Pu = 0 \text{ in } \Omega \text{ and } u = 0 \text{ in } A \text{ imply } u = 0 \text{ in } \Omega. \tag{1.1}$$

A quantitative version of the UCP property (1.1) is called a stability result or a quantitative estimate of unique continuation (QEUC):

If two solutions u_1 and u_2 to $Pu = 0$ in Ω are close in A
then they are close in Ω.

In the present introduction we shall outline the main steps to derive the main type of UCP and QEUC for parabolic equations.

Such UCP and QEUC can be divided in two large classes:

(i) Backward uniqueness and backward stability estimates for solutions to parabolic equations;
(ii) spacelike unique continuation properties and their quantitative versions which, in turns, can be divided in two subclasses:
(ii_1) Uniqueness and stability estimates for non-characteristic Cauchy problems, weak unique continuation properties and their quantitative versions for solutions to parabolic equations;
(ii_2) Strong unique continuation properties (SUCP) and their quantitative versions in the interior and at the boundary, for solutions to parabolic equations.

To make clear the exposition we introduce some notation. Let $a(x,t) = \left\{a^{ij}(x,t)\right\}_{i,j=1}^{n}$ be a real symmetric matrix valued function which satisfies a uniform ellipticity condition. We denote by $b(x,t)$ and $c(x,t)$ two measurable functions from $\mathbb{R}^{n+1}$ with values in $\mathbb{R}^n$ and $\mathbb{R}$, respectively. We assume that b and c are bounded. We denote by L the parabolic operator

$$Lu = \operatorname{div}(a(x,t)\nabla u) - \partial_t u + b(x,t)\cdot\nabla u + c(x,t)\,u.$$

(i) *Backward uniqueness and backward stability estimates.*

Let D be a bounded domain in $\mathbb{R}^n$ with a sufficiently smooth boundary and let T be a positive number. Let L be as above, and let $\mathcal{H}$ be a functional space in which it makes sense to look for the solution to $Lu = 0$, in $D \times (0,T)$, for instance, $\mathcal{H} := \mathcal{H}_0 = L^2\left(0,T; H^2(D)\cap H_0^1(D)\right)\cap H^1\left(0,T; L^2(D)\right)$. The backward problem for the operator L can be formulated as the problem of determining u satisfying

$$\begin{cases} Lu = 0, & \text{in } D \times (0, T), \\ u \in \mathcal{H}, \\ u(., T) = h, & \text{in } D, \end{cases} \tag{1.2}$$

where h is a given function. It is a simple matter to show that the problem (1.2) is ill-posed for lack of the continuous dependence of solution u on the datum h. Here we illustrate such ill-posedness in the simplest case of $L = \partial_x^2 - \partial_t$, $n = 1$, $D = (-\pi, \pi)$ and $\mathcal{H} = \mathcal{H}_0$. Let us consider the following sequence of functions

$$u^{(k)}(x, t) = \mathrm{e}^{-k} \mathrm{e}^{k^2(T-t)} \sin kx, \quad k \in \mathbb{N}.$$

We have, for every $k \in \mathbb{N}$,

$$\begin{cases} \left(\partial_x^2 - \partial_t\right) u^{(k)} = 0, & \text{in } (-\pi, \pi) \times (0, T), \\ u^{(k)} \in \mathcal{H}_0, \\ u^{(k)}(., T) = \mathrm{e}^{-k} \mathrm{e}^{k^2(T-t)} \sin kx, & \text{in } (-\pi, \pi), \end{cases}$$

moreover we have, for every $N \in \mathbb{N}$,

$$\left\| \partial_x^N u^{(k)}(., T) \right\|_{L^\infty(-\pi,\pi)} \to 0, \quad \text{as } k \to \infty$$

and, for any $t \in (0, T)$, we have

$$\left\| u^{(k)}(., t) \right\|_{L^1(-\pi,\pi)} \to 0, \quad \text{as } k \to \infty.$$

In other words, any small perturbation of the datum h of the problem (1.2) may produce uncontrollable effects on the solution u whenever such a perturbation and effect are evaluated in any functional spaces which are reasonable for the applications.

In the case in which D is an unbounded domain, for instance $D = \mathbb{R}^n$, a classical example of Tychonoff [36,90], shows that if the space $\mathcal{H}$ does not include a suitable growth condition at infinity with respect to the x variables then the uniqueness for problem (1.2) may fail. In both cases (D bounded or unbounded domain) it is of great interest for investigation about the uniqueness of the backward problem (1.2). More precisely it is of interest to investigate the properties of the space $\mathcal{H}$ and the minimal assumptions of the coefficients of L, especially on the leading coefficients a^{ij}, which guarantee the uniqueness for problem (1.2). Also, due to the ill-posedness character illustrated above, it is very important for the applications to ask for some additional information on the solutions to (1.2) which ensure a continuous dependence of such solutions on the datum h in some suitable functional spaces. As an example of the matter let us consider, again, the simplest case $L = \partial_x^2 - \partial_t$, $n = 1$, $D = (-\pi, \pi)$ and $\mathcal{H} = \mathcal{H}_0$. In such a case, assuming the existence of a solution u to the problem

$$\begin{cases} \left(\partial_x^2 - \partial_t\right) u = 0, & \text{in } (-\pi, \pi) \times (0, T), \\ u \in \mathcal{H}_0, \\ u(., T) = h, & \text{in } (-\pi, \pi), \end{cases} \tag{1.3}$$

it is simple to determine u. Indeed, denoting by $f = u(., 0)$ we have

$$u(x, t) = \sum_{k=1}^{\infty} f_k \mathrm{e}^{-k^2 t} \sin kx,$$

where

$$f_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin kx \mathrm{d}x, \quad k \in \mathbb{N}.$$

Since

$$u(x, T) = h(x), \quad \text{for every } x \in (-\pi, \pi),$$

we have

$$f_k = \mathrm{e}^{k^2 T} h_k, \quad k \in \mathbb{N},$$

where

$$h_k = \frac{1}{\pi} \int_{-\pi}^{\pi} h(x) \sin kx \mathrm{d}x.$$

Therefore, by using the Hölder inequality we have

$$\begin{aligned} \|u(., t)\|_{L^2(-\pi,\pi)} &= \pi \sum_{k=1}^{\infty} h_k^2 \mathrm{e}^{2k^2(T-t)} \\ &= \pi \sum_{k=1}^{\infty} \left(h_k^2 \mathrm{e}^{2k^2 T}\right)^{1-\frac{t}{T}} h_k^{2\frac{t}{T}} \\ &\leq \left(\pi \sum_{k=1}^{\infty} h_k^2 \mathrm{e}^{2k^2 T}\right)^{1-\frac{t}{T}} \left(\pi \sum_{k=1}^{\infty} h_k^2\right)^{\frac{t}{T}}, \end{aligned}$$

that is

$$\|u(., t)\|_{L^2(-\pi,\pi)} \leq \left(\|u(., 0)\|_{L^2(-\pi,\pi)}\right)^{1-\frac{t}{T}} \left(\|h\|_{L^2(-\pi,\pi)}\right)^{\frac{t}{T}}. \tag{1.4}$$

The estimate (1.4) is a stability estimate for the backward problem (1.3). A most important consequence of (1.4) is that if it is *a priori* known that the solution u to problem (1.3) satisfies the condition $\|u(.,0)\|_{L^2(-\pi,\pi)} \leq M$, where M is a given positive number, then $\|u(.,t)\|_{L^2(-\pi,\pi)}$, $t \in (0,T)$, is small whenever h is small.

Inequality (1.4) can be also obtained employing the so-called *logarithmic convexity* method. With such a method inequality (1.4) is derived as a consequence of the convexity of the function

$$F(t) = \log \|u(.,t)\|_{L^2(-\pi,\pi)}, \quad t \in [0,T].$$

The *logarithmic convexity* method has been introduced by Agmon and Nirenberg [4], see also [3], in order to prove the uniqueness for the backward problem for abstract evolution equations. Subsequently the *logarithmic convexity* method has been used by other authors, [13,40], in order to generalize the results of [4]. However, the uniqueness for the backward problem for abstract evolution equations has been proved for the first time by Lions and Malgrange [75] by using the method of Carleman estimates, [47,48]. The method of Carleman estimates has been also used by Lees and Protter in [71] to prove the uniqueness for the backward (1.2) under the following Lipschitz continuous hypothesis on the matrix valued function a

$$|a(x,t) - a(y,\tau)| \leq C(|x-y| + |t-\tau|). \tag{1.5}$$

We emphasize that, in order to have the uniqueness for the backward problem (1.2), the hypothesis (1.5) is not the minimal one. Indeed, Del Santo and Prizzi in [26–28] have proved that the uniqueness for the backward problem (1.2) is ensured whenever the matrix valued function a is non-Lipschitz continuous with respect to t. More precisely, denoting by μ the modulus of continuity of a, with respect to the t variable, in [26,27] it is proved that to have the uniqueness for problem (1.2) it is sufficient that μ satisfies the so called Osgood condition

$$\int_0^1 \frac{1}{\mu(s)} \mathrm{d}s = +\infty. \tag{1.6}$$

In [26,27,76] it is also proved that the uniqueness for the backward problem (1.2) may fail whenever a is Hölder continuous of order α, for any $\alpha \in (0,1)$, with respect to the variable t.

(ii) *spacelike unique continuation properties and their quantitative versions.*

Let D be a bounded domain in $\mathbb{R}^n$ and let Γ be a sufficiently smooth portion of the boundary ∂D. Let φ, ψ be given functions defined on $\Gamma \times (0,T)$, where T is a positive number. The *non-characteristic Cauchy problem* for $Lu = 0$ can be formulated as follows. Determine u such that

$$\begin{cases} Lu = 0, & \text{in } D \times (0,T), \\ u = \varphi, & \text{on } \Gamma \times (0,T), \\ a^{ij}\partial_j u \nu_i = \psi, & \text{on } \Gamma \times (0,T), \end{cases} \tag{1.7}$$

where $\nu = (\nu_1, \dots, \nu_n)$ is the exterior unit normal to ∂D.

We say that the Cauchy problem (1.7) enjoys the uniqueness property if u vanishes whenever φ, ψ do. It is well known, [43,69,79,83], that the non-characteristic Cauchy problem is a severely ill posed problem, as a small error on the data φ, ψ may have uncontrollable effects on the solution u of (1.7). Therefore it is very important for the applications to have a *stability estimate* for the solutions of (1.7) whenever such solutions belong to a certain class of functions. There exists a large quantity of literature, for parabolic and other types of equations, on the issue of stability, we refer to the books [52] and [69] for a first introduction on the subject.

The uniqueness for the Cauchy problem (1.7) is equivalent to, [78], the so called *weak unique continuation property* for operator L. Such a property asserts that, denoting by ω an open subset of D, if u is a solution to $Lu = 0$ in $D \times (0, T)$ such that $u = 0$ in $\omega \times (0, T)$, then $u = 0$ in $D \times (0, T)$. If Γ is smooth enough, say $C^{1,1}$, then stability estimates for the Cauchy problem (1.7) can be derived by a quantitative version of the above mentioned weak unique continuation property, see Section 4.3 of the present paper for details. Conversely, such a weak unique continuation property, is a consequence of the following *spacelike strong unique continuation property* [7] for the operator L. Such a property asserts that, if u is a solution to $Lu = 0$ in $D \times (0, T)$, $t_0 \in (0, T)$ and $u(., t_0)$ doesn't vanish identically in D then, for every $x_0 \in D$, there exist $C > 0$ and $K \geq 0$ such that for every r, $0 < r < \operatorname{dist}(x_0, \partial D)$, we have

$$\int_{B_r(x_0)} u^2(x, t_0)\,\mathrm{d}x \geq Cr^K. \tag{1.8}$$

In other words, if the the operator L enjoys such a spacelike strong unique continuation property then $u(., t_0)$ either vanishes in D or it cannot have a zero of infinite order in a point of D. We emphasize that in (1.8), C and K may depend on u. The natural quantitative versions of a spacelike strong unique continuation property are the doubling inequality for $u(., t_0)$, [33], and the two-sphere one-cylinder inequality. In a rough form the latter can be expressed as

$$\|u(., t_0)\|_{L^2(B_\rho(x_0))} \leq C\,\|u(., t_0)\|^{\theta}_{L^2(B_r(x_0))}\,\|u\|^{1-\theta}_{L^2(B_R(x_0)\times(t_0-R^2, t_0))}, \tag{1.9}$$

where $\theta = \left(C \log \frac{R}{Cr}\right)^{-1}$, $0 < r < \rho < R$, the cylinder $B_R(x_0) \times \left(t_0 - R^2, t_0\right)$ is contained in $D \times (0, T)$ and C depends neither on u nor on r. See Theorem 4.2.6 for a precise statement.

Likewise, L enjoys a strong unique continuation property at the *boundary* if a solution u to $Lu = 0$ in $D \times (0, T)$, which satisfies a homogeneous boundary condition, for instance $u = 0$ on $I \times (0, T)$, $I \subset \partial\Omega$, $t_0 \in (0, T)$ and $u(., t_0)$ doesn't vanish identically in D then, for every $x_0 \in I$, there exists $C > 0$ and $K \geq 0$ such that, for any sufficiently small r we have

$$\int_{B_r(x_0)\cap\Omega} u^2(x, t_0)\,\mathrm{d}x \geq Cr^K. \tag{1.10}$$

We emphasize that also in (1.10) as in (1.8), the quantities C and K may depend on u, but they do not depend on r.

Before proceeding, we dwell a bit on the regularity of the leading coefficients of L that guarantees the uniqueness and the stability of solutions to problem (1.7). Indeed, by some examples of Miller [77] and Pliš [80] in the elliptic case, it is clear that the minimal regularity of a with respect to the space variables has to be Lipschitz continuous. On the other hand, if a depends on x and t optimal regularity assumptions (to the author knowledge) are not known. Even the cases $n = 1$ or $n = 2$ present open issues of optimal regularity for a.

Now we attempt to give a view of the literature and of the main results concerning the spacelike unique continuation properties for parabolic equations and their quantitative versions. To this aim we collect the contributions on the subject in the following way.

(A) *Pioneering works and the case* $n = 1$;
(B) *Weak unique continuation properties and the Cauchy problem for* $n > 1$;
(C) *spacelike strong unique continuation properties.*

Concerning (A) we make no claim of bibliographical completeness. A good source of references on such points is the book [16].

(A) *Pioneering works.* Such works concern mainly the case where L is the heat operator $\partial_t - \Delta$ and those papers that are related to investigation of the regularity properties of solutions to $\partial_t u - \Delta u = 0$, [39,46,84]. In this group of contributions we have to include the classical papers and books in which the ill-posed character of the Cauchy problem for parabolic equations was investigated for the first time [43,50,83]. The case $n = 1$ is a special, but nevertheless interesting case. Of course, much more than in the general case ($n \geq 1$), operational transformations, particular tricks and even resolution formulae are available in the case $n = 1$, [14,15,21,45]. In such a case the regularity assumption, especially concerning the leading coefficient, can be relaxed [44,61,65].

(B) *Weak unique continuation property and Cauchy problem for* $n > 1$. The greater part of the papers of this group share the Carleman estimates technique for proving the uniqueness and the stability estimates. The basic idea of such a technique has been introduced in the paper [19] to prove the uniqueness of a solution of a Cauchy problem for elliptic systems in two variables with nonanalytic coefficients. Nowadays the general theory of Carleman estimates is presented in several books and papers [29,37,47–49,51,52, 54,58,59,69,86–89,95]. In particular, in [51], a general theory of Carleman estimates for anisotropic operators (of which L is an example) has been developed. Nirenberg [78] has proved, in a very general context, the uniqueness for the Cauchy problem (and the weak unique continuation property) when the entries of matrix a are constants. In such a paper Nirenberg has posed the following question. Let u be a solution to

$$Lu = 0, \quad \text{in } D \times (0, T),$$

where D is as above. Let ω be an open subset of D.

Is it true that $u(., t_0) = 0$ *in* ω *implies* $u(., t_0) = 0$ *in* D?

In other words, the question posed by Nirenberg is whether a weak form of the above mentioned spacelike strong unique continuation holds true.

John [50] has proved some stability estimate for the Cauchy problem for the operator L under the same hypotheses of [78]. Lees and Protter, [71,82], have proved uniqueness for the Cauchy problem, when the entries of the matrix a are assumed twice continuously differentiable. To the author's knowledge a stability estimate of Hölder type for the Cauchy problem for the operator L, under the same hypotheses of [71,82], was proved for the first time in [9], see also [8]. Many authors have contributed to reducing the regularity assumption on a in order to obtain the weak unique continuation for operator L, among those we recall [37,85], the above mentioned [51] and the references therein. In the context of a quantitative version of weak unique continuation properties for operator L, we mention the so-called *three cylinder inequalities*. Roughly speaking these are estimates of type

$$\|u\|_{L^2(B_\rho(x_0)\times(t_0,T-t_0))} \leq C\,\|u\|^{\gamma}_{L^2(B_r(x_0)\times(0,T))}\,\|u\|^{1-\gamma}_{L^2(B_R(x_0)\times(0,T))}, \tag{1.11}$$

where $\gamma \in (0,1)$, $0 < r < \rho < R$, $t_0 \in (0, T/2)$, the cylinder $B_R(x_0)\times(0,T)$ is contained in $D\times(0,T)$ and C depends neither on u nor on r, but it may depend on t_0. Such an estimate has been proved in [24,42,91] when $a \in C^3$. More recently, [32,92,93], the regularity assumption on a (up to Lipschitz continuity with respect to x and t) has been reduced and (1.11) has been proved with an optimal exponent γ, that is $\gamma = \left(C \log \frac{R}{Cr}\right)^{-1}$. Inequality (1.11) and its version at the boundary have been used in [25] to prove optimal stability estimates for the inverse parabolic problem (Dirichlet case) with an unknown boundary in the case of a non-time-varying boundary [53].

(C) *Spacelike strong unique continuation property.* Such a property and the inequality (1.9) have been proved for the first time in 1974 by Landis and Oleinik [68] in the case where all the coefficients of operator L do not depend on t. In [68] the authors discovered a method and named it the "elliptic continuation technique", to derive unique continuation properties and their quantitative versions, by properties and inequalities which hold true in the elliptic context [3,10,11,22,38,56,66,67,70]. They have employed this method also for parabolic equations of order higher than two and for systems. The elliptic continuation technique depends strongly on the time independence of the coefficients of the equation. Some fundamental ideas of this technique can be traced back to the pioneering work of Ito and Yamabe [55]. Roughly speaking, the above method consists of the following idea. Let u be a solution to the parabolic equation

$$\operatorname{div}(a(x)\nabla u) - \partial_t u = 0 \tag{1.12}$$

and let t_0 be fixed, then $u(.,t_0)$ can be continued to a solution $U(x,y)$, $y \in (-\delta,\delta)$, $\delta > 0$, of an elliptic equation in the variables x and y. In this way some unique continuation properties (and their quantitative versions) of solutions to elliptic equations can be transferred to solutions to equation (1.12). In [68] the regularity assumptions on the matrix a are very strong, but in 1990 Lin [73] employed the same technique to prove the spacelike strong unique continuation properties for solution to (1.12), assuming a merely Lipschitz continuous. In [17,18] the elliptic continuation technique has been used to prove some two-sphere one-cylinder inequalities at the boundary (with Dirichlet or Neumann conditions) for solution to (1.12).

A very important step towards the proof of the spacelike strong unique continuation property for parabolic equations with time dependent coefficients, consists in proving the following strong unique continuation property: let u be a solution of $Lu = 0$ in $D \times (0, T)$, and $(x_0, t_0) \in D \times (0, T)$ then $u(., t_0)$ vanishes whenever

$$u(x, t) = O\left(\left(|x - x_0|^2 + |t - t_0|\right)^{N/2}\right), \quad \text{for every } N \in \mathbb{N}.$$

Such a property has been proved for the first time by Poon [81] when $Lu = \partial_t u - \Delta u + b(x, t) \cdot \nabla u + c(x, t) u$, $D = \mathbb{R}^n$ and u satisfies some growth condition at infinity. Hence in [81] the strong unique continuation property has not yet been proved as a local property. Such a local strong unique continuation property has been proved, by using Carleman estimate techniques, in [30] for the operator $Lu = \partial_t u - \operatorname{div}(a(x, t) \nabla u) + b(x, t) \cdot \nabla u + c(x, t) u$, when a is Lipschitz continuous with respect to the parabolic distance and b, c are bounded. In [7] the spacelike strong unique continuation has been derived as a consequence of the theorem of [30], already mentioned, and of the local behaviour of solutions to parabolic equations proved in [6]. In [35] the same spacelike strong unique continuation property has been proved as a consequence of a refined version of the Carleman estimate proved in [30]. The natural quantitative versions, that is the two-sphere one cylinder inequality and doubling inequality on characteristic hyperplane, have been proved in [33,94]. In the present paper, Sections 4.1 and 4.2 are devoted to a complete proof of the two-sphere one-cylinder inequality (at the interior and at the boundary) and a simplified version of the proof of the Carleman estimate proved in [30] and [35]. Quite recently the spacelike strong unique continuation property has been proved in [60], with some improvement on the regularity assumption of the coefficients of operator L. On this topics see also [20].

In [30,35] the strong unique continuation properties at *the boundary* and the spacelike strong unique continuation properties at *the boundary* for second order parabolic equations have been proved. Some quantitative versions of the latter property have been proved in [33]. The investigation about such unique continuation properties at the boundary presents, with respect to the investigation for the analogous interior properties, additional problems arising from the boundary conditions. A crucial step toward the resolution of such problems is based on the results obtained in the investigation of the strong unique continuation properties at the boundary for second order elliptic equations, [1,2,5,56,57,63].

The plan of this chapter is the following. In Section 2 we give the main notation and definitions. In Section 3 we give some results concerning the uniqueness of the backward problem for second order parabolic equations. Such a Section 3 is also conceived as an introduction to the Carleman estimates methods, in particular for the derivation of a unique continuation property using a Carleman estimate. In Section 4 we prove some spacelike unique continuation properties and their quantitative versions. Such a Section 4 is subdivided into three Subsections. In Section 4.1 we consider the parabolic equations with time independent coefficients. In Section 4.2 we prove the two-sphere one-cylinder inequalities in the interior and at the boundary. In Section 4.3 we apply the above mentioned two-sphere one-cylinder inequalities to get sharp stability estimates for Cauchy problem for second order parabolic equations.

2. Main notations and definitions

For every $x \in \mathbb{R}^n$, $n \geq 2$, $x = (x_1, \ldots, x_n)$, we shall set $x = \left(x', x_n\right)$, where $x' = (x_1, \ldots, x_{n-1}) \in \mathbb{R}^{n-1}$. We shall use $X = (x, t)$ to denote a point in $\mathbb{R}^{n+1}$, where $x \in \mathbb{R}^n$ and $t \in \mathbb{R}$. For every $x \in \mathbb{R}^n$, $X = (x, t) \in \mathbb{R}^{n+1}$ we shall set $|x| = \left(\sum_{i=1}^n x_i^2\right)^{1/2}$ and $|X| = \left(|x|^2 + |t|\right)^{1/2}$. Let r be a positive number. For every $x_0 \in \mathbb{R}^n$ we shall denote by $B_r(x_0) = \{x \in \mathbb{R}^n : |x - x_0| < r\}$ (the n-dimensional open sphere of radius r centered at x_0) and $B'_r\left(x'_0\right) = \left\{x \in \mathbb{R}^{n-1} : \left|x' - x'_0\right| < r\right\}$ (the $(n-1)$-dimensional open sphere of radius r centered at x'_0). We set, generally, $B_r = B_r(0)$ and $B'_r = B'_r(0)$. We shall denote by $B_r^+ = \{x \in B_r : x_n > 0\}$.

Given a sufficiently smooth function u of x and t, we shall denote by $\partial_i u = \frac{\partial u}{\partial x_i}$, $\partial^2_{ij} u = \frac{\partial^2 u}{\partial x_i \partial x_j}$, $i, j = 1, \ldots, n$ and by $\partial_t u = \frac{\partial u}{\partial t}$.

Let $X_0 = (x_0, t_0)$, $Y_0 = (y_0, \tau_0)$ be two points of $\mathbb{R}^{n+1}$, we shall say that $S : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ is a *a rigid transformation of space coordinates under which we have* $X_0 \equiv Y_0$ if $S(X) = (\sigma(x), t - t_0 + \tau_0)$, where σ is an isometry of $\mathbb{R}^n$ such that $\sigma(x_0) = y_0$.

DEFINITION 2.0.1. Let Ω be a domain in $\mathbb{R}^{n+1}$. Given a positive integer number k and $\alpha \in (0, 1]$, we shall say that a portion Γ of $\partial\Omega$ is of class $C^{k,\alpha}$ with constants ρ_0, $E > 0$ if, for any $X_0 \in \Gamma$, there exists a rigid transformation of space coordinates under which we have $X_0 \equiv 0$ and

$$\Omega \cap \left(B'_{\rho_0}(0) \times \left(-\rho_0^2, \rho_0^2\right)\right) = \left\{X \in B'_{\rho_0}(0) \times \left(-\rho_0^2, \rho_0^2\right) : x_n > \varphi\left(x', t\right)\right\},$$

where $\varphi \in C^{k,\alpha}\left(B'_{\rho_0}(0) \times \left(-\rho_0^2, \rho_0^2\right)\right)$ satisfying

$$\varphi(0,0) = |\nabla_{x'}\varphi(0,0)| = 0$$

and

$$\|\varphi\|_{C^{k,\alpha}\left(B'_{\rho_0}(0)\times\left(-\rho_0^2,\rho_0^2\right)\right)} \leq E\rho_0.$$

We shall occasionally use the sum index convention. In addition, in order to simplify the writing and calculations we shall use some of the standard notations in Riemannian geometry, but always dropping the corresponding volume element in the definition of the Laplace–Beltrami operator associated to a Riemannian metric. More precisely, let $\left\{g^{ij}(x,t)\right\}_{i,j=1}^n$ be a symmetric matrix valued function which satisfies a uniform ellipticity condition, letting $g(x,t) = \left\{g_{ij}(x,t)\right\}_{i,j=1}^n$ denote the inverse of the matrix $\left\{g^{ij}(x,t)\right\}_{i,j=1}^n$ we have $g^{-1}(x,t) = \left\{g^{ij}(x,t)\right\}_{i,j=1}^n$ and we use the following notation when considering either a function v or two vector fields ξ and η:

1. $\xi \cdot \eta = \sum_{i,j=1}^{n} g_{ij}(x,t)\,\xi_i \eta_j$, $|\xi|_g^2 = \sum_{i,j=1}^{n} g_{ij}(x,t)\,\xi_i \xi_j$,
2. $\nabla v = (\partial_1 v, \ldots \partial_n v)$, $\nabla_g v(x,t) = g^{-1}(x,t)\,\nabla v(x,t)$, $\operatorname{div}(\xi) = \sum_{i=1}^{n} \partial_i \xi_i$, $\Delta_g = \operatorname{div}\left(\nabla_g\right)$.

With this notation the following formulae hold true when u, v and w are smooth functions

$$\partial_i\left(g^{ij}(x,t)\,\partial_j u\right) + \partial_t u = \Delta_g u + \partial_t u,\ \Delta_g\left(v^2\right) = 2v\Delta_g v + 2\left|\nabla_g v\right|_g^2 \quad (2.1)$$

and

$$\int_{\Omega\times(0,1)} v\Delta_g w \mathrm{d}X = -\int_{\Omega\times(0,1)} \nabla_g v \cdot \nabla_g w \mathrm{d}X + \int_{\partial\Omega\times(0,1)} \left(\nabla_g w \cdot \nu\right) v \mathrm{d}S, \quad (2.2)$$

where Ω is a domain of $\mathbb{R}^n$, whose boundary $\partial\Omega$ is of Lipschitz class, $\mathrm{d}X = \mathrm{d}x dt$ and $\mathrm{d}S$ is an $(n+1)$-dimensional surface element.

For any matrix M in $G((0,T))$, we shall denote its transpose by M^* and its trace by $\operatorname{tr}(M)$.

We shall fix the space dimension $n \geq 1$ throughout the paper. Therefore we shall omit the dependence of various quantities on n.

We shall use the the letters C, C_0, $C_1 \ldots$ to denote constants. The value of the constants may change from line to line, but we shall specifiy their dependence everywhere they appear.

3. Uniqueness and stability estimate for the backward problem

In this section we give two results of uniqueness for the backward problem: in the first one (Theorem 3.0.2) we consider the heat operator in the strip $\mathbb{R}^n \times (0,1)$; in the second one (Theorem 3.0.4) we consider the second order parabolic operator with variable leading coefficients in a bounded space-time cylinder. In Lemma 3.0.1 and Lemma 3.0.3, concerning the heat operator and the parabolic operator with variable leading coefficients, respectively, we prove two Carleman estimates. In Lemma 3.0.1 the Carleman estimate has been proved, in a slightly different way in [31]. The Carleman estimate of Lemma 3.0.3 has been proved, in a slight different form, in [71]. At the end of the present section we give a sketch of the proof of uniqueness and of a stability estimate, (3.65), for the backward problem by using the *logarithmic convexity* method which we have referred to in the introduction to this chapter.

In all the sections, in order to simplify the calculations, we always consider the backward parabolic operator. For instance, in Lemma 3.0.1 and in Theorem 3.0.2 we consider the operator

$$P_0 u = \Delta u + \partial_t u. \tag{3.1}$$

In Lemma 3.0.1 and Theorem 3.0.2, below, we shall use the following notation. Let a be a nonnegative number, we let

$$\mathcal{K}_a = \left\{u \in C^\infty\left(\mathbb{R}^n \times [a, +\infty)\right) \mid u\,(x, a) = 0, \operatorname{supp} u \text{ compact}\right\}.$$

LEMMA 3.0.1. *Let P_0 be the operator (3.1). There exists a constant C_0, $C_0 \geq 1$, such that for every $\mu > 0$, every $a \in \left(0, \frac{1}{4\mu}\right)$, every α, $\alpha \geq 1$ and for every $u \in \mathcal{K}_0$ such that supp $u \subset \mathbb{R}^n \times \left[0, \frac{1}{2\mu} - a\right)$, the following inequality holds true*

$$\begin{aligned}
&\alpha\mu \int (t+a)\left((t+a)\,\mathrm{e}^{-\mu(t+a)}\right)^{-2\alpha} u^2 \mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X \\
&\quad + \mu \int (t+a)^2\left((t+a)\,\mathrm{e}^{-\mu(t+a)}\right)^{-2\alpha} |\nabla u|^2\, \mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X \\
&\quad\quad \leq C_0 \int (t+a)^2\left((t+a)\,\mathrm{e}^{-\mu(t+a)}\right)^{-2\alpha} (P_0 u)^2\, \mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X.
\end{aligned} \tag{3.2}$$

PROOF. For the sake of brevity in what follows we prove inequality (3.2) when the function $u\,(x, t)$ is replaced by $\widetilde{u}\,(x, t) := u\,(x, t-a)$, so that $\widetilde{u}$ is a function belonging to $\mathcal{K}_a$. Also, the sign "$\sim$" over u will be dropped. Finally, we shall denote by $\int (.)\,\mathrm{d}X$ the integral $\int_{\mathbb{R}^n \times [a,+\infty)} (.)\,\mathrm{d}X$.

Let $\sigma \in C^2\,([0, +\infty))$ be a positive function that we will choose later on. Denote

$$\begin{aligned}
&\phi\,(x, t) = -\frac{|x|^2}{8t} - (\alpha+1)\log\sigma\,(t)\,, \\
&v = \mathrm{e}^{\phi} u, \\
&Lv = \mathrm{e}^{\phi}\,(\Delta + \partial_t)\left(\mathrm{e}^{-\phi} v\right).
\end{aligned}$$

Denoting by S and A the symmetric and skew-symmetric parts of the operator tL, respectively, we have

$$\begin{aligned}
&Sv = t\left(\Delta v + \left(|\nabla\phi|^2 - \partial_t\phi\right) v\right) - \frac{1}{2}v, \\
&Av = -t\,(2\nabla v \cdot \nabla\phi + v\Delta\phi) + \frac{1}{2}\,(\partial_t\,(tv) + t\partial_t v)
\end{aligned}$$

and

$$tLv = Sv + Av.$$

We have

$$\int t^2\,(P_0 v)^2\,\mathrm{d}X = \int (Sv)^2\,\mathrm{d}X + \int (Av)^2\,\mathrm{d}X + 2\int SvAv\mathrm{d}X.$$

Denote by

$$I := 2\int SvAv\mathrm{d}X.$$

We have

$$\begin{aligned} I = & -2\int t^2\left(\Delta v + \left(|\nabla\phi|^2 - \partial_t\phi\right)v\right)(2\nabla v\cdot\nabla\phi + v\Delta\phi)\,\mathrm{d}X \\ & + \int t\left(\Delta v + \left(|\nabla\phi|^2 - \partial_t\phi\right)v\right)(\partial_t(tv) + t\partial_t v)\,\mathrm{d}X \\ & + \int\left\{-\frac{1}{2}v(\partial_t(tv) + t\partial_t v) + tv(\Delta\phi v + 2\nabla v\cdot\nabla\phi)\right\}\mathrm{d}X \\ := & I_1 + I_2 + I_3. \end{aligned} \tag{3.3}$$

Evaluation of I_1.

We have

$$\begin{aligned} I_1 = & -4\int t^2\Delta v\,(\nabla v\cdot\nabla\phi)\,\mathrm{d}X - 4\int t^2\left(|\nabla\phi|^2 - \partial_t\phi\right)v\nabla v\cdot\nabla\phi\mathrm{d}X \\ & -2\int t^2\Delta\phi\left(v\Delta v + \left(|\nabla\phi|^2 - \partial_t\phi\right)v^2\right)\mathrm{d}X := I_{11} + I_{12} + I_{13}. \end{aligned} \tag{3.4}$$

By using the Rellich–Nečas–Pohozaev identity

$$\begin{aligned} 2(\beta\cdot\nabla v)\,\Delta v = & \,2\mathrm{div}\,((\beta\cdot\nabla v)\nabla v) - \mathrm{div}\left(\beta\,|\nabla v|^2\right) \\ & + \mathrm{div}\,(\beta)\,|\nabla v|^2 - 2\partial_i\beta^k\partial_i v\partial_k v, \end{aligned}$$

with the vector field $\beta = \nabla\phi$, we have

$$I_{11} = \int\left(4t^2\partial^2_{ik}\phi\partial_i v\partial_k v - 2t^2\Delta\phi\,|\nabla v|^2\right)\mathrm{d}X. \tag{3.5}$$

Concerning the term I_{12}, using integration by parts we have

$$\begin{aligned} I_{12} & = -2\int t^2\left(|\nabla\phi|^2 - \partial_t\phi\right)\nabla\phi\cdot\nabla\left(v^2\right)\mathrm{d}X \\ & = 2\int t^2\mathrm{div}\left(\left(|\nabla\phi|^2 - \partial_t\phi\right)\nabla\phi\right)v^2\mathrm{d}X \\ & = 2\int t^2v^2\left\{\Delta\phi\left(|\nabla\phi|^2 - \partial_t\phi\right) + \nabla\phi\cdot\nabla\left(|\nabla\phi|^2 - \partial_t\phi\right)\right\}\mathrm{d}X. \end{aligned}$$

Now we have

$$\nabla\phi\cdot\nabla\left(|\nabla\phi|^2 - \partial_t\phi\right) = 2\partial^2_{jk}\phi\partial_k\phi\partial_j\phi - \frac{1}{2}\partial_t\left(|\nabla\phi|^2\right),$$

hence

$$I_{12} = 2\int t^2 v^2 \Delta\phi \left(|\nabla\phi|^2 - \partial_t\phi\right) \mathrm{d}X$$
$$+ 4\int t^2 v^2 \partial^2_{ik}\phi \partial_i\phi \partial_k\phi \mathrm{d}X - \int t^2 v^2 \partial_t \left(|\nabla\phi|^2\right) \mathrm{d}X. \tag{3.6}$$

Now let us consider I_{13}. First notice that $\Delta\phi = -\frac{n}{4t}$. Then, by using the identity $v\Delta v = \mathrm{div}\,(v\nabla v) - |\nabla v|^2$, and by integration by parts, we obtain

$$I_{13} = \frac{n}{2}\int t \left(\mathrm{div}\,(v\nabla v) - |\nabla v|^2 + \left(|\nabla\phi|^2 - \partial_t\phi\right) v^2\right) \mathrm{d}X$$
$$= \frac{n}{2}\int \left(-t\,|\nabla v|^2 + \left(|\nabla\phi|^2 - \partial_t\phi\right) v^2\right) \mathrm{d}X,$$

so

$$I_{13} = \frac{n}{2}\int \left(-t\,|\nabla v|^2 + \left(|\nabla\phi|^2 - \partial_t\phi\right) v^2\right) \mathrm{d}X. \tag{3.7}$$

By (3.4)–(3.7) and taking into account that $\Delta\phi = -\frac{n}{4t}$, we have

$$I_1 = 4\int t^2 \partial^2_{ik}\phi \partial_i\phi \partial_k\phi v^2 \mathrm{d}X + 4\int t^2 \partial^2_{ik}\phi \partial_i v \partial_k v \mathrm{d}X$$
$$- \int t^2 v^2 \partial_t \left(|\nabla\phi|^2\right) \mathrm{d}X. \tag{3.8}$$

Evaluation of I_2.

We have

$$I_2 = \int t v \Delta v \mathrm{d}X + \int 2t^2 \left(|\nabla\phi|^2 - \partial_t\phi\right) v \partial_t v \mathrm{d}X$$
$$+ \int 2t^2 \Delta v \partial_t v \mathrm{d}X + \int t \left(|\nabla\phi|^2 - \partial_t\phi\right) v^2 \mathrm{d}X. \tag{3.9}$$

Now, using the identities $v\Delta v = \mathrm{div}\,(v\nabla v) - |\nabla v|^2$, $2v\partial_t v = \partial_t \left(v^2\right)$, $2\partial_t v \Delta v = 2\,\mathrm{div}\,(\partial_t v \nabla v) - \partial_t \left(|\nabla v|^2\right)$ in the first, second and third integrals, respectively, on the right-hand side of (3.9) we integrate by parts and obtain

$$I_2 = \int t\,|\nabla v|^2 \mathrm{d}X - \int t^2 \partial_t \left(|\nabla\phi|^2 - \partial_t\phi\right) v^2 \mathrm{d}X$$
$$- \int t \left(|\nabla\phi|^2 - \partial_t\phi\right) v^2 \mathrm{d}X. \tag{3.10}$$

Evaluation of I_3.

We have

$$I_3 = \int \left(-\frac{1}{2} + t\Delta\phi\right) v^2 \mathrm{d}X - \frac{1}{2}\int t\partial_t \left(v^2\right) \mathrm{d}X + \int 2tv\nabla v \cdot \nabla\phi \mathrm{d}X. \quad (3.11)$$

We use the identity $2v\nabla v \cdot \nabla\phi = \operatorname{div}\left(v^2\nabla\phi\right) - v^2\Delta\phi$ in the third integral at the right-hand side of (3.11) then we integrate by parts and we obtain

$$I_3 = 0. \quad (3.12)$$

Now, denote by

$$\begin{aligned} J_1 &= \int 4t^2\partial^2_{ik}\phi\partial_i\phi\partial_k\phi v^2 \mathrm{d}X \\ &\quad - \int \left\{t^2\partial_t\left(2\left|\nabla\phi\right|^2 - \partial_t\phi\right) v^2 + t\left(\left|\nabla\phi\right|^2 - \partial_t\phi\right) v^2\right\} \mathrm{d}X, \quad (3.13) \\ J_2 &= \int \left\{4t^2\partial^2_{ik}\phi\partial_i v\partial_k v + t\left|\nabla v\right|^2\right\} \mathrm{d}X. \end{aligned}$$

By (3.3), (3.8), (3.10) and (3.12) we have

$$I = J_1 + J_2. \quad (3.14)$$

Let us now examine the terms J_1, J_2, at the right-hand side of (3.14). We have easily

$$\partial_i\phi = -\frac{x_i}{4t}, \quad \partial^2_{ij}\phi = -\frac{\delta_{ij}}{4t}, \quad \left|\nabla\phi\right|^2 = \frac{|x|^2}{16t^2}, \quad (3.15)$$

$$\Delta\phi = -\frac{n}{4t}, \quad \partial_t\phi = \frac{|x|^2}{8t^2} - (\alpha+1)\frac{\sigma'(t)}{\sigma(t)}, \quad (3.16)$$

$$\partial^2_t\phi = -\frac{|x|^2}{4t^3} - (\alpha+1)\left(\frac{\sigma'(t)}{\sigma(t)}\right)', \quad \partial_t\left(\left|\nabla\phi\right|^2\right) = -\frac{|x|^2}{8t^3}. \quad (3.17)$$

Now we examine the term under the integral sign at the right-hand side of (3.13). We have

$$\begin{aligned} &4t^2\partial^2_{ik}\phi\partial_i\phi\partial_k\phi v^2 - t^2\partial_t\left(2\left|\nabla\phi\right|^2 - \partial_t\phi\right) v^2 - t\left(\left|\nabla\phi\right|^2 - \partial_t\phi\right) v^2 \\ &\qquad = (\alpha+1) v^2 t^2 \frac{\sigma'}{\sigma}\frac{\mathrm{d}}{\mathrm{d}t}\left(\log\frac{\sigma}{t\sigma'}\right). \end{aligned}$$

Therefore

$$J_1 = (\alpha+1)\int t^2 v^2 \frac{\sigma'}{\sigma}\frac{\mathrm{d}}{\mathrm{d}t}\left(\log\frac{\sigma}{t\sigma'}\right) \mathrm{d}X. \quad (3.18)$$

In addition, since

$$4t^2\partial^2_{ik}\phi\partial_i v\partial_k v + t\left|\nabla v\right|^2 = 0,$$

we have

$$J_2 = 0. \tag{3.19}$$

By (3.14), (3.18) and (3.19) we get

$$I = (\alpha + 1) \int t^2 v^2 \frac{\sigma'}{\sigma} \frac{\mathrm{d}}{\mathrm{d}t} \left(\log \frac{\sigma}{t\sigma'} \right) \mathrm{d}X. \tag{3.20}$$

Let μ be a positive number and let

$$\sigma(t) = t\mathrm{e}^{-\mu t}.$$

By (3.20) we have

$$I = (\alpha + 1)\, \mu \int t v^2 \mathrm{d}X, \tag{3.21}$$

for every $v \in \mathcal{K}_a$ such that $\operatorname{supp} v \subset \mathbb{R}^n \times \left[a, \frac{1}{2\mu}\right)$.

Now, let us consider the symmetric part S of tP_0. We have

$$\begin{aligned}
\int (Sv)^2 \,\mathrm{d}X &= \int ((Sv + \mu t v) - \mu t v)^2 \,\mathrm{d}X \\
&\geq -2\mu \int (Sv + \mu t v)\, t v \mathrm{d}X \\
&= -2\mu \int t^2 v \Delta v \mathrm{d}X - 2\mu \int \left(t \left(|\nabla \phi|^2 - \partial_t \phi \right) - \frac{1}{2} + \mu t \right) t v^2 \mathrm{d}X \\
&= 2\mu \int t^2 \left(|\nabla v|^2 + |\nabla \phi|^2 v^2 \right) \mathrm{d}X \\
&\quad - 2\mu \int \left(t \left(2 |\nabla \phi|^2 - \partial_t \phi \right) - \frac{1}{2} + \mu t \right) t v^2 \mathrm{d}X \\
&= 2\mu \int t^2 \left(|\nabla v|^2 + |\nabla \phi|^2 v^2 \right) \mathrm{d}X \\
&\quad - 2\mu \int \left((\alpha + 1)(1 - \mu t) - \frac{1}{2} + \mu t \right) t v^2 \mathrm{d}X.
\end{aligned}$$

Hence by the inequality just obtained and by (3.15) and (3.16) and (3.21) we have

$$\begin{aligned}
\int t^2 (P_0 v)^2 \,\mathrm{d}X &\geq \frac{1}{4} \int (Sv)^2 \,\mathrm{d}X + 2 \int S v A v \mathrm{d}X \\
&\geq \frac{\mu}{2} \int t^2 \left(|\nabla v|^2 + |\nabla \phi|^2 v^2 \right) \mathrm{d}X + \frac{\mu}{2} (\alpha + 1) \int t v^2 \mathrm{d}X,
\end{aligned}$$

for every $v \in \mathcal{K}_a$ such that $\operatorname{supp} v \subset \mathbb{R}^n \times \left[a, \frac{1}{2\mu}\right)$. Finally coming back to the function u we obtain (3.2). □

In the next theorem we shall use the following notation. Denote by H the set of functions u defined on $\mathbb{R}^n \times (0,1)$ such that for every $R > 0$ its restrictions $u_{|B_R\times(0,1)}$ belong to $H^{2,1}(B_R \times (0,1))$. Also, for a given positive number $M > 0$ define

$$\mathcal{H}_M = \left\{ u \in H \mid u(x,0) = 0, u\mathrm{e}^{-M|x|^2} \in L^\infty\left(\mathbb{R}^n \times (0,1)\right)\right\}.$$

THEOREM 3.0.2. *Let P_0 be the operator (3.1). Let u be a function belonging to $\mathcal{H}_M$ which satisfies the inequality*

$$|P_0 u| \le \Lambda\left(|\nabla u| + |u|\right), \quad \textit{in } \mathbb{R}^n \times (0,1), \tag{3.22}$$

where Λ is a positive number. Then

$$u \equiv 0, \quad \textit{in } \mathbb{R}^n \times (0,1).$$

PROOF. Let $\mu \ge 1$, be a number that we shall choose later on and let $a = \frac{1}{2}\min\left\{\frac{1}{4\mu}, \frac{1}{4M}\right\}$. Denote by $\delta = \frac{1}{2\mu} - a$. Let $\eta \in C^1([0,1])$ be such that $0 \le \eta \le 1$, $\eta(t) = 1$, for every $t \in \left[0, \frac{\delta}{4}\right]$, $\eta(t) = 0$, for every $t \in \left[\frac{\delta}{2}, 1\right]$ and, for a constant c_0,

$$\left|\eta'(t)\right| \le \frac{c_0}{\delta}, \quad \text{for every } t \in \left[\frac{\delta}{4}, \frac{\delta}{2}\right].$$

Let $u \in \mathcal{H}_M$. By Lemma 3.0.1 and by the standard density theorem we have that, for every $\alpha \ge 1$, the following estimate holds true

$$\begin{aligned}
&\alpha\mu \int_{\mathbb{R}^n\times(0,1)} (t+a)\left((t+a)\mathrm{e}^{-\mu(t+a)}\right)^{-2\alpha} (u\eta)^2 \mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\mathrm{d}X \\
&\quad + \mu \int_{\mathbb{R}^n\times(0,1)} (t+a)^2\left((t+a)\mathrm{e}^{-\mu(t+a)}\right)^{-2\alpha} |\nabla(u\eta)|^2 \mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\mathrm{d}X \\
&\qquad \le C_0 \int_{\mathbb{R}^n\times(0,1)} (t+a)^2\left((t+a)\mathrm{e}^{-\mu(t+a)}\right)^{-2\alpha} (P_0(u\eta))^2 \mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\mathrm{d}X,
\end{aligned} \tag{3.23}$$

where C_0 is the constant that appears in (3.2).

By (3.22) we have

$$|P_0(\eta u)| \le \Lambda\left(|\nabla u| + |u|\right)\eta + \Lambda|u|\left|\eta'\right|, \quad \text{in } \mathbb{R}^n \times (0,1). \tag{3.24}$$

By (3.23) and (3.24) we have

$$\begin{aligned}
&\alpha\mu \int_{\mathbb{R}^n\times(0,1)} (t+a)\left((t+a)\mathrm{e}^{-\mu(t+a)}\right)^{-2\alpha} (u\eta)^2 \mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\mathrm{d}X \\
&\quad + \mu \int_{\mathbb{R}^n\times(0,1)} (t+a)^2\left((t+a)\mathrm{e}^{-\mu(t+a)}\right)^{-2\alpha} |\nabla(u\eta)|^2 \mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\mathrm{d}X
\end{aligned}$$

$$\leq 2C_0\Lambda^2 \int_{\mathbb{R}^n\times(0,1)} (t+a)^2 \left((t+a)\,\mathrm{e}^{-\mu t}\right)^{-2\alpha} (u\eta)^2\,\mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X$$
$$+2C_0\Lambda^2 \int_{\mathbb{R}^n\times(0,1)} (t+a)^2 \left((t+a)\,\mathrm{e}^{-\mu(t+a)}\right)^{-2\alpha} |\nabla(u\eta)|^2\,\mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X$$
$$+2C_0\Lambda^2 \int_{\mathbb{R}^n\times(0,1)} (t+a)^2 \left((t+a)\,\mathrm{e}^{-\mu(t+a)}\right)^{-2\alpha} u^2\eta'^2\mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X, \quad (3.25)$$

for every $\alpha \geq 1$.

If $\alpha \geq \left(4\Lambda^2+1\right)C_0$ and $\mu \geq \left(2\Lambda^2+1\right)C_0$, the first and second terms on the left-hand side dominate the first and the second terms on the right-hand side of (3.25), respectively. Let us fix $\mu = \left(2\Lambda^2+1\right)C_0$, and we get

$$\alpha \int_{\mathbb{R}^n\times(0,1)} (t+a) \left((t+a)\,\mathrm{e}^{-\mu(t+a)}\right)^{-2\alpha} (u\eta)^2\,\mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X$$
$$\leq \frac{C_1}{\delta^2} \int_{\mathbb{R}^n\times\left(\frac{\delta}{4},\frac{\delta}{2}\right)} (t+a)^2 \left((t+a)\,\mathrm{e}^{-\mu(t+a)}\right)^{-2\alpha} u^2\mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X, \quad (3.26)$$

for every $\alpha \geq \left(4\Lambda^2+1\right)C_0$, where C_1 depends on Λ only.

Now, let us consider the following trivial inequalities

$$\int_{\mathbb{R}^n\times\left(\frac{\delta}{4},\frac{\delta}{2}\right)} (t+a)^2 \left((t+a)\,\mathrm{e}^{-\mu(t+a)}\right)^{-2\alpha} u^2\mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X$$
$$\leq \left(\left(\frac{\delta}{4}+a\right)\mathrm{e}^{-\mu\left(\frac{\delta}{4}+a\right)}\right)^{-2\alpha} \int_{\mathbb{R}^n\times\left(\frac{\delta}{4},\frac{\delta}{2}\right)} u^2\mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X \quad (3.27)$$

and

$$\int_{\mathbb{R}^n\times(0,1)} (t+a) \left((t+a)\,\mathrm{e}^{-\mu(t+a)}\right)^{-2\alpha} (u\eta)^2\,\mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X$$
$$\geq \int_{\mathbb{R}^n\times(0,\delta/8)} (t+a) \left((t+a)\,\mathrm{e}^{-\mu(t+a)}\right)^{-2\alpha} (u\eta)^2\,\mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X$$
$$\geq a\left(\left(\frac{\delta}{8}+a\right)\mathrm{e}^{-\mu\left(\frac{\delta}{8}+a\right)}\right)^{-2\alpha} \int_{\mathbb{R}^n\times(0,\delta/8)} u^2\mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X. \quad (3.28)$$

By using (3.27) to estimate from above the right-hand side of (3.26) and by using (3.28) to estimate from below the left-hand side of (3.26) we get

$$\int_{\mathbb{R}^n\times(0,\delta/8)} u^2\mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X$$
$$\leq \frac{C_1}{a\delta^2\alpha} \left(\frac{\left(\frac{\delta}{4}+a\right)\mathrm{e}^{-\mu\left(\frac{\delta}{4}+a\right)}}{\left(\frac{\delta}{8}+a\right)\mathrm{e}^{-\mu\left(\frac{\delta}{8}+a\right)}}\right)^{-2\alpha} \int_{\mathbb{R}^n\times\left(\frac{\delta}{4},\frac{\delta}{2}\right)} u^2\mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X, \quad (3.29)$$

for every $\alpha \geq \left(4\Lambda^2+1\right) C_0$. When α goes to infinity the right-hand side of (3.29) goes to 0 and it follows that $u \equiv 0$ in $\mathbb{R}^n \times (0, \delta/8)$. By iteration we get $u \equiv 0$ in $\mathbb{R}^n \times (0,1)$ and the proof is completed. □

Let P be the following backward parabolic operator

$$Pu = \partial_i \left(g^{ij}(x,t)\, \partial_j u\right) + \partial_t u, \tag{3.30}$$

where $\left\{g^{ij}(x,t)\right\}_{i,j=1}^n$ is a symmetric $n \times n$ matrix whose entries are real functions. When $\xi \in \mathbb{R}^n$ and $(x,t), (y,\tau) \in \mathbb{R}^{n+1}$ we assume that

$$\lambda |\xi|^2 \leq \sum_{i,j=1}^n g^{ij}(x,t)\, \xi_i \xi_j \leq \lambda^{-1} |\xi|^2 \tag{3.31}$$

and

$$\left(\sum_{i,j=1}^n \left(g^{ij}(x,t) - g^{ij}(y,\tau)\right)^2\right)^{1/2} \leq \Lambda \left(|x-y| + |t-\tau|\right), \tag{3.32}$$

where λ and Λ are positive numbers with $\lambda \in (0,1]$.

Let Ω be an open bounded domain in $\mathbb{R}^n$ with Lipschitz boundary $\partial\Omega$. Denote by ν the unit exterior normal to $\partial\Omega$ and denote by $\mathcal{B}$ one of the following boundary operators on $\partial\Omega \times [0,1)$.

(a) $\mathcal{B}u = u$ (boundary Dirichlet operator),
(b) $\mathcal{B}u = \nabla_g u \cdot \nu$ (boundary Neumann operator),
(c) $\mathcal{B}u = \nabla_g u \cdot \nu + \gamma u$ (boundary Robin operator). We assume that γ is a measurable function on $\partial\Omega \times (0,1)$ satisfying the following conditions

$$|\gamma(x,t)| \leq \lambda^{-1}, \quad \text{for every } (x,t) \in \partial\Omega \times (0,1) \tag{3.33}$$

and

$$|\gamma(x,t) - \gamma(x,\tau)| \leq \Lambda |t-\tau|, \quad \text{for every } (x,t), (x,\tau) \in \partial\Omega \times (0,1). \tag{3.34}$$

In the proof of Lemma 3.0.3 and Theorem 3.0.4 below, we shall use the notation introduced in Section 2. In particular, with such notation we have

$$Pu = \Delta_g u + \partial_t u.$$

In the next lemma we shall denote by $\mathcal{H}_{\mathcal{B}}$ the following functional space

$$\mathcal{H}_{\mathcal{B}} = \left\{u \in H^{2,1}(\Omega \times (0,1)) \mid u(x,0) = 0, \mathcal{B}u = 0\right\}.$$

LEMMA 3.0.3. *Let P be the operator defined in (3.30). Assume that (3.31)–(3.34) are satisfied. Then there exist constants C_1, $C_1 \geq 1$, δ_0, $0 < \delta_0 < 1$, depending on λ and Λ only, such that for every $a \in (0, \delta_0)$, $\alpha \geq C_1$ and $w \in \mathcal{H}_{\mathcal{B}}$, with* $\operatorname{supp} w \subset \overline{\Omega} \times [0, \delta_0 - a)$, *the following inequality holds true*

$$\alpha \int_{\Omega\times(0,1)} (t+a)^{-2\alpha-1} w^2 \mathrm{d}X + \int_{\Omega\times(0,1)} (t+a)^{-2\alpha} \left|\nabla_g w\right|_g^2 \mathrm{d}X \leq C_1 \int_{\Omega\times(0,1)} (t+a)^{-2\alpha+1} (Pw)^2 \mathrm{d}X. \tag{3.35}$$

PROOF. Let a be a positive number that we shall choose later on. Let w be a function belonging to $\mathcal{H}_{\mathcal{B}}$, with $\operatorname{supp} w \subset \overline{\Omega} \times [0, 1)$. For any $\alpha > 0$, denote by

$$v(x,t) = (t+a)^{-\alpha} w(x,t).$$

For a number s that we shall choose later on, denote by

$$Lv := (t+a)^{s-\alpha} Pw = (t+a)^s \Delta_g v + \alpha (t+a)^{s-1} v + (t+a)^s \partial_t v.$$

We have

$$Lv = Sv + Av, \tag{3.36}$$

where

$$Sv = (t+a)^s \Delta_g v + \left(\alpha - \frac{s}{2}\right)(t+a)^{s-1} v$$

and

$$Av = \frac{s}{2}(t+a)^{s-1} v + (t+a)^s \partial_t v.$$

Observe that S and A are the symmetric and the skew-symmetric part of the operator L, respectively.

For the sake of brevity, denote by $\int (.)\,\mathrm{d}X$ the integral $\int_{\Omega\times(0,1)} (.)\,\mathrm{d}X$. By (3.36) we obtain

$$\int (Lv)^2 \mathrm{d}X = \int (Av)^2 \mathrm{d}X + \int (Sv)^2 \mathrm{d}X + 2\int AvSv\mathrm{d}X. \tag{3.37}$$

Let us examine the third integral on the right-hand side of (3.37). We have

$$\begin{aligned}\int AvSv\mathrm{d}X &= \left(\alpha - \frac{s}{2}\right)\left(\int \left(\frac{s}{2}(t+a)^{2s-2} v^2 + (t+a)^{2s-1} v\partial_t v\right)\mathrm{d}X\right) \\ &\quad + \frac{s}{2}\int (t+a)^{2s-1} v\Delta_g v\mathrm{d}X + \int (t+a)^{2s} \partial_t v\Delta_g v\mathrm{d}X \\ &:= I_1 + I_2 + I_3.\end{aligned} \tag{3.38}$$

Evaluation of I_1.

By using the identity $v\partial_t v = \frac{1}{2}\partial_t\left(v^2\right)$ and integrating by parts, we get

$$I_1 = \frac{1-s}{2}\left(\alpha - \frac{s}{2}\right)\int (t+a)^{2s-2}\, v^2 \mathrm{d}X. \tag{3.39}$$

Evaluation of I_2.

By (2.2) we have

$$\begin{aligned} I_2 = &-\frac{s}{2}\int (t+a)^{2s-1}\left|\nabla_g v\right|_g^2 \mathrm{d}X \\ &+\frac{s}{2}\int_{\partial\Omega\times(0,1)} (t+a)^{2s-1}\left(\nabla_g v\cdot \nu\right) v\mathrm{d}S. \end{aligned} \tag{3.40}$$

Evaluation of I_3.

By using the identity

$$\partial_t v\Delta_g v = -\frac{1}{2}\partial_t\left(\left|\nabla_g v\right|_g^2\right) + \frac{1}{2}\partial_t g^{ij}\partial_i v\partial_j v + \partial_i\left(g^{ij}\partial_i v\partial_t v\right)$$

and integrating by parts, we obtain

$$\begin{aligned} I_3 = &\, s\int (t+a)^{2s-1}\left|\nabla_g v\right|_g^2 \mathrm{d}X + \frac{1}{2}\int (t+a)^{2s}\,\partial_t g^{ij}\partial_i v\partial_j v\mathrm{d}X \\ &+\int_{\partial\Omega\times(0,1)} (t+a)^{2s}\left(\nabla_g v\cdot\nu\right)\partial_t v\mathrm{d}S. \end{aligned} \tag{3.41}$$

Now, denoting by

$$\begin{aligned} \mathcal{D} = &\,\frac{s}{2}\int_{\partial\Omega\times(0,1)} (t+a)^{2s-1}\left(\nabla_g v\cdot\nu\right) v\mathrm{d}S \\ &+\int_{\partial\Omega\times(0,1)} (t+a)^{2s}\left(\nabla_g v\cdot\nu\right)\partial_t v\mathrm{d}S, \end{aligned} \tag{3.42}$$

by (3.38)–(3.41) we have

$$\begin{aligned} \int AvSv\mathrm{d}X = &\,\frac{1-s}{2}\left(\alpha-\frac{s}{2}\right)\int (t+a)^{2s-2}\, v^2\mathrm{d}X \\ &+\frac{s}{2}\int (t+a)^{2s-1}\left|\nabla_g v\right|_g^2 \mathrm{d}X \\ &+\frac{1}{2}\int (t+a)^{2s}\,\partial_t g^{ij}\partial_i v\partial_j v\mathrm{d}X + \mathcal{D}. \end{aligned}$$

The identity above and (3.32) yields

$$\begin{aligned} \int AvSv\mathrm{d}X \geq &\,\frac{1-s}{2}\left(\alpha-\frac{s}{2}\right)\int (t+a)^{2s-2}\, v^2\mathrm{d}X \\ &+\int (t+a)^{2s-1}\left(\frac{s}{2} - C_0\,(t+a)\right)\left|\nabla_g v\right|_g^2 \mathrm{d}X + \mathcal{D}, \end{aligned} \tag{3.43}$$

where C_0, $C_0 \geq 1$, depends on λ and Λ only.

Now, let $s = \frac{1}{2}$ and let a be such that $0 < a < \frac{1}{8C_0}$. By (3.43) we have

$$\int AvSv\mathrm{d}X \geq \frac{\alpha}{8}\int (t+a)^{-1} v^2\mathrm{d}X + \frac{1}{8}\int \left|\nabla_g v\right|_g^2 \mathrm{d}X + \mathcal{D}, \tag{3.44}$$

for every $v \in \mathcal{H}_{\mathcal{B}}$ such that $\operatorname{supp} v \subset \overline{\Omega} \times \left[0, \frac{1}{8C_0} - a\right)$ and every $\alpha > \frac{1}{4}$.

Now let us examine the term $\mathcal{D}$ at the right-hand side of (3.44).

If either $\mathcal{B}v = v$ or $\mathcal{B}v = \nabla_g v \cdot \nu$, then by (3.42) we have $\mathcal{D} = 0$.

If $\mathcal{B}v = \nabla_g v \cdot \nu + \gamma v$ then by (3.42) and integration by parts, we have

$$\mathcal{D} = \frac{1}{4}\int_{\partial\Omega\times(0,1)} \gamma v^2\mathrm{d}S - \frac{1}{2}\int_{\partial\Omega\times(0,1)} (t+a)\left(\partial_t\gamma\right) v^2\mathrm{d}S.$$

Now recall the following trace inequality, [5]: there exists a constant c_1 depending on the Lipschitz character of $\partial\Omega$ only, such that for any $\varepsilon > 0$ we have

$$\int_{\partial\Omega} h^2 ds \leq \varepsilon \int_{\Omega} |\nabla h|^2 \mathrm{d}x + \frac{c_1}{\varepsilon}\int_{\Omega} h^2\mathrm{d}x,$$

for every $h \in C^1\left(\overline{\Omega}\right)$, where d$s$ is the n-dimensional surface element.

Therefore

$$\begin{aligned}\mathcal{D} &\leq c_2 \int_{\partial\Omega\times(0,1)} v^2\mathrm{d}S\\ &\leq c_3\varepsilon \int_{\Omega\times(0,1)} \left|\nabla_g v\right|_g^2 \mathrm{d}X + \frac{c_1c_3}{\varepsilon}\int_{\Omega\times(0,1)} v^2\mathrm{d}X,\end{aligned}$$

where c_2 and c_3 depend on λ and Λ only.

Now we choose $\varepsilon = \frac{1}{16c_3}$ in the inequality just obtained. By such an inequality and by (3.44) we get

$$\int AvSv\mathrm{d}X \geq \frac{\alpha}{16}\int (t+a)^{-1} v^2\mathrm{d}X + \frac{1}{16}\int \left|\nabla_g v\right|_g^2 \mathrm{d}X, \tag{3.45}$$

for every $v \in \mathcal{H}_{\mathcal{B}}$, such that $\operatorname{supp} v \subset \overline{\Omega} \times [0, \delta_0 - a)$ and every $\alpha > \max\left\{\frac{1}{4}, 16c_1c_3^2\right\}$, where $\delta_0 = \frac{1}{8C_0}$.

Finally by (3.37), (3.44) and (3.45) we get

$$\int (Lv)^2 \mathrm{d}X \geq \frac{\alpha}{16}\int (t+a)^{-1} v^2\mathrm{d}X + \frac{1}{16}\int \left|\nabla_g v\right|_g^2 \mathrm{d}X,$$

for every $v \in \mathcal{H}_{\mathcal{B}}$, such that $\operatorname{supp} v \subset \overline{\Omega} \times [0, \delta_0 - a)$, and every $\alpha > \max\left\{\frac{1}{4}, 16c_1c_2^2\right\}$.

Finally coming back to the function w, we obtain (3.35). □

THEOREM 3.0.4. *Let P be the operator defined in* (3.30). *Assume that* (3.31) and (3.32) *are satisfied and assume that, when $\mathcal{B}u = \nabla_g u \cdot \nu + \gamma u$,* (3.33) and (3.34) *are satisfied.*

Let u be a function belonging to $H^{2,1}(\Omega \times (0,1))$ *which satisfies the inequality*

$$|Pu| \leq \Lambda \left(\left|\nabla_g u\right|_g + |u| \right), \quad \text{in } \Omega \times (0,1) \tag{3.46}$$

and

$$\begin{aligned} &\mathcal{B}u = 0, \quad \text{on } \partial\Omega \times (0,1), \\ &u(x,0) = 0, \quad \text{in } \Omega, \end{aligned} \tag{3.47}$$

then

$$u \equiv 0, \quad \text{in } \Omega \times (0,1).$$

PROOF. Let δ_0 be defined in Lemma 3.0.3, and let δ be a number belonging to $(0, \delta_0)$ that we shall choose later on. Let $\eta \in C^1([0,1])$ be such that $0 \leq \eta \leq 1$, $\eta(t) = 1$, for every $t \in \left[0, \frac{\delta}{4}\right]$, $\eta(t) = 0$, for every $t \in \left[\frac{\delta}{2}, 1\right]$ and, for a constant c_0,

$$\left|\eta'(t)\right| \leq \frac{c_0}{\delta}, \quad \text{for every } t \in \left[\frac{\delta}{4}, \frac{\delta}{2}\right]. \tag{3.48}$$

Let $a = \frac{\delta}{2}$. We have $\eta u \in \mathcal{H}_{\mathcal{B}}$ and $\operatorname{supp} \eta u \subset \overline{\Omega} \times \left[0, \frac{\delta}{2}\right)$, so we can apply inequality (3.35) to the function ηu. Since, by (3.46), we have

$$|P(\eta u)| \leq \Lambda \left(\left|\nabla_g u\right|_g + |u| \right) \eta + \Lambda |u| \left|\eta'\right|, \quad \text{in } \Omega \times (0,1),$$

we have

$$\begin{aligned} &\alpha \int_{\Omega\times(0,1)} (t+a)^{-2\alpha-1} (u\eta)^2 \, \mathrm{d}X + \int_{\Omega\times(0,1)} (t+a)^{-2\alpha} \left|\nabla_g (u\eta)\right|_g^2 \mathrm{d}X \\ &\quad \leq 2C_1\Lambda^2 \int_{\Omega\times(0,1)} (t+a)^{-2\alpha+1} (u\eta)^2 \, \mathrm{d}X \\ &\qquad + 2C_1\Lambda^2 \int_{\Omega\times(0,1)} (t+a)^{-2\alpha+1} \left|\nabla_g (u\eta)\right|_g^2 \mathrm{d}X \\ &\qquad + 2C_1\Lambda^2 \int_{\Omega\times(0,1)} (t+a)^{-2\alpha+1} u^2 \eta'^2 \mathrm{d}X, \end{aligned} \tag{3.49}$$

for every $\alpha \geq C_1$.

If α is large enough, then the first term in the left-hand side of (3.49) dominates the first term of (3.49), so recalling (3.48) we have

$$\begin{aligned} &\alpha \int_{\Omega\times(0,1)} (t+a)^{-2\alpha-1} u^2 \eta^2 \mathrm{d}X \\ &\qquad + \int_{\Omega\times(0,1)} (t+a)^{-2\alpha} \left(1 - 2C_1\Lambda^2\delta\right) \left|\nabla_g u\right|_g^2 \eta^2 \mathrm{d}X \\ &\quad \leq \frac{2C_1 c_0^2 \Lambda^2}{\delta^2} \int_{B_1\times\left(\frac{\delta}{4}, \frac{\delta}{2}\right)} (t+a)^{-2\alpha+1} u^2 \mathrm{d}X, \end{aligned} \tag{3.50}$$

for every $\alpha \geq C_2$ where C_2, $C_2 \geq 1$, depends on λ and Λ only.

Let us choose $\delta = \min\left\{\delta_0, \frac{1}{2C_1\Lambda^2}\right\}$ and by (3.50) we obtain

$$\alpha \int_{\Omega\times(0,1)} (t+a)^{-2\alpha-1} u^2\eta^2 \mathrm{d}X \leq \frac{2C_1c_0^2\Lambda^2}{\delta^2} \int_{\Omega\times\left(\frac{\delta}{4},\frac{\delta}{2}\right)} (t+a)^{-2\alpha+1} u^2 \mathrm{d}X, \tag{3.51}$$

for every $\alpha \geq C_2$.

Now, let us consider the following trivial inequalities

$$\int_{\Omega\times\left(\frac{\delta}{4},\frac{\delta}{2}\right)} (t+a)^{-2\alpha+1} u^2 \mathrm{d}X \leq \left(\frac{3\delta}{4}\right)^{-2\alpha+1} \int_{\Omega\times\left(\frac{\delta}{4},\frac{\delta}{2}\right)} u^2 \mathrm{d}X \tag{3.52}$$

and

$$\int_{\Omega\times(0,1)} (t+a)^{-2\alpha-1} u^2\eta^2 \mathrm{d}X \geq \int_{\Omega\times(0,\delta/8)} (t+a)^{-2\alpha-1} u^2\eta^2 \mathrm{d}X \geq \left(\frac{5\delta}{8}\right)^{-2\alpha-1} \int_{\Omega\times(0,\delta/8)} u^2 \mathrm{d}X. \tag{3.53}$$

By using (3.52) to estimate from above the right-hand side of (3.51) and by using (3.53) to estimate from below the left-hand side of (3.51), we get

$$\int_{\Omega\times(0,\delta/8)} u^2 \mathrm{d}X \leq \frac{C_3}{\alpha}\left(\frac{5}{6}\right)^{2\alpha} \int_{\Omega\times\left(\frac{\delta}{4},\frac{\delta}{2}\right)} u^2 \mathrm{d}X, \tag{3.54}$$

for every $\alpha \geq C_2$, where $C_3 = 2C_1c_0^2\Lambda^2$. When α goes to infinity the right-hand side of (3.54) goes to 0 and it follows that $u = 0$ in $B_1 \times (0, \delta/8)$. By iteration we get $u \equiv 0$ in $B_1 \times (0, 1)$ and the proof is completed. □

Logarithmic convexity method.

In what follows we sketch the proof of the uniqueness and of the stability estimate for the backward problem based on the *logarithmic convexity* method. To this aim we recall the following elementary fact. Let $f(s)$ be a positive smooth function defined on the interval $[0, a]$, where a is a positive number, if f satisfies the inequality

$$F_f(s) := f(s)\ddot{f}(s) - \left(\dot{f}(s)\right)^2 \geq 0, \quad \text{for every } s \in [0, a], \tag{3.55}$$

then

$$f(s) \leq (f(0))^{1-\frac{s}{a}} (f(a))^{\frac{s}{a}}, \quad \text{for every } s \in [0, a]. \tag{3.56}$$

Indeed, (3.55) is equivalent to the convexity of $\log f(s)$.

Let $\left\{g^{ij}(x,t)\right\}_{i,j=1}^{n}$ satisfy (3.31) and (3.32) and let $u \in H^{2,1}(\Omega \times (0,1))$ be a solution to the equation

$$Pu = \partial_i\left(g^{ij}(x,t)\,\partial_j u\right) + \partial_t u = 0, \quad \text{in } \Omega \times (0,1)\,. \tag{3.57}$$

Assume that u satisfies the boundary Dirichlet condition

$$u = 0, \quad \text{on } \partial\Omega \times (0,1)\,. \tag{3.58}$$

Denote by

$$\phi(t) = \int_{B_1} u^2(x,t)\,\mathrm{d}x, \quad \text{for } t \in [0,1]\,. \tag{3.59}$$

Let $\theta(s)$ be a function that we shall choose later on, θ is an increasing positive function with values in $[0,1]$, $\theta(0) = 0$. Denote by

$$f(s) = \phi(\theta(s))\,.$$

By the chain rule we have

$$F_f(s) = F_\phi(\theta(s))\left(\dot{\theta}(s)\right)^2 + \phi'((\theta(s)))\,\phi(\theta(s))\,\ddot{\theta}(s)\,. \tag{3.60}$$

By differentiating ϕ we obtain

$$\phi'(t) = 2\int_\Omega u(x,t)\,\partial_t u(x,t)\,\mathrm{d}x. \tag{3.61}$$

In addition, by (3.57) and integrating by parts, we have

$$\begin{aligned}\int_{B_1} u(x,t)\,\partial_t u(x,t)\,\mathrm{d}x &= -\int_\Omega u(x,t)\,\partial_i\left(g^{ij}(x,t)\,\partial_j u(x,t)\right)\mathrm{d}x \\ &= \int_\Omega g^{ij}(x,t)\,\partial_j u(x,t)\,\partial_i u(x,t)\,\mathrm{d}x.\end{aligned}$$

Therefore, besides (3.61), we have

$$\phi'(t) = 2\int_\Omega g^{ij}(x,t)\,\partial_j u(x,t)\,\partial_i u(x,t)\,\mathrm{d}x. \tag{3.62}$$

Now we use this expression to differentiate $\phi'(t)$. By integrating by parts and using (3.57), we get

$$\begin{aligned}\phi''(t) &= 4\int_\Omega g^{ij}(x,t)\,\partial_j u(x,t)\,\partial_i\partial_t u(x,t)\,\mathrm{d}x \\ &\quad + 2\int_\Omega \partial_t g^{ij}(x,t)\,\partial_j u(x,t)\,\partial_i u(x,t)\,\mathrm{d}x \\ &= -4\int_\Omega \partial_i\left(g^{ij}(x,t)\,\partial_j u(x,t)\right)\partial_t u(x,t)\,\mathrm{d}x \\ &\quad + 2\int_\Omega \partial_t g^{ij}(x,t)\,\partial_j u(x,t)\,\partial_i u(x,t)\,\mathrm{d}x \\ &= 4\int_\Omega (\partial_t u(x,t))^2\,\mathrm{d}x + 2\int_\Omega \partial_t g^{ij}(x,t)\,\partial_j u(x,t)\,\partial_i u(x,t)\,\mathrm{d}x.\end{aligned}$$

Hence

$$\begin{aligned}F_\phi(t) &= 4\left(\int_\Omega (\partial_t u(x,t))^2\,\mathrm{d}x\right)\left(\int_\Omega u^2(x,t)\,\mathrm{d}x\right) \\ &\quad - 4\left(\int_\Omega u(x,t)\,\partial_t u(x,t)\,\mathrm{d}x\right)^2 \\ &\quad + 2\left(\int_\Omega u^2(x,t)\,\mathrm{d}x\right)\left(\int_\Omega \partial_t g^{ij}(x,t)\,\partial_j u(x,t)\,\partial_i u(x,t)\,\mathrm{d}x\right).\end{aligned}$$

By the Cauchy–Schwartz inequality and (3.32), we obtain

$$\begin{aligned}F_\phi(t) &\geq 2\left(\int_\Omega u^2(x,t)\,\mathrm{d}x\right)\left(\int_\Omega \partial_t g^{ij}(x,t)\,\partial_j u(x,t)\,\partial_i u(x,t)\,\mathrm{d}x\right) \\ &\geq -C_1\left(\int_\Omega u^2(x,t)\,\mathrm{d}x\right)\left(\int_\Omega |\nabla u(x,t)|^2\,\mathrm{d}x\right),\end{aligned}$$

where C_1, $C_1 \geq 1$, depends on Λ only. Hence, recalling (3.60), (3.59) and (3.62),we have

$$\begin{aligned}F_f(s) &\geq \left(2\lambda\,\ddot{\theta}(s) - C_1\left(\dot{\theta}(s)\right)^2\right)\left(\int_{B_1} u^2(x,\theta(s))\,\mathrm{d}x\right) \\ &\quad \times\left(\int_{B_1} |\nabla u(x,\theta(s))|^2\,\mathrm{d}x\right).\end{aligned} \tag{3.63}$$

Now, we choose θ such that

$$\theta(s) = -\frac{1}{C_2}\log(1 - C_2 s), \quad \text{for } s \in \left[0, \frac{1}{2C_2}\right],$$

where $C_2 = \frac{C_1}{2\lambda}$. The function θ satisfies the equations

$$2\lambda\,\ddot{\theta}(s) - C_1\left(\dot{\theta}(s)\right)^2 = 0, \quad \theta(0) = 0, \quad \dot{\theta}(0) = 1.$$

Therefore by (3.63) we have

$$F_f(s) \geq 0, \quad \text{for every } s \in \left[0, \frac{1}{2C_2}\right]. \tag{3.64}$$

By (3.56) and (3.64) we conclude that, denoting by $\delta = \min\left\{\frac{1}{2C_2}, \frac{1-e^{-C_2}}{C_2}\right\}$, $T = \theta(\delta)$, $\mu(t) = \frac{1-e^{-C_2 t}}{\delta C_2}$, the following *stability estimate* holds true, for every $t \in [0, T]$,

$$\int_\Omega u^2(x,t)\,dx \leq \left(\int_\Omega u^2(x,0)\,dx\right)^{1-\mu(t)} \left(\int_\Omega u^2(x,T)\,dx\right)^{\mu(t)}. \tag{3.65}$$

The above inequality implies, in particular, that if u, $u \in H^{2,1}(\Omega \times (0,1))$, satisfies (3.57) and (3.58) and

$$u(.,0) = 0, \quad \text{in } \Omega,$$

then $u(.,t) = 0$, $t \in [0,T)$ and, by iteration, we have $u \equiv 0$ in $\Omega \times (0,1)$.

4. Spacelike unique continuation properties and their quantitative versions

In Section 4.1 we consider the case of parabolic operators whose coefficients do not depend on t and we present the so-called elliptic continuation technique due to Landis and Oleinik, [68]. In the next two sub-sections we consider the parabolic operators whose coefficients do depend on t. In Section 4.2 we prove the two-sphere one-cylinder inequality at the interior and at the (time varying) boundary. In Section 4.3 we prove some sharp stability estimates for the Cauchy problem for parabolic equations.

4.1. *Parabolic equations with time independent coefficients*

In this section we present some quantitative estimates of unique continuation for solutions to parabolic equations whose coefficients do not depend on t. The method for deriving such estimates has been introduced in [68] and, concerning the smoothness assumption on the coefficients, has been improved in [73]. In what follows we give an outline of the above method and we limit ourselves to present detailed proofs only of the main points of the method, we omit the most technical proofs for which we refer to the quoted literature and especially to [17]. Throughout this section, for any positive number r, we shall denote by $\widetilde{B}_r$ the $(n+1)$-dimensional open sphere of radius r centered at 0 and we shall denote by $\widetilde{B}_r^+ = \left\{x \in \widetilde{B}_r : x_{n+1} > 0\right\}$.

In order to simplify the exposition we consider the equation

$$Lu := \operatorname{div}(a(x)\nabla u) - \partial_t u = 0, \quad \text{in } B_1 \times (0,1], \tag{4.1}$$

where $a(x) = \left\{a^{ij}(x)\right\}_{i,j=1}^{n}$ is a symmetric $n \times n$ matrix whose entries are real-valued functions. When $\xi \in \mathbb{R}^n$ and $x, y \in \mathbb{R}^n$ we assume that

$$\lambda |\xi|^2 \leq \sum_{i,j=1}^{n} a^{ij}(x)\, \xi_i \xi_j \leq \lambda^{-1} |\xi|^2 \tag{4.2}$$

and

$$\left(\sum_{i,j=1}^{n} \left(a^{ij}(x) - a^{ij}(y)\right)^2\right)^{1/2} \leq \Lambda\, |x - y|\,, \tag{4.3}$$

where λ and Λ are positive numbers with $\lambda \in (0, 1]$.

Let $u \in H^{2,1}(B_1 \times (0, 1))$ be a solution to (4.1). Denoting by $\overline{u}$ the following extension of u

$$\overline{u} = \begin{cases} u(x,t)\,, & \text{if } (x,t) \in B_1 \times (0,1)\,, \\ -3u(x, 2-t) + 4u\left(x, \dfrac{3}{2} - \dfrac{1}{2}t\right), & \text{if } (x,t) \in B_1 \times [1,2)\,, \end{cases} \tag{4.4}$$

we have $\overline{u} \in H^{2,1}(B_1 \times (0, 2))$ and, [34],

$$\|\overline{u}\|_{H^{2,1}(B_1 \times (0,2))} \leq C\, \|u\|_{H^{2,1}(B_1 \times (0,1))}\,, \tag{4.5}$$

where C is an absolute constant.

Let $\eta \in C^1([0, +\infty))$ be a function satisfying $0 \leq \eta \leq 1$, $\eta = 1$ in $[0, 1]$, $\eta = 0$ in $[2, +\infty)$ and $\left|\eta'\right| \leq c$, in $[1, 2]$. Moreover, denoting by $\widetilde{u}$ the trivial extension of $\eta(t)\,\overline{u}(x,t)$ to $(0, +\infty)$ (i.e. $\widetilde{u}(x,t) = 0$ if $t \geq 2$), let us denote by u_1, u_2 the weak solutions to the following initial-boundary value problems, respectively,

$$\begin{cases} Lu_1 = 0, & \text{in } B_1 \times (0, +\infty)\,, \\ u_1 = 0, & \text{on } \partial B_1 \times (0, +\infty)\,, \\ u_1(.,0) = u(.,0)\,, & \text{in } B_1 \end{cases} \tag{4.6}$$

and

$$\begin{cases} Lu_2 = 0, & \text{in } B_1 \times (0, +\infty)\,, \\ u_2 = \widetilde{u}, & \text{on } \partial B_1 \times (0, +\infty)\,, \\ u_2(.,0) = 0, & \text{in } B_1. \end{cases} \tag{4.7}$$

Since $u_1 + u_2 = u$ on $\partial B_1 \times (0, 1]$ and $(u_1 + u_2)(.,0) = u(.,0)$ in B_1, by the uniqueness theorem for initial-boundary value problems for parabolic equations, we have

$$u_1(x,1) + u_2(x,1) = u(x,1), \quad \text{for every } x \in B_1. \tag{4.8}$$

In what follows, we shall denote by $C_{\mathcal{P}}$ the constant appearing in the following Poincaré inequality

$$\int_{B_1} f^2(x)\,\mathrm{d}x \le C_{\mathcal{P}} \int_{B_1} |\nabla f(x)|^2\,\mathrm{d}x, \quad \text{for every } f \in H_0^1(B_1),$$

where we recall that $C_{\mathcal{P}} \le 1$, [41], and $C_{\mathcal{P}} = \frac{1}{k_0^2}$, where k_0 is the smallest positive root of the Bessel function of first kind $J_{\frac{n-2}{2}}$, [23]. Let us denote

$$b_1 = \frac{\lambda}{C_{\mathcal{P}}}, \tag{4.9}$$

and

$$H = \sup_{t\in[0,1]} \|u(.,t)\|_{H^1(B_1)}. \tag{4.10}$$

PROPOSITION 4.1.1. *Let u and u_2 be as above. We have*

$$\|u_2(.,t)\|_{H^1(B_1)} \le C_1 H \mathrm{e}^{-b_1(t-2)_+}, \quad \textit{for every } t \in (0,+\infty), \tag{4.11}$$

($(t-2)_+$ is equal to $t-2$ if $t \ge 2$ and it is equal to 0 if $t < 2$) where C_1, $C_1 > 1$, depends on λ and Λ only.

PROOF. Let

$$v = u_2 - \widetilde{u},$$

where $\widetilde{u}$ is the function defined above. Denoting $F = -L\widetilde{u}$, we have

$$\begin{cases} Lv = F, & \text{in } B_1 \times (0,+\infty), \\ v = 0, & \text{on } \partial B_1 \times (0,+\infty), \\ v(.,0) = -u(.,0), & \text{in } B_1. \end{cases} \tag{4.12}$$

We claim

$$\|u_2(.,t)\|_{L^2(B_1)} \le CH, \quad \text{for every } t \in [0,2], \tag{4.13}$$

$$\|\nabla u_2(.,t)\|_{L^2(B_1)} \le CH, \quad \text{for every } t \in [0,2], \tag{4.14}$$

where C depends on λ and Λ only.

In order to prove (4.13) let us multiply the first equation in ((4.12) by v and integrate over $B_1 \times (0,t)$. We get, for every $t \in [0,2]$,

$$\int_{B_1} v^2(x,t)\,\mathrm{d}x \le 2\|F\|^2_{L^2(B_1\times(0,2))} + \int_{B_1} u^2(x,0)\,\mathrm{d}x + 2\int_0^t \mathrm{d}\tau \int_{B_1} v^2(x,\tau)\,\mathrm{d}x.$$

By applying Gronwall's inequality we have

$$\int_{B_1} v^2(x,t)\,\mathrm{d}x \le \mathrm{e}^4\left(2\|F\|^2_{L^2(B_1\times(0,2))} + H^2\right), \quad \text{for every } t\in[0,2]. \tag{4.15}$$

Now by (4.5) and the regularity estimate for parabolic equations we have

$$\|F\|_{L^2(B_1\times(0,2))} \le CH, \tag{4.16}$$

where C depends on λ and Λ only. From this inequality and (4.15) we obtain (4.13).

In order to prove (4.14), let us multiply the first equation in (4.12) by $\partial_t v$ and integrate over $B_1\times(0,t)$. We obtain

$$\lambda\int_{B_1} |\nabla v(x,t)|^2\,\mathrm{d}x \le \lambda^{-1}\int_{B_1} |\nabla u(x,0)|^2\,\mathrm{d}x + 2\|F\|^2_{L^2(B_1\times(0,2))},$$

so, by (4.10) and (4.16) we have (4.14). The claims are thus proved.

Now let us denote by μ_k, $k\in\mathbb{N}$, the decreasing sequence of eigenvalues associated with the problem

$$\begin{cases}\operatorname{div}(a\nabla\varphi) = \mu\varphi, & \text{in } B_1,\\ \varphi = 0, & \text{on } \partial B_1\end{cases} \tag{4.17}$$

and by φ_k, $k\in\mathbb{N}$, the corresponding eigenfunctions normalized by

$$\int_{B_1} \varphi_k^2(x)\,\mathrm{d}x = 1, \quad k\in\mathbb{N}.$$

We have

$$0 > -b_1 \ge \mu_1 \ge \mu_2 \ge \cdots \ge \mu_k \ge \cdots. \tag{4.18}$$

Since $u_2 = 0$, on $\partial B_1\times[2,+\infty)$, we have

$$u_2(x,t) = \sum_{k=1}^{\infty} \beta_k\varphi_k(x)\,\mathrm{e}^{\mu_k(t-2)}, \quad \text{for every } t\ge 2, \tag{4.19}$$

where

$$\beta_k = \int_{B_1} u_2(x,2)\,\varphi_k(x)\,\mathrm{d}x, \quad k\in\mathbb{N}.$$

Since $u_2(.,2) = v(.,2)$ in B_1, by (4.13), (4.18) and (4.19) we have

$$\int_{B_1} u_2^2(x,t)\,\mathrm{d}x = \sum_{k=1}^{\infty} \beta_k^2 \mathrm{e}^{2\mu_k(t-2)} \le \mathrm{e}^{-2b_1(t-2)} \int_{B_1} u_2^2(x,2)\,\mathrm{d}x \le CH^2 \mathrm{e}^{-2b_1(t-2)}, \quad \text{for every } t \ge 2, \tag{4.20}$$

where C depends on λ and Λ only.

Moreover, for every $t \ge 2$, we have

$$\int_{B_1} a(x) \nabla u_2(x,t) \cdot \nabla u_2(x,t)\,\mathrm{d}x = -\int_{B_1} \partial_t u_2(x,t)\, u_2(x,t)\,\mathrm{d}x = \sum_{k=1}^{\infty} |\mu_k|\, \beta_k^2 \mathrm{e}^{2\mu_k(t-2)}. \tag{4.21}$$

By choosing $t = 2$ in (4.21) and using (4.14) we get

$$\sum_{k=1}^{\infty} |\mu_k|\, \beta_k^2 \le CH^2, \tag{4.22}$$

where C depends on λ and Λ only.

By (4.18), (4.21) and (4.22) we derive

$$\int_{B_1} |\nabla u_2(x,t)|^2\,\mathrm{d}x \le CH^2 \mathrm{e}^{-2b_1(t-2)}, \quad \text{for every } t \ge 2, \tag{4.23}$$

where C depends on λ and Λ only.

Finally, from (4.13), (4.14), (4.20) and (4.23) we get (4.11). □

Let us still denote by u_2 the extension by 0 of u_2 to $B_1 \times \mathbb{R}$ and let us consider the Fourier transform of u_2 with respect to the t variable

$$\widehat{u}_2(x,\mu) = \int_{-\infty}^{+\infty} \mathrm{e}^{-i\mu t} u_2(x,t)\,\mathrm{d}t = \int_{0}^{+\infty} \mathrm{e}^{-i\mu t} u_2(x,t)\,\mathrm{d}t, \quad \mu \in \mathbb{R}. \tag{4.24}$$

We have that $\widehat{u}_2$ satisfies

$$\operatorname{div}(a(x) \nabla \widehat{u}_2(x,\mu)) - \mathrm{i}\mu \widehat{u}_2(x,\mu) = 0, \quad \text{if } (x,\mu) \in B_1 \times \mathbb{R}. \tag{4.25}$$

PROPOSITION 4.1.2. *Let $\widehat{u}_2$ be as above. We have, for every $\mu \in \mathbb{R}$,*

$$\|\widehat{u}_2(.,\mu)\|_{L^2(B_{1/2})} \le cC_1 \lambda^{-1} H \left(2 + \frac{1}{b_1}\right) \mathrm{e}^{-|\mu|^{1/2}\delta}, \tag{4.26}$$

where c is an absolute constant, C_1 is the constant that appears in ((4.11) and δ is given by

$$\delta = \frac{\lambda}{8\pi e}. \tag{4.27}$$

PROOF. Let us denote, for every $\mu, \xi \in \mathbb{R}$, $x \in B_1$

$$v\,(x, \xi; \mu) = \mathrm{e}^{i|\mu|^{1/2}\xi}\widehat{u}_2\,(x, \mu)\,.$$

For every $\mu \in \mathbb{R} \setminus \{0\}$, the function $v\,(.,.;\mu)$ solves the uniformly elliptic equation

$$\mathrm{div}\,(a\,(x)\,\nabla v\,(x, \xi; \mu)) + \mathrm{i}\,\mathrm{sgn}\,(\mu)\,\partial_\xi^2 v\,(x, \xi; \mu) = 0, \quad \text{in } B_1 \times \mathbb{R}. \tag{4.28}$$

Let $k \in \mathbb{N}$ and denote by $r_m = 1 - \frac{m}{2k}$, for every $m \in \{0, 1, \ldots, k\}$. Moreover, set

$$h_m\,(s) = \begin{cases} 0, & \text{if } |s| > r_m, \\ \dfrac{1}{2}\,(1 + \cos 2\pi k\,(r_{m+1} - s))\,, & \text{if } r_{m+1} \le |s| \le r_m, \\ 1, & |s| < r_{m+1}, \end{cases}$$
$$\eta_m\,(x, \xi) = h_m\,(|x|)\,h_m\,(|\xi|)$$

and

$$v_m\,(x, \xi; \mu) = \partial_\xi^m v\,(x, \xi; \mu)\,, \quad m \in \{0, 1, \ldots, k\}\,.$$

We have that $v_m\,(.,.;\mu)$ satisfies

$$\mathrm{div}\,(a\,(x)\,\nabla v_m\,(x, \xi; \mu)) + \mathrm{isgn}\,(\mu)\,\partial_\xi^2 v_m\,(x, \xi; \mu) = 0, \quad \text{in } B_1 \times \mathbb{R}. \tag{4.29}$$

Multiplying Eq. (4.29) by $\overline{v}_m\,(x, \xi; \mu)\,\eta_m^2\,(x, \xi)$ (here $\overline{v}_m$ denotes the complex conjugate of v_m) and integrating over $D_m = B_{r_m} \times (-r_m, r_m)$, we obtain

$$\int_{D_m} (a\nabla v_m \cdot \nabla \overline{v}_m)\,\eta_m^2 \mathrm{d}x\mathrm{d}\xi + \int_{D_m} \left|\partial_\xi v_m\right|^2 \eta_m^2 \mathrm{d}x\mathrm{d}\xi \le C_2 m^2 \int_{D_m} |v_m|^2\,\eta_m^2 \mathrm{d}x\mathrm{d}\xi,$$

where $C_2 = \frac{8\sqrt{2}\pi^2}{\lambda}$.

Therefore, for every $m \in \{0, 1, \ldots, k\}$, we have

$$\int_{D_{m+1}} \left|\partial_\xi^{m+1} v\right|^2 \mathrm{d}x\mathrm{d}\xi \le \frac{C_2 m^2}{\lambda} \int_{D_m} \left|\partial_\xi^m v\right|^2 \mathrm{d}x\mathrm{d}\xi. \tag{4.30}$$

By iteration of (4.30) for $m = 0, 1, \ldots, k-1$, we get

$$\int_{B_{1/2}\times(-1/2,1/2)} \left|\partial_\xi^k v\right|^2 \mathrm{d}x\mathrm{d}\xi \le 2\left(\frac{C_2 k^2}{\lambda}\right)^k \int_{B_1} |\widehat{u}_2\,(x, \mu)|^2\,\mathrm{d}x. \tag{4.31}$$

Now, let us estimate the integral at the right-hand side of (4.31). By (4.11) and the Schwartz inequality we have

$$\begin{aligned}\int_{B_1} |\widehat{u}_2(x,\mu)|^2 \,\mathrm{d}x &= \int_{B_1} \left|\int_0^{+\infty} \mathrm{e}^{-i\mu t} u_2(x,t)\,\mathrm{d}t\right|^2 \mathrm{d}x \\ &\le \int_{B_1} \mathrm{d}x \left(\int_0^{+\infty} \mathrm{e}^{-b_1(t-2)_+}\mathrm{d}t\right)\left(\int_0^{+\infty} \mathrm{e}^{-b_1(t-2)_+} u_2^2(x,t)\,\mathrm{d}t\right) \\ &\le cC_1^2\left(2+\frac{1}{b_1}\right)^2,\end{aligned}$$

where c is an an absolute constant and C_1 is the constant that appears in (4.11).

By the inequality just obtained and by (4.31) we get, for every $k \in \mathbb{N}$,

$$\int_{B_{1/2}\times(-1/2,1/2)} \left|\partial_\xi^k v\right|^2 \mathrm{d}x\mathrm{d}\xi \le cC_1^2 H^2 \left(2+\frac{1}{b_1}\right)^2 \left(\frac{C_2 k^2}{\lambda}\right)^k, \tag{4.32}$$

where c is an absolute constant.

For fixed $\mu \in \mathbb{R} \setminus \{0\}$ and $\psi \in L^2\left(B_{1/2}, \mathbb{C}\right)$, let us denote by Ψ the function

$$\Psi(\xi) = \int_{B_{1/2}} v(x,\xi;\mu)\,\overline{\psi(x)}\mathrm{d}x, \quad \xi \in \left(-\frac{1}{2},\frac{1}{2}\right).$$

Recall now the inequality

$$\|f\|_{L^\infty(J)} \le c\left(|J|\,\|f\|_{L^2(J)}^2 + |J|^{-1}\left\|f'\right\|_{L^2(J)}^2\right)^{1/2}, \tag{4.33}$$

where J is a bounded interval of $\mathbb{R}$, $|J|$ is the length of J and c is an absolute constant.

By (4.32) and (4.33), we have that for every $k \in \mathbb{N}\cup\{0\}$ and $\xi \in \left(-\frac{1}{2},\frac{1}{2}\right)$,

$$\left|\Psi^{(k)}(\xi)\right| \le cC_1 H \|\psi\|_{L^2(B_{1/2})}\left(2+\frac{1}{b_1}\right)\left(\left(C_2\lambda^{-1}\right)^{1/2} k\right)^k. \tag{4.34}$$

By using inequality (4.34) and the power series of Ψ at any point ξ_0 such that $\mathrm{Re}(\xi_0) \in \left(-\frac{1}{2},\frac{1}{2}\right)$, $\mathrm{Re}(\xi_0) = 0$, we have that the function Ψ can be analytically extended in the rectangle

$$Q = \left\{\xi \in \mathbb{C} : \mathrm{Re}(\xi) \in \left(-\frac{1}{2},\frac{1}{2}\right), \mathrm{Im}(\xi) \in (-2\delta, 2\delta)\right\},$$

where $\delta = \frac{\lambda}{8\pi e}$. Denoting by $\widetilde{\Psi}$ such an analytic extension of Ψ to Q we have

$$\left|\widetilde{\Psi}(-\mathrm{i}\delta)\right| \le c\lambda^{-1} C_1 H \|\psi\|_{L^2(B_{1/2})}\left(2+\frac{1}{b_1}\right), \tag{4.35}$$

where c is an absolute constant.

Conversely, by the definition of v, we have

$$\widetilde{\Psi}\left(-\mathrm{i}\delta\right)=\int_{B_{1/2}}\mathrm{e}^{|\mu|^{1/2}\delta}\widehat{u}_2\left(x,\mu\right)\overline{\psi\left(x\right)},$$

so that, by (4.35), we obtain (4.26). □

Estimate (4.26) allows us to define, for every $x\in B_{1/2}$ and $y\in\left(-\sqrt{2}\delta,\sqrt{2}\delta\right)$, the function

$$w_2\left(x,y\right)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\mathrm{e}^{\mathrm{i}\mu}\widehat{u}_2\left(x,\mu\right)\cosh\left(\sqrt{-\mathrm{i}\mu}\,y\right)\mathrm{d}\mu, \tag{4.36}$$

where

$$\sqrt{-i\mu}=|\mu|^{1/2}\,\mathrm{e}^{-\frac{\pi}{4}i\,\mathrm{sgn}(\mu)}.$$

It turns out that w_2 satisfies the following elliptic equation

$$\mathrm{div}\left(a\left(x\right)\nabla w_2\left(x,y\right)\right)+\partial_y^2 w_2\left(x,y\right)=0,\quad\text{in }B_{1/2}\times\left(-\sqrt{2}\delta,\sqrt{2}\delta\right) \tag{4.37}$$

and the following conditions at $y=0$

$$w_2\left(x,0\right)=u_2\left(x,1\right),\quad\text{for every }x\in B_{1/2}, \tag{4.38}$$

$$\partial_y w_2\left(x,0\right)=0,\quad\text{for every }x\in B_{1/2}. \tag{4.39}$$

Moreover, notice that w_2 is an even function with respect to the variable y.

Now let us consider the function u_1. Since such a function is the solution of the initial-boundary value problem (4.6), we have

$$u_1\left(x,t\right)=\sum_{k=1}^{\infty}\alpha_k\mathrm{e}^{\mu_k t}\varphi_k\left(x\right), \tag{4.40}$$

where μ_k and $\varphi_k\left(x\right)$, for $k\in\mathbb{N}$, are, respectively, the decreasing sequence of eigenvalues and the corresponding eigenfunctions associated with the problem

$$\begin{cases}\mathrm{div}\left(a\nabla\varphi\right)=\mu\varphi, & \text{in }B_1,\\ \varphi=0, & \text{on }\partial B_1\end{cases}$$

and

$$\alpha_k=\int_{B_1}u_1\left(x,0\right)\varphi_k\left(x\right)\mathrm{d}x,\quad k\in\mathbb{N}.$$

Let us define

$$w_1(x,y)=\sum_{k=1}^{\infty}\alpha_k e^{\mu_k}\varphi_k(x)\cosh\left(\sqrt{|\mu_k|}y\right). \tag{4.41}$$

It is easy to check that w_1 is a solution to Eq. (4.37) and satisfies the following conditions

$$w_1(x,0)=u_1(x,1), \quad \text{for every } x\in B_1, \tag{4.42}$$
$$\partial_y w_1(x,0)=0, \quad \text{for every } x\in B_1. \tag{4.43}$$

Moreover, notice that w_1 is an even function with respect to the variable y.

Now, let us define

$$w=w_1+w_2, \quad \text{in } B_{1/2}\times\left(-\sqrt{2}\delta,\sqrt{2}\delta\right), \tag{4.44}$$

we have that w is again a solution to Eq. (4.37), it is an even function with respect to the variable y and, as a consequence of (4.8), (4.38) and (4.42), we have that $w(x,0)=u(x,1)$ for every $x\in B_{1/2}$. In addition by (4.39) and (4.43) we get $\partial_y w(x,0)=0$, for every $x\in B_{1/2}$. Also, observe that by (4.26), (4.36) and (4.41) we derive

$$\|w(.,y)\|_{L^2(B_{1/2})}\le c\frac{C_1H}{\lambda\delta}\left(1+\frac{1}{b_1}\right), \quad \text{for every } y\in\left(-\frac{\delta}{2},\frac{\delta}{2}\right). \tag{4.45}$$

Indeed, if $y\in\left(-\frac{\delta}{2},\frac{\delta}{2}\right)$ then, by the Schwartz inequality, we have

$$\begin{aligned}\|w_2(.,y)\|^2_{L^2(B_{1/2})}&=\int_{B_{1/2}}\mathrm{d}x\left|\frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{i\mu}\widehat{u}_2(x,\mu)\cosh\left(\sqrt{-i\mu}y\right)\mathrm{d}\mu\right|^2\\&\le\frac{1}{(2\pi)^2}\int_{B_{1/2}}\left(\int_{-\infty}^{+\infty}|\widehat{u}_2(x,\mu)|^2e^{2\delta|\mu|^{1/2}}\mathrm{d}\mu\right)\\&\quad\times\left(\int_{-\infty}^{+\infty}e^{-\delta\left(2-\sqrt{2}\right)|\mu|^{1/2}}\mathrm{d}\mu\right)\\&\le c\left(\frac{C_1H}{\lambda\delta}\left(1+\frac{1}{b_1}\right)\right)^2,\end{aligned}$$

where c is an absolute constant and C_1 is the constant that appears in (4.11).
Moreover, for every $y\in\mathbb{R}$ we have

$$\begin{aligned}\|w_1(.,y)\|^2_{L^2(B_1)}&=\sum_{k=1}^{\infty}\alpha_k^2e^{2\mu_k}\cosh^2\left(\sqrt{|\mu_k|}y\right)\\&\le\sum_{k=1}^{\infty}\alpha_k^2=\|u(.,0)\|^2_{L^2(B_1)}\le H^2.\end{aligned}$$

By the inequalities obtained above, we get (4.45).

In the following proposition we summarize the results obtained above.

PROPOSITION 4.1.3. *Let w be the function defined in (4.44). Then w is a solution to the equation*

$$\operatorname{div}\left(a\left(x\right)\nabla w\left(x,y\right)\right)+\partial_y^2 w\left(x,y\right)=0,\quad \text{in } B_{1/2}\times\left(-\sqrt{2}\delta,\sqrt{2}\delta\right) \tag{4.46}$$

and w satisfies the following conditions

$$w\left(x,0\right)=u\left(x,1\right),\quad \text{for every } x\in B_1, \tag{4.47}$$

$$\partial_y w\left(x,0\right)=0,\quad \text{for every } x\in B_1. \tag{4.48}$$

Moreover w is an even function with respect to the variable y, and w satisfies inequality (4.45).

The following lemma is nothing more than a stability estimate for the elliptic Cauchy problem (4.46)-(4.48). A detailed proof of such a lemma can be found in Lemma 3.1.5 of [17].

LEMMA 4.1.4. *Let w be the function defined in (4.44). Then there exist constants C, $C>1$, γ_0, β, $\gamma_0,\beta\in(0,1)$, depending on λ only, such that for every $r\leq\frac{1}{2}\gamma_0$ the following estimate holds true*

$$\int_{\widetilde{B}_r} w^2 \mathrm{d}x\mathrm{d}y\leq C\left(\int_{B_{\rho/2}} w^2\left(x,0\right)\mathrm{d}x\right)^{\beta}\left(\int_{\widetilde{B}_{\widetilde{\rho}}} w^2\mathrm{d}x\mathrm{d}y\right)^{1-\beta}, \tag{4.49}$$

where $\rho=\frac{8\sqrt{2}e\pi}{3\lambda}r$ and $\widetilde{\rho}=2\sqrt{2}\rho$.

Now we recall the three sphere inequality for elliptic equations [62]. In what follows we denote by $\widetilde{a}\left(x,y\right)=\left\{\widetilde{a}^{ij}\left(x,y\right)\right\}_{i,j=1}^{n+1}$ a symmetric $(n+1)\times(n+1)$ matrix whose entries are real valued functions. When $\xi\in\mathbb{R}^{n+1}$ and $(x,y)\in\mathbb{R}^{n+1}$ we assume that

$$\widetilde{\lambda}\left|\xi\right|^2\leq\sum_{i,j=1}^{n+1}\widetilde{a}^{ij}\left(x,y\right)\xi_i\xi_j\leq\widetilde{\lambda}^{-1}\left|\xi\right|^2 \tag{4.50}$$

and, for every $\left(x_1,y_1\right),\left(x_2,y_2\right)\in\mathbb{R}^{n+1}$,

$$\left(\sum_{l,j=1}^{n+1}\left(\widetilde{a}^{ij}\left(x_1,y_1\right)-\widetilde{a}^{ij}\left(x_2,y_2\right)\right)^2\right)^{1/2}\leq\widetilde{\Lambda}\left(\left|x_1-x_2\right|+\left|y_1-y_2\right|\right), \tag{4.51}$$

where $\widetilde{\lambda}$ and $\widetilde{\Lambda}$ are positive numbers with $\widetilde{\lambda}\in(0,1]$.

LEMMA 4.1.5 (*three sphere inequality for elliptic equations*). *Let* $\widetilde{a}(x,y) = \left\{\widetilde{a}^{lj}(x,y)\right\}_{l,j=1}^{n+1}$ *satisfy* (4.50) and (4.51). *Let* $\widetilde{w} \in H^1\left(\widetilde{B}_1\right)$ *be a weak solution to*

$$\operatorname{div}\left(\widetilde{a}\nabla\widetilde{w}\right) = 0, \quad \text{in } \widetilde{B}_1. \tag{4.52}$$

Then there exist constants γ_1, $\gamma_1 \in (0,1)$, C, $C > 1$, *depending on* $\widetilde{\lambda}$ *and* $\widetilde{\Lambda}$ *only, such that for every* r_1, r_2, r_3 *that satisfy* $0 < r_1 \le r_2 \le \frac{\lambda}{2} r_3 \le \gamma_1$ *we have*

$$\int_{\widetilde{B}_{r_2}} \widetilde{w}^2 \mathrm{d}x\mathrm{d}y \le C\left(\frac{r_3}{r_2}\right)^C \left(\int_{\widetilde{B}_{r_1}} \widetilde{w}^2 \mathrm{d}x\mathrm{d}y\right)^{\vartheta_0} \left(\int_{\widetilde{B}_{r_3}} \widetilde{w}^2 \mathrm{d}x\mathrm{d}y\right)^{1-\vartheta_0}, \tag{4.53}$$

where

$$\vartheta_0 := \vartheta_0(r_1, r_2, r_3) = \frac{\log\left(\frac{1}{2} + \frac{\widetilde{\lambda} r_3}{2r_2}\right)}{\log\left(\frac{1}{2} + \frac{\widetilde{\lambda} r_3}{2r_2}\right) + C\log\frac{2r_2}{\widetilde{\lambda} r_1}}. \tag{4.54}$$

THEOREM 4.1.6 (*two-sphere one cylinder inequality for Eq.* (4.1)). *Let* $u \in H^{2,1}(B_1 \times (0,1))$ *be a solution to Eq.* (4.1) *and let* (4.2), (4.3) *be satisfied. Then there exist constants* γ_2, $\gamma_2 \in (0,1)$, C, $C > 1$, *depending on* λ *and* Λ *only, such that for every* r_1, r_2, r_3 *that satisfy* $0 < r_1 \le r_2 \le \gamma_2 r_3 \le \gamma_2^2$, *we have*

$$\int_{B_{r_2}} u^2(x,1)\,\mathrm{d}x \le K(r_2, r_3)\, H^{2(1-\beta\vartheta_1)} \left(\int_{B_{r_1}} u^2(x,1)\,\mathrm{d}x\right)^{\beta\vartheta_1}, \tag{4.55}$$

where

$$K(r_2, r_3) = C\frac{r_3}{r_3 - r_2}\left(\frac{r_3}{r_2}\right)^C,$$

$$\vartheta_1 = \frac{\log\left(\frac{1}{2} + \frac{\lambda r_3}{2\widetilde{r}_2}\right)}{\log\left(\frac{1}{2} + \frac{\lambda r_3}{2\widetilde{r}_2}\right) + C\log\frac{2\widetilde{r}_2}{\lambda b r_1}}, \tag{4.56}$$

$$\widetilde{r}_2 = (1 - s_\lambda)\, r_2 + s_\lambda r_3, \qquad s_\lambda = \frac{\lambda}{4-\lambda}, \qquad b = \frac{3\lambda}{4\sqrt{2}e\pi} \tag{4.57}$$

and β *is the same exponent that appears in inequality* (4.49).

PROOF. Let us consider the function w introduced in (4.44). Recall that w satisfies (4.46)–(4.48). Let r_1, r_2, r_3 satisfy

$$0 < r_1 \le r_2 \le \left(4\max\left\{\sqrt{2}, \lambda^{-1}\right\}\right)^{-1} r_3 \le \gamma_*,$$

where $\gamma_* = \min\left\{\gamma_0, \frac{\gamma_1\lambda\delta}{\sqrt{2}}\right\}$, γ_0 and γ_1 have been defined in Lemmas 4.1.4 and 4.1.5, respectively, δ is given by (4.27). Let $\widetilde{r}_2$ and b be defined by ((4.57), we have $0 < br_1 < r_1 \leq r_2 \leq \widetilde{r}_2 \leq \frac{\lambda}{2} r_3 \leq \gamma_1$. By applying Lemma 4.1.5 to the triplet of radii $br_1, \widetilde{r}_2, r_3$ we get

$$\int_{\widetilde{B}_{\widetilde{r}_2}} w^2 \mathrm{d}x\mathrm{d}y \leq C\left(\frac{r_3}{r_2}\right)^C \left(\int_{\widetilde{B}_{br_1}} w^2 \mathrm{d}x\mathrm{d}y\right)^{\vartheta_1} \left(\int_{\widetilde{B}_{r_3}} w^2 \mathrm{d}x\mathrm{d}y\right)^{1-\vartheta_1}, \tag{4.58}$$

where $\vartheta_1 = \vartheta_0\left(br_1, r_2', r_3\right)$, with ϑ_0 defined by (4.56) and C depending on λ and Λ only.

Now, let us recall the following trace inequality. Given $r, \rho, 0 < \rho < r, f \in H^1\left(B_r^+\right)$, we have

$$\int_{B_\rho} f^2(x,0)\,\mathrm{d}x \leq c\left(\frac{r}{r^2-\rho^2}\int_{\widetilde{B}_r^+} f^2(x,y)\,\mathrm{d}x\mathrm{d}y + r\int_{\widetilde{B}_r^+}\left(\partial_y f(x,y)\right)^2 \mathrm{d}x\mathrm{d}y\right), \tag{4.59}$$

where c is an absolute constant.

Set $r_2^* = \frac{1}{2}\left(r_2 + \widetilde{r}_2\right)$. By the Caccioppoli inequality and (4.59) we have

$$\begin{aligned}
\int_{\widetilde{B}_{\widetilde{r}_2}} w^2 \mathrm{d}x\mathrm{d}y &= \frac{1}{2}\int_{\widetilde{B}_{\widetilde{r}_2}} w^2 \mathrm{d}x\mathrm{d}y + \frac{1}{2}\int_{\widetilde{B}_{\widetilde{r}_2}} w^2 \mathrm{d}x\mathrm{d}y \\
&\geq \frac{1}{2}\int_{\widetilde{B}_{r_2^*}} w^2 \mathrm{d}x\mathrm{d}y + C\left(r_2 - \widetilde{r}_2\right)^2 \int_{\widetilde{B}_{r_2^*}} \left(|\nabla_x w|^2 + \left(\partial_y w\right)^2\right)\mathrm{d}x\mathrm{d}y \\
&\geq C'\left(r_3 - r_2\right)\left(\frac{r_2^*}{r_2^{*2} - r_2^2}\int_{\widetilde{B}_{r_2^*}} w^2 \mathrm{d}x\mathrm{d}y + r_2^*\int_{\widetilde{B}_{r_2^*}}\left(\partial_y w\right)^2 \mathrm{d}x\mathrm{d}y\right) \\
&\geq C''\left(r_3 - r_2\right)\int_{B_{r_2}} w^2(x,0)\,\mathrm{d}x = C''\left(r_3 - r_2\right)\int_{B_{r_2}} u^2(x,1)\,\mathrm{d}x,
\end{aligned}$$

where C, C', C'' depend on λ only.

By the inequality just obtained and recalling (4.44) and (4.58), we obtain (4.55). □

COROLLARY 4.1.7 (*Spacelike strong unique continuation for Eq.* (4.1)). *Let* $u \in H^{2,1}\left(B_1 \times (0,1)\right)$ *satisfy Eq.* (4.1).

If for every $k \in \mathbb{N}$ *we have*

$$\int_{B_r} u^2(x,1)\,\mathrm{d}x = O\left(r^{2k}\right), \quad \textit{as } r \to 0, \tag{4.60}$$

then

$$u(.,1) = 0, \quad \textit{in } B_1.$$

PROOF. Let us fix $\rho \in \left(0, \gamma_2^2\right]$, where γ_2 has been introduced in Theorem 4.1.6. By applying Lemma 4.1.5 to the triplet of radii $r_1 = r$, $r_2 = \rho$ and $r_3 = \gamma_2$, by (4.60) and passing to the limit as r tends to 0, we get

$$\int_{B_\rho} u^2 (x, t_0) \, dx \leq C e^{-C_1 k}, \quad \text{for every } k \in \mathbb{N}, \tag{4.61}$$

where C depends on λ, Λ and $\|u\|_{L^2(B_1 \times (0,1))}$, only, and C_1 depends on λ and Λ, only. Passing to the limit as $k \to \infty$, (4.61) yields $u(., 1) = 0$ in B_ρ. By iteration the thesis follows. □

4.2. *Two-sphere one-cylinder inequalities and spacelike strong unique continuation*

In this section we prove the Carleman estimate (Theorem 4.2.3) for parabolic operators proved in [30], [35], see also [33]. In order to prove such an inequality we adapt to the case of variable coefficients the approach used in Lemma 3.0.1 for the heat operator. In Theorems 4.2.6 and 4.2.8 we prove the two-sphere one-cylinder inequality at the interior, and at the boundary, respectively. As a consequence of such theorems we get in 4.2.7 and 4.2.10 the spacelike strong unique continuation property in the interior and at the boundary. In the present section it is convenient to carry out the calculations for the backward parabolic operator. Thus we denote by

$$Pu = \partial_i \left(g^{ij}(x, t) \, \partial_j u\right) + \partial_t u,$$

where $\left\{g^{ij}(x, t)\right\}_{i,j=1}^n$ is a symmetric $n \times n$ matrix whose entries are real functions. When $\xi \in \mathbb{R}^n$ and $(x, t), (y, \tau) \in \mathbb{R}^{n+1}$ we assume that

$$\lambda |\xi|^2 \leq \sum_{i,j=1}^n g^{ij}(x, t) \, \xi_i \xi_j \leq \lambda^{-1} |\xi|^2$$

and

$$\left(\sum_{i,j=1}^n \left(g^{ij}(x, t) - g^{ij}(y, \tau)\right)^2\right)^{1/2} \leq \Lambda \left(|x - y|^2 + |t - \tau|\right)^{1/2}, \tag{4.62}$$

where λ and Λ are positive numbers with $\lambda \in (0, 1]$.

We shall use the notation introduced in Section 2. In particular, with such notation we have

$$Pu = \Delta_g u + \partial_t u.$$

Also, we shall use the following Rellich–Nečas–Pohozaev identity

$$2\left(\beta\cdot\nabla_g v\right)\Delta_g v = 2\,\mathrm{div}\left(\left(\beta\cdot\nabla_g v\right)\nabla_g v\right) - \mathrm{div}\left(\beta\left|\nabla_g v\right|_g^2\right) + \mathrm{div}\left(\beta\right)\left|\nabla_g v\right|_g^2 - 2\partial_i\beta^k g^{ij}\partial_j v\partial_k v + \beta^k\partial_k g^{ij}\partial_i v\partial_j v, \tag{4.63}$$

where $\beta = \left(\beta^1,\dots,\beta^n\right)$ is a smooth vector field.

LEMMA 4.2.1. *Assume that* $\theta : (0,1)\to(0,+\infty)$ *satisfies*

$$0\le\theta\le C_0,\ \left|t\theta'(t)\right|\le C_0\theta(t),\qquad \int_0^1\left(1+\log\frac{1}{t}\right)\frac{\theta(t)}{t}\mathrm{d}t\le C_0, \tag{4.64}$$

for some constant C_0. *Let* $\gamma\ge1$ *and*

$$\sigma(t) = t\exp\left(-\int_0^{\gamma t}\left(1-\exp\left(-\int_0^s\frac{\theta(\eta)}{\eta}\mathrm{d}\eta\right)\right)\frac{\mathrm{d}s}{s}\right). \tag{4.65}$$

Then σ *is the solution to the Cauchy problem*

$$\frac{\mathrm{d}}{\mathrm{d}t}\log\left(\frac{\sigma}{t\sigma'}\right) = \frac{\theta(\gamma t)}{t},\ \sigma(0)=0,\quad \sigma'(0)=1 \tag{4.66}$$

and satisfies the following properties when $0\le\gamma t\le1$

$$t\mathrm{e}^{-C_0}\le\sigma(t)\le t, \tag{4.67}$$

$$\mathrm{e}^{-C_0}\le\sigma'(t)\le1. \tag{4.68}$$

PROOF. The proof is straightforward. □

LEMMA 4.2.2. *For every positive number* μ *there exists a constant* C *such that for all* $y\ge0$ *and* $\varepsilon\in(0,1)$,

$$y^\mu\mathrm{e}^{-y}\le C\left(\varepsilon+\left(\log\frac{1}{\varepsilon}\right)^\mu\mathrm{e}^{-y}\right). \tag{4.69}$$

PROOF. Consider the function $\varphi(y) = y-\varepsilon\mathrm{e}^y$ on $[0,+\infty)$. Since the maximum of φ is $\log\frac{1}{\varepsilon}-1$, $y-\varepsilon\mathrm{e}^y<\log\frac{1}{\varepsilon}$ and we get (4.69) when $\mu=1$. If $\mu>1$, we use the inequality just proved and the convexity of y^μ to get

$$\left(\frac{y}{\mu}\right)^\mu\le\left(\varepsilon^{1/\mu}\mathrm{e}^{y/\mu}+\log\frac{1}{\varepsilon^{1/\mu}}\right)^\mu\le2^{\mu-1}\left(\varepsilon\mathrm{e}^y+\left(\frac{1}{\mu}\log\frac{1}{\varepsilon}\right)^\mu\right),$$

that gives (4.69).

If $0<\mu<1$ we obtain (4.69) by the inequality for $\mu=1$ and the inequality $(a+b)^\mu\le a^\mu+b^\mu$. □

In the next part of this section we shall denote by

$$\theta(t) = t^{1/2}\left(\log\frac{1}{t}\right)^{3/2}, \quad t \in (0,1]. \tag{4.70}$$

It is easy to check that θ satisfies (4.64) of Lemma 4.2.1. We denote by σ the function defined in (4.65) where θ is given by (4.70) and $\gamma = \frac{\alpha}{\delta^2}$ with $\alpha \geq 1$, $\delta \in (0,1)$. In addition, for a given number $a > 0$, we shall denote by $\sigma_a(t) = \sigma(t+a)$.

THEOREM 4.2.3. *Let P be the operator defined in (3.30) and assume that (3.31)* and *(4.62) are satisfied. Assume that $g^{ij}(0,0) = \delta^{ij}$, $i,j = 1,\dots,n$. Then there exist constants C, $C \geq 1$, $\eta_0 \in (0,1)$ and $\delta_1 \in (0,1)$ depending on λ only Λ only, such that for every α, $\alpha \geq 2$, a, $0 < a \leq \frac{\delta^2}{4\alpha}$, $\delta \in (0,\delta_1]$ and $u \in C_0^\infty(\mathbb{R}^n \times [0,+\infty))$, with supp $u \subset B_{\eta_0} \times \left[0, \frac{\delta^2}{2\alpha}\right)$, the following inequality holds true*

$$\begin{aligned}
&\alpha \int_{\mathbb{R}^{n+1}_+} \sigma_a^{-2\alpha-2}(t)\,\theta(\gamma(t+a))\,u^2 \mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X \\
&\quad + \int_{\mathbb{R}^{n+1}_+} (t+a)\,\sigma_a^{-2\alpha-2}(t)\,\theta(\gamma(t+a))\left|\nabla_g u\right|_g^2 \mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X \\
&\leq C \int_{\mathbb{R}^{n+1}_+} (t+a)^2\,\sigma_a^{-2\alpha-2}(t)\,|Pu|^2\,\mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X \\
&\quad + C^\alpha \alpha^\alpha \int_{\mathbb{R}^{n+1}_+} \left(u^2 + (t+a)\left|\nabla_g u\right|_g^2\right)\mathrm{d}X \\
&\quad + C\alpha\sigma^{-2\alpha-1}(a) \int_{\mathbb{R}^n} u^2(x,0)\,\mathrm{e}^{-\frac{|x|^2}{4a}}\,\mathrm{d}x \\
&\quad - \frac{\sigma^{-2\alpha}(a)}{C} \int_{\mathbb{R}^n} \left|\left(\nabla_g u\right)(x,0)\right|_{g(.,0)}^2 \mathrm{e}^{-\frac{|x|^2}{4a}}\,\mathrm{d}x.
\end{aligned} \tag{4.71}$$

PROOF. We divide the proof into two steps. In the first step we prove that if the matrix g does not depend on t then there exists $\delta_0 \in (0,1)$ such that, for every $\delta \in (0,\delta_0]$, inequality (4.71) holds true. In the second step we prove inequality (4.71), for a suitable δ_1, in the general case. For the sake of brevity, in what follows we prove the inequality when the matrix $g(x,t)$ and the function $u(x,t)$ are replaced by $\widetilde{g}(x,t) := g(x,t-a)$ and $\widetilde{u}(x,t) := u(x,t-a)$, respectively, so that $\widetilde{g}^{ij}(0,a) = \delta^{ij}$ and $\widetilde{u}$ is a function belonging to $C_0^\infty(\mathbb{R}^n \times [a,+\infty))$. Also, the sign "$\sim$" over g and u will be dropped. Finally, we shall denote by $\int (.)\,\mathrm{d}X$ the integral $\int_{\mathbb{R}^n \times [a,+\infty)} (.)\,\mathrm{d}X$.

STEP 1. Set $g(x) = g(x,a)$ and denote

$$P_1 u = \partial_i\left(g^{ij}(x)\,\partial_j u\right) + \partial_t u,$$

$$\phi(x,t) = -\frac{|x|^2}{8t} - (\alpha+1)\log\sigma(t),$$
$$v = e^{\phi}u,$$
$$Lv = e^{\phi}P_1\left(e^{-\phi}v\right).$$

Denoting by S and A the symmetric and skew-symmetric part of the operator tL, respectively, we have

$$Sv = t\left(\Delta_g v + \left(\left|\nabla_g\phi\right|_g^2 - \partial_t\phi\right)v\right) - \frac{1}{2}v,$$
$$Av = \frac{1}{2}\left(\partial_t(tv) + t\partial_t v\right) - t\left(v\Delta_g\phi + 2\nabla_g v\cdot\nabla_g\phi\right)$$

and

$$tLv = Sv + Av.$$

Furthermore, noticing that

$$\Delta_g\phi = -\frac{n}{4t} + \psi_0(x,t),$$

where

$$\psi_0(x,t) = \frac{1}{4t}\left(n - tr\left(g^{-1}\right) - x_j\partial_i g^{ij}\right),$$

we write

$$tLv = Sv + A_0v - t\psi_0 v, \tag{4.72}$$

where

$$A_0v = \frac{1}{2}\left(\partial_t(sv) + t\partial_t v\right) - t\left(-\frac{n}{4t}v + 2\nabla_g v\cdot\nabla_g\phi\right).$$

Noticing that $|\psi_0| \leq \frac{C\Lambda|x|}{t}$, where C is an absolute constant, we have by (4.72)

$$\begin{aligned}2\int t^2|Lv|^2\,dX &\geq \int|Sv + A_0v|^2\,dX - 4\int t^2\psi_0^2v^2dX\\ &\geq \int|Sv|^2\,dX + \int|A_0v|^2\,dX + 2\int SvA_0vdX\\ &\quad - 4C^2\Lambda^2\int|x|^2v^2dX.\end{aligned} \tag{4.73}$$

Denote by

$$I := 2\int SvA_0vdX.$$

We have

$$
\begin{aligned}
I = & -2\int t^2\left(\Delta_g v+\left(\left|\nabla_g\phi\right|_g^2-\partial_t\phi\right)v\right)\left(-\frac{n}{4t}v+2\nabla_g v\cdot\nabla_g\phi\right)\mathrm{d}X\\
&+\int t\left(\Delta_g v+\left(\left|\nabla_g\phi\right|_g^2-\partial_t\phi\right)v\right)\left(\partial_t\left(tv\right)+t\partial_t v\right)\mathrm{d}X\\
&+\int\left\{-\frac{1}{2}v\left(\partial_t\left(sv\right)+t\partial_t v\right)+tv\left(-\frac{n}{4t}v+2\nabla_g v\cdot\nabla_g\phi\right)\right\}\mathrm{d}X\\
:= & I_1+I_2+I_3.
\end{aligned}
\tag{4.74}
$$

Evaluation of I_1.

We have

$$
\begin{aligned}
I_1 = & -4\int t^2\Delta_g v\nabla_g v\cdot\nabla_g\phi\mathrm{d}X-4\int t^2\left(\left|\nabla_g\phi\right|_g^2-\partial_t\phi\right)v\nabla_g v\cdot\nabla_g\phi\mathrm{d}X\\
&+\frac{n}{2}\int\left(tv\Delta_g v+\left(\left|\nabla_g\phi\right|_g^2-\partial_t\phi\right)v^2\right)\mathrm{d}X := I_{11}+I_{12}+I_{13}.
\end{aligned}
\tag{4.75}
$$

By using identity (4.63) with the vector field $\beta=\nabla_g\phi$ we have

$$
\begin{aligned}
I_{11} = & \int\left\{4t^2\partial_i\left(g^{kj}\partial_j\phi\right)g^{ih}\partial_h v\partial_k v-2t^2\Delta_g\phi\left|\nabla_g v\right|_g^2\right.\\
&\left.-2t^2\left(g^{kj}\partial_j\phi\right)\left(\partial_k g^{hl}\partial_h v\partial_l v\right)\right\}\mathrm{d}X.
\end{aligned}
\tag{4.76}
$$

Concerning the term I_{12}, by integration by parts we have

$$
\begin{aligned}
I_{12} &= -2\int t^2\left(\left|\nabla_g\phi\right|_g^2-\partial_t\phi\right)\nabla_g\phi\cdot\nabla_g\left(v^2\right)\mathrm{d}X\\
&= 2\int t^2\mathrm{div}\left(\left(\left|\nabla_g\phi\right|_g^2-\partial_t\phi\right)\nabla_g\phi\right)v^2\mathrm{d}X\\
&= 2\int t^2v^2\left\{\Delta_g\phi\left(\left|\nabla_g\phi\right|_g^2-\partial_t\phi\right)+g^{ij}\partial_i\phi\partial_j\left(\left|\nabla_g\phi\right|_g^2-\partial_t\phi\right)\right\}\mathrm{d}X.
\end{aligned}
$$

Now we have

$$
\begin{aligned}
& g^{ij}\partial_i\phi\partial_j\left(\left|\nabla_g\phi\right|_g^2-\partial_t\phi\right)\\
&\quad=\left(2\partial^2_{jk}\phi g^{kh}\partial_h\phi+\partial_j g^{hk}\partial_k\phi\partial_h\phi\right)g^{ij}\partial_i\phi-\frac{1}{2}\partial_t\left(\left|\nabla_g\phi\right|_g^2\right),
\end{aligned}
$$

hence

$$
\begin{aligned}
I_{12} = & 2\int t^2v^2\Delta_g\phi\left(\left|\nabla_g\phi\right|_g^2-\partial_t\phi\right)\mathrm{d}X\\
&+4\int t^2v^2\left(\partial^2_{jk}\phi g^{kh}\partial_h\phi+\frac{1}{2}\partial_j g^{hk}\partial_k\phi\partial_h\phi\right)g^{ij}\partial_i\phi\mathrm{d}X
\end{aligned}
$$

$$-\int t^2 v^2 \partial_t \left(\left|\nabla_g \phi\right|_g^2\right) \mathrm{d}X. \tag{4.77}$$

Concerning I_{13}, by the identity $v\Delta_g v = \operatorname{div}\left(v\nabla_g v\right) - \left|\nabla_g v\right|_g^2$ and by integration by parts we obtain

$$\begin{aligned} I_{13} &= \frac{n}{2}\int t\left(\operatorname{div}\left(v\nabla_g v\right) - \left|\nabla_g v\right|_g^2 + \left(\left|\nabla_g \phi\right|_g^2 - \partial_t \phi\right) v^2\right) \mathrm{d}X \\ &= \frac{n}{2}\int \left(-t\left|\nabla_g v\right|_g^2 + \left(\left|\nabla_g \phi\right|_g^2 - \partial_t \phi\right) v^2\right) \mathrm{d}X, \end{aligned}$$

so

$$I_{13} = \frac{n}{2}\int \left(-t\left|\nabla_g v\right|_g^2 + \left(\left|\nabla_g \phi\right|_g^2 - \partial_t \phi\right) v^2\right) \mathrm{d}X. \tag{4.78}$$

Now, let us introduce the notation

$$\begin{aligned} &\psi_1 = \left(\frac{n}{2}t + 2t^2\Delta_g \phi\right)\left(\left|\nabla_g \phi\right|_g^2 - \partial_t \phi\right) + 2t^2 \partial_j g^{hk} \partial_k \phi \partial_h \phi g^{ij} \partial_i \phi, \\ &Q\left(\nabla_g v\right) = -2t^2 \psi_0 \left|\nabla_g v\right|_g^2 - 2t^2 \left(g^{kj}\partial_j \phi\right)\left(\partial_k g^{hl} \partial_h v \partial_l v\right). \end{aligned}$$

By (4.75)–(4.78) we have

$$\begin{aligned} I_1 = {}& 4\int t^2 \left\{\partial_i\left(g^{kj}\partial_j \phi\right) g^{ih} \partial_h v \partial_k v + \partial_{jk}^2 \phi g^{kh} \partial_h \phi g^{ij} \partial_i \phi v^2\right\} \mathrm{d}X \\ & - \int t^2 v^2 \partial_t \left(\left|\nabla_g \phi\right|_g^2\right) \mathrm{d}X + \int \psi_1 v^2 \mathrm{d}X + \int Q\left(\nabla_g v\right) \mathrm{d}X. \end{aligned} \tag{4.79}$$

Evaluation of I_2.

We have

$$\begin{aligned} I_2 = {}& \int t v \Delta_g v \mathrm{d}X + \int 2t^2 \left(\left|\nabla_g \phi\right|_g^2 - \partial_t \phi\right) v \partial_t v \mathrm{d}X \\ & + \int 2t^2 \Delta_g v \partial_t v \mathrm{d}X + \int t\left(\left|\nabla_g \phi\right|_g^2 - \partial_t \phi\right) v^2 \mathrm{d}X. \end{aligned} \tag{4.80}$$

Now, we use the identities

$$v\Delta_g v = \operatorname{div}\left(v\nabla_g v\right) - \left|\nabla_g v\right|_g^2, \qquad 2v\partial_t v = \partial_t\left(v^2\right)$$

and

$$2\partial_t v \Delta_g v = 2\operatorname{div}\left(\partial_t v \nabla_g v\right) - \partial_t\left(\left|\nabla_g v\right|_g^2\right),$$

in the first, second and third integrals, respectively, on the right-hand side of (4.80), then we integrate by parts and obtain

$$
\begin{aligned}
I_2 = & \int t \left|\nabla_g v\right|_g^2 \mathrm{d}X - \int t^2 \partial_t \left(\left|\nabla_g \phi\right|_g^2 - \partial_t \phi\right) v^2 \mathrm{d}X \\
& - \int t \left(\left|\nabla_g \phi\right|_g^2 - \partial_t \phi\right) v^2 \mathrm{d}X \\
& - a^2 \int_{\mathbb{R}^n} \left(\left|\nabla_g \phi\left(x, a\right)\right|_g^2 - \partial_t \phi\left(x, a\right)\right) v^2\left(x, a\right) \mathrm{d}x \\
& + a^2 \int_{\mathbb{R}^n} \left|\nabla_g v\left(x, a\right)\right|_g^2 \mathrm{d}x. \qquad (4.81)
\end{aligned}
$$

Evaluation of I_3

We have

$$
I_3 = -\int \left(\frac{1}{2} + \frac{n}{4}\right) v^2 \mathrm{d}X - \frac{1}{2} \int t \partial_t \left(v^2\right) \mathrm{d}X + \int 2 t v \nabla_g v \cdot \nabla_g \phi \mathrm{d}X. \qquad (4.82)
$$

We use the identity

$$
2 v \nabla_g v \cdot \nabla_g \phi = \operatorname{div}\left(v^2 \nabla_g \phi\right) - v^2 \Delta_g \phi,
$$

in the third integral on the right-hand side of (4.82), then we integrate by parts and obtain

$$
I_3 = -\int t \psi_0 v^2 \mathrm{d}X + \frac{a}{2} \int_{\mathbb{R}^n} v^2\left(x, a\right) \mathrm{d}x. \qquad (4.83)
$$

Now, denote by

$$
\begin{aligned}
J_1 = & \int 4 t^2 \partial_{jk}^2 \phi g^{kh} \partial_h \phi g^{ij} \partial_i \phi v^2 \mathrm{d}X \\
& - \int \left\{t^2 \partial_t \left(2\left|\nabla_g \phi\right|_g^2 - \partial_t \phi\right) v^2 + t\left(\left|\nabla_g \phi\right|_g^2 - \partial_t \phi\right) v^2\right\} \mathrm{d}X, \qquad (4.84)
\end{aligned}
$$

$$
J_2 = \int \left\{4 t^2 \partial_i \left(g^{kj} \partial_j \phi\right) g^{ih} \partial_h v \partial_k v + t\left|\nabla_g v\right|_g^2 + Q\left(\nabla_g v\right)\right\} \mathrm{d}X, \qquad (4.85)
$$

$$
J_3 = \int \left(\psi_1 - t \psi_0\right) v^2 \mathrm{d}X, \qquad (4.86)
$$

$$
\begin{aligned}
\mathcal{B}_1 = & -a^2 \int_{\mathbb{R}^n} \left(\left|\nabla_g \phi\left(x, a\right)\right|_g^2 - \partial_t \phi\left(x, a\right)\right) v^2\left(x, a\right) \mathrm{d}x \\
& + a^2 \int_{\mathbb{R}^n} \left|\nabla_g v\left(x, a\right)\right|_g^2 \mathrm{d}x + \frac{a}{2} \int_{\mathbb{R}^n} v^2\left(x, a\right) \mathrm{d}x. \qquad (4.87)
\end{aligned}
$$

By (4.74), (4.79), (4.81) and (4.82) we have

$$
I = J_1 + J_2 + J_3 + \mathcal{B}_1. \qquad (4.88)
$$

Let us now estimate from below the terms J_1, J_2, J_3, at the right-hand side of (4.88). We have easily

$$\partial_i\phi = -\frac{x_i}{4t}, \qquad \partial^2_{ij}\phi = -\frac{\delta_{ij}}{4t}, \qquad \left|\nabla_g\phi\right|^2_g = g^{ij}(x)\frac{x_ix_j}{16t^2}, \tag{4.89}$$

$$\Delta_g\phi = -\frac{tr\left(g^{-1}(x)\right) + x_j\partial_i g^{ij}(x)}{4t}, \qquad \partial_t\phi = \frac{|x|^2}{8t^2} - (\alpha+1)\frac{\sigma'(t)}{\sigma(t)}, \tag{4.90}$$

$$\partial^2_t\phi = -\frac{|x|^2}{4t^3} - (\alpha+1)\left(\frac{\sigma'(t)}{\sigma(t)}\right)', \qquad \partial_t\left(\left|\nabla_g\phi\right|^2_g\right) = -g^{ij}(x)\frac{x_ix_j}{8t^3}. \tag{4.91}$$

Observing that if $g^{ij}(x) = \delta^{ij}$, $i, j = 1, 2, \ldots, n$, $x \in \mathbb{R}^n$, then the term under the integral sign at the right-hand side of (4.84) is equal to $(\alpha+1)\,v^2\frac{\sigma'}{\sigma}\frac{\mathrm{d}}{\mathrm{d}t}\left(\log\frac{\sigma}{t\sigma'}\right)$, it is easy to check that

$$J_1 \ge (\alpha+1)\int t^2v^2\frac{\sigma'}{\sigma}\frac{\mathrm{d}}{\mathrm{d}t}\left(\log\frac{\sigma}{t\sigma'}\right)\mathrm{d}X - C\Lambda\int\frac{|x|^3}{t}v^2\mathrm{d}X, \tag{4.92}$$

where C is an absolute constant.
Likewise, observing that if $g^{ij}(x) = \delta^{ij}$, $i, j = 1, 2, \ldots, n$, $\in \mathbb{R}^n$, then the term under the integral sign at the right-hand side of (4.85) vanishes, so that the following inequality holds true

$$J_2 \ge -C\Lambda\int |x|\,t\left|\nabla_g v\right|^2_g\mathrm{d}X, \tag{4.93}$$

where C is an absolute constant.
Now, let us consider J_3. Let $\delta \in (0, 1)$ be a number that we shall choose later on, let σ be the function (4.65) where $\gamma = \frac{\alpha}{\delta^2}$, $\alpha \ge 1$ and θ is given by (4.70). We have

$$|\psi_1 - t\psi_0| \le |\psi_1| + t|\psi_0| \le C\left(|x| + \frac{|x|^3}{t} + (\alpha+1)\,t\,|x|\frac{\sigma'}{\sigma}\right),$$

where C depends on λ, Λ only. By (4.67) and (4.68) we have

$$J_3 \ge -C\int\left(\frac{|x|^3}{t} + (\alpha+1)|x|\right)v^2\mathrm{d}X, \tag{4.94}$$

for every $v \in C^\infty_0(\mathbb{R}^n \times [a, +\infty))$ such that $\operatorname{supp} v \subset B_1 \times \left[a, \frac{1}{\gamma}\right)$, where C depends on λ and Λ only.

Now by Lemma 4.2.1 and by (4.92)–(4.94) we get

$$I \ge \frac{(\alpha+1)}{\mathrm{e}^{C_0}}\int\theta(\gamma t)\,v^2\mathrm{d}X - C\int\left((\alpha+1)|x| + \frac{|x|^3}{t}\right)v^2\mathrm{d}X$$
$$- C\int|x|\,t\left|\nabla_g v\right|^2_g\mathrm{d}X + \mathcal{B}_1, \tag{4.95}$$

for every $v \in C_0^\infty(\mathbb{R}^n \times [a, +\infty))$ such that $\operatorname{supp} v \subset B_1 \times \left[a, \frac{1}{\gamma}\right)$, $\alpha \geq 1$, where C_0 is defined in Lemma 4.2.1 and C depends on λ and Λ only.

Let $\varepsilon_0 \in (0, 1]$ be a number that we shall choose later. We get

$$\begin{aligned}
\int |Sv|^2 \, dX \geq \varepsilon_0 \int |Sv|^2 \, dX &= \varepsilon_0 \int |(Sv + \theta(\gamma t) v) - \theta(\gamma t) v|^2 \, dX \\
&\geq -2\varepsilon_0 \int t\theta(\gamma t) v \Delta_g v dX \\
&\quad - 2\varepsilon_0 \int \left(t\left(|\nabla_g \phi|_g^2 - \partial_t \phi \right) - \frac{1}{2} + \theta(\gamma t) \right) \theta(\gamma t) v^2 dX \\
&= 2\varepsilon_0 \int t\theta(\gamma t) \left(|\nabla_g v|_g^2 + |\nabla_g \phi|_g^2 v^2 \right) dX \\
&\quad - 2\varepsilon_0 \int \left(t\left(2|\nabla_g \phi|_g^2 - \partial_t \phi \right) - \frac{1}{2} + \theta(\gamma t) \right) \theta(\gamma t) v^2 dX.
\end{aligned} \tag{4.96}$$

Now by (4.67), (4.68) and (4.89)–(4.91) we obtain

$$\begin{aligned}
&\left| \left(t\left(2|\nabla_g \phi|_g^2 - \partial_t \phi \right) - \frac{1}{2} + \theta(\gamma t) \right) \theta(\gamma t) \right| \\
&\quad \leq C \left(\frac{|x|^3}{t} + (\alpha + 1)\theta(\gamma t) \right), \quad \text{for every } 0 < \gamma t \leq \frac{1}{2}, \alpha \geq 1,
\end{aligned} \tag{4.97}$$

where C depends on λ and Λ only.

By (4.96) and (4.97) we have

$$\begin{aligned}
\int |Sv|^2 \, dX &\geq 2\varepsilon_0 \int t\theta(\gamma t) \left(|\nabla_g v|_g^2 + |\nabla_g \phi|_g^2 v^2 \right) dX \\
&\quad - C_1 \varepsilon_0 \int \left(\frac{|x|^3}{t} + (\alpha + 1)\theta(\gamma t) \right) v^2 dX,
\end{aligned} \tag{4.98}$$

for every $v \in C_0^\infty(\mathbb{R}^n \times [a, +\infty))$ such that *supp* $v \subset B_1 \times \left[a, \frac{1}{2\gamma}\right)$, $\alpha \geq 1$, where C_1 depends on λ and Λ only.

By the inequality just obtained and (4.95) we get

$$\begin{aligned}
&2 \int Sv A_0 v dX + \int |Sv|^2 \, dX \geq (\alpha + 1) \left(\frac{1}{e^{C_0}} - C_1 \varepsilon_0 \right) \int \theta(\gamma t) v^2 dX \\
&\quad + 2\varepsilon_0 \int t\theta(\gamma t) \left(|\nabla_g v|_g^2 + |\nabla_g \phi|_g^2 v^2 \right) dX \\
&\quad - C \int \left((\alpha + 1)|x| + \frac{|x|^3}{t} \right) v^2 dX \\
&\quad - C \int |x| t \left(|\nabla_g v|_g^2 + |\nabla_g \phi|_g^2 v^2 \right) + \mathcal{B}_1,
\end{aligned} \tag{4.99}$$

for every $v \in C_0^\infty(\mathbb{R}^n \times [a, +\infty))$ such that $\operatorname{supp} v \subset B_1 \times \left[a, \frac{1}{2\gamma}\right)$, $\alpha \geq 1$, where C depends on λ and Λ only and C_1 is the same constant that appears in (4.98).

Now we choose $\varepsilon_0 = \frac{1}{2C_1 e^{C_0}}$ at the right-hand side of (4.99), then, recalling (4.73) and coming back to the function u, we have

$$\int t^2 |P_1 u|^2 e^{2\phi} dX \geq \frac{(\alpha+1)}{C} \int \theta(\gamma t) u^2 e^{2\phi} dX + \frac{1}{C} \int t\theta(\gamma t) \left|\nabla_g u\right|_g^2 e^{2\phi} dX + \mathcal{B}_1 - C \int |x| t \left|\nabla_g u\right|_g^2 e^{2\phi} dX - C \int \left((\alpha+1)|x| + \frac{|x|^3}{t}\right) u^2 e^{2\phi} dX, \tag{4.100}$$

for every $u \in C_0^\infty(\mathbb{R}^n \times [a, +\infty))$ such that $\operatorname{supp} u \subset B_1 \times \left[a, \frac{1}{2\gamma}\right)$, $\alpha \geq 1$, where $C > 1$, depends on λ and Λ only.

Now, we use Lemma 4.2.2 to prove the following estimates

$$\int |x| t \left|\nabla_g u\right|_g^2 e^{2\phi} dX \leq C \left(e^{C_0}\gamma\right)^{2\alpha+\frac{3}{2}} \int t \left|\nabla_g u\right|_g^2 dX + C\delta \int \theta(\gamma t) t \left|\nabla_g u\right|_g^2 e^{2\phi} dX, \tag{4.101}$$

$$\int \left((1+\alpha)|x| + \frac{|x|^3}{t}\right) u^2 e^{2\phi} dX \leq C \left(e^{C_0}\gamma\right)^{2\alpha+\frac{5}{2}} \int u^2 dX + C\alpha\delta \int \theta(\gamma t) u^2 e^{2\phi} dX, \tag{4.102}$$

where C is an absolute constant.

By applying Lemma 4.2.2 with $\mu = \frac{1}{2}$, $y = \frac{|x|^2}{4t}$, $\varepsilon = (\gamma t)^{\frac{3}{2}+2\alpha}$ and by using (4.67) we have

$$\begin{aligned} |x| e^{2\phi} &= (4s)^{1/2} \left(\frac{|x|^2}{4t}\right)^{1/2} e^{-\frac{|x|^2}{4t}} \sigma^{-2(\alpha+1)} \\ &\leq C \left((\gamma t)^{\alpha+\frac{3}{2}} t^{1/2} \sigma^{-2(\alpha+1)} + t^{1/2} \left(\left(\frac{3}{2} + 2\alpha\right) \log \frac{1}{\gamma t} \right)^{1/2} \right. \\ &\quad \left. \times \sigma^{-2(\alpha+1)} e^{-\frac{|x|^2}{4t}} \right) \\ &\leq C' \left(\left(e^{C_0}\gamma\right)^{2\alpha+\frac{5}{2}} + \delta\theta(\gamma t) e^{2\phi} \right), \end{aligned} \tag{4.103}$$

where $C \geq 1$, hence (4.101) follows.

Likewise, by applying Lemma 4.2.2 with $\mu = \frac{3}{2}$, $y = \frac{|x|^2}{4t}$, $\varepsilon = (\gamma t)^{\frac{3}{2}+2\alpha}$ we have

$$\frac{|x|^3}{t} e^{2\phi} = 4\,(4t)^{1/2} \left(\frac{|x|^2}{4t}\right)^{3/2} e^{-\frac{|x|^2}{4t}} \sigma^{-2(\alpha+1)}$$
$$\leq C\left(\left(e^{C_0}\gamma\right)^{2\alpha+2} + \alpha\delta\theta\,(\gamma t)\, e^{2\phi}\right). \tag{4.104}$$

By (4.103) and (4.104) we get (4.102).

Finally, by (4.100), (4.102) and (4.101) we have that there exists $\delta_0 \in (0, 1]$ such that for every $\delta \in (0, \delta_0]$, $\alpha \geq 1$, $u \in C_0^\infty(\mathbb{R}^n \times [a, +\infty))$ such that $\operatorname{supp} u \subset B_1 \times \left[a, \frac{1}{2\gamma}\right)$, the following estimate holds true

$$\int t^2 |P_1 u|^2 e^{2\phi} dX + C\left(e^{C_0}\gamma\right)^{2\alpha+\frac{5}{2}} \int \left(u^2 + t\left|\nabla_g u\right|_g^2\right) e^{2\phi} dX$$
$$\geq \frac{\alpha+1}{C} \int \theta\,(\gamma t)\, u^2 e^{2\phi} dX + \frac{1}{C} \int \theta\,(\gamma t)\, t\left|\nabla_g u\right|_g^2 e^{2\phi} dX + \mathcal{B}_1, \tag{4.105}$$

where $C \geq 1$ depends on λ and Λ only.

STEP 2. It is simple to check the following identity, for $k \in \{1, \ldots, n\}$

$$P_1\left((\partial_k u)^2\right) = 2\partial_k\,(\partial_k u P_0 u) - 2\partial_i\left(\partial_k u\left(\partial_k g^{ij}(x, a)\right)\partial_j u\right)$$
$$- 2\partial_{kk}^2 u P_0 u + 2\left(\partial_k g^{ij}(x, a)\right)\partial_j u \partial_{ik}^2 u + 2g^{ij}(x, a)\,\partial_{ik}^2 u \partial_{jk}^2 u,$$

(recall that $(P_1 u)(x, t) = \partial_j\left(g^{ij}(x, a)\,\partial_i u(x, t)\right) + \partial_t u(x, t)$). Now we multiply both sides of the above inequality by $\sigma^{1-2\alpha} e^{-\frac{|x|^2}{4t}}$ and we integrate by parts. We obtain, for any $k \in \{1, \ldots, n\}$,

$$2\int\left(g^{ij}(x, a)\,\partial_{ik}^2 u \partial_{jk}^2 u\right)\sigma^{1-2\alpha} e^{-\frac{|x|^2}{4t}} dX$$
$$\leq \int (\partial_k u)^2 \left|P_0^*\left(\sigma^{1-2\alpha} e^{-\frac{|x|^2}{4t}}\right)\right| dX$$
$$+ 2\int |\partial_k u P_0 u| \left|\partial_k\left(\sigma^{1-2\alpha} e^{-\frac{|x|^2}{4t}}\right)\right| dX$$
$$+ 2\int \left|\partial_k g^{ij}(x, a)\,\partial_j u \partial_k u\right| \left|\partial_i\left(\sigma^{1-2\alpha} e^{-\frac{|x|^2}{4t}}\right)\right| dX$$
$$+ 2\int \left|\partial_{kk}^2 u P_0 u\right| \sigma^{1-2\alpha} e^{-\frac{|x|^2}{4t}} dX$$
$$+ 2\int \left|\partial_k g^{ij}(x, a)\,\partial_j u \partial_{ik}^2 u\right| \sigma^{1-2\alpha} e^{-\frac{|x|^2}{4t}} dX$$
$$- \int_{\mathbb{R}^n} (\partial_k u(x, a))^2 \sigma^{1-2\alpha}(a)\, e^{-\frac{|x|^2}{4a}} dx, \tag{4.106}$$

where $P_1^* = \partial_j \left(g^{ij}(x,a)\,\partial_i u(x,s)\right) - \partial_t u(x,t)$.
We have

$$\left|P_1^*\left(\sigma^{1-2\alpha}\mathrm{e}^{-\frac{|x|^2}{4t}}\right)\right| = \left|P_1^*\left(\left(\sigma^{1-2\alpha}t^{n/2}\right)\left(t^{-n/2}\mathrm{e}^{-\frac{|x|^2}{4t}}\right)\right)\right|$$
$$\leq C\left(\left(\frac{|x|}{t}+\frac{|x|^3}{t^2}\right)\sigma^{1-2\alpha}+\alpha\sigma^{-2\alpha}\right)\mathrm{e}^{-\frac{|x|^2}{4t}}$$

and

$$\left|\nabla\left(\sigma^{1-2\alpha}\mathrm{e}^{-\frac{|x|^2}{4t}}\right)\right| \leq C'\frac{|x|}{t}\sigma^{1-2\alpha}\mathrm{e}^{-\frac{|x|^2}{4t}},$$

where C depends on λ and Λ only and C' is an absolute constant.
By the above inequalities and (4.106) we get

$$\begin{aligned}\int\left|D^2u\right|^2\sigma^{1-2\alpha}\mathrm{e}^{-\frac{|x|^2}{4t}}\mathrm{d}X \leq{}& C\alpha\int\sigma^{-2\alpha}\left|\nabla_g u\right|_g^2\mathrm{e}^{-\frac{|x|^2}{4t}}\mathrm{d}X\\ &+C\int\left(\frac{|x|}{t}+\frac{|x|^3}{t^2}\right)\sigma^{1-2\alpha}\left|\nabla_g u\right|_g^2\mathrm{e}^{-\frac{|x|^2}{4t}}\mathrm{d}X\\ &+C\int\frac{|x|}{t}\sigma^{1-2\alpha}\left|\nabla_g u\right|_g|P_1u|\,\mathrm{e}^{-\frac{|x|^2}{4t}}\mathrm{d}X\\ &+C\int\left|D^2u\right||P_0u|\,\sigma^{1-2\alpha}\mathrm{e}^{-\frac{|x|^2}{4t}}\mathrm{d}X\\ &+C\int\left|D^2u\right|\left|\nabla_g u\right|_g\sigma^{1-2\alpha}\mathrm{e}^{-\frac{|x|^2}{4t}}\mathrm{d}X\\ &-\sum_{k=1}^n\int_{\mathbb{R}^n}\sigma^{1-2\alpha}(a)\left(\partial_k u(x,a)\right)^2\mathrm{e}^{-\frac{|x|^2}{4a}}\mathrm{d}x\\ ={}& I^{(1)}+I^{(2)}+I^{(3)}+I^{(4)}+I^{(5)}-\mathcal{B}_2,\end{aligned} \tag{4.107}$$

(here $\left|D^2u\right|^2=\sum_{i,j=1}^n\left(\partial_{ij}^2u\right)^2$) where C depends on λ and Λ only.

By the trivial inequality $\frac{\theta(t)}{t}\geq\frac{1}{C}$, for every $t\in\left(0,\frac{1}{2}\right]$, $C\geq 1$, we have $C\delta^2\frac{\theta(\gamma t)}{t}\geq\alpha$, whenever $0<\gamma t\leq\frac{1}{2}$. By the last inequality and (4.67) we have

$$I^{(1)}\leq C\delta^2\int t\theta(\gamma t)\,\sigma^{-2-2\alpha}\left|\nabla_g u\right|_g^2\mathrm{e}^{-\frac{|x|^2}{4t}}\mathrm{d}X, \tag{4.108}$$

for every $\alpha\geq 1$, $u\in C_0^\infty(\mathbb{R}^n\times[a,+\infty))$ such that $\operatorname{supp}u\subset B_1\times\left[a,\frac{1}{2\gamma}\right)$, where C depends on λ and Λ only.
To estimate $I^{(2)}$ we observe that (4.65) gives $\sigma(t)\leq\frac{C}{\gamma}$, whenever $0<\gamma t\leq 1$. By this inequality, (4.67) and Lemma 4.2.2 we have

$$I^{(2)} \leq \frac{C}{\gamma} \int \left(|x| + \frac{|x|^3}{t} \right) \sigma^{-1-2\alpha} \left| \nabla_g u \right|_g^2 \mathrm{e}^{-\frac{|x|^2}{4t}} \mathrm{d}X$$
$$\leq \frac{C}{\gamma} \left(\mathrm{e}^{C_0} \gamma \right)^{2\alpha + \frac{3}{2}} \int t \left| \nabla_g u \right|_g^2 \mathrm{d}X$$
$$+ C\delta^3 \int t\theta\left(\gamma t\right) \sigma^{-2-2\alpha} \left| \nabla_g u \right|_g^2 \mathrm{e}^{-\frac{|x|^2}{4t}} \mathrm{d}X, \tag{4.109}$$

for every $\alpha \geq 1$, $u \in C_0^\infty\left(\mathbb{R}^n \times [a, +\infty)\right)$ such that $\operatorname{supp} u \subset B_1 \times \left[a, \frac{1}{2\gamma}\right)$, where C depends on λ and Λ only.

In order to estimate $I^{(3)}$, $I^{(4)}$ and $I^{(5)}$ we use the inequalities $2ab \leq a^2 + b^2$, $\sigma(t) \leq \frac{C}{\gamma}$, $t \leq \theta(\gamma t)$, when $0 < \gamma t \leq \frac{1}{2}$, and Lemma 4.2.2 and we obtain, for every $\alpha \geq 1$, $u \in C_0^\infty\left(\mathbb{R}^n \times [a, +\infty)\right)$ such that $\operatorname{supp} u \subset B_1 \times \left[a, \frac{1}{2\gamma}\right)$

$$I^{(3)} \leq C \left(\mathrm{e}^{C_0} \gamma \right)^{2\alpha+1} \int t \left| \nabla_g u \right|_g^2 \mathrm{d}X$$
$$+ C\delta^2 \int t\theta\left(\gamma t\right) \sigma^{-2-2\alpha} \left| \nabla_g u \right|_g^2 \mathrm{e}^{-\frac{|x|^2}{4t}} \mathrm{d}X$$
$$+ C\delta^2 \int t^2 \sigma^{-2-2\alpha} \left| P_1 u \right|^2 \mathrm{e}^{-\frac{|x|^2}{4t}} \mathrm{d}X, \tag{4.110}$$

$$I^{(4)} \leq \frac{1}{4} \int \left| D^2 u \right|^2 \sigma^{1-2\alpha} \mathrm{e}^{-\frac{|x|^2}{4t}} \mathrm{d}X$$
$$+ C\delta^2 \int t^2 \sigma^{-2-2\alpha} \left| P_1 u \right|^2 \mathrm{e}^{-\frac{|x|^2}{4t}} \mathrm{d}X, \tag{4.111}$$

$$I^{(5)} \leq \frac{1}{4} \int \left| D^2 u \right|^2 \sigma^{1-2\alpha} \mathrm{e}^{-\frac{|x|^2}{4t}} \mathrm{d}X$$
$$+ C\delta^2 \int t\theta\left(\gamma t\right) \sigma^{-2-2\alpha} \left| \nabla_g u \right|_g^2 \mathrm{e}^{-\frac{|x|^2}{4t}} \mathrm{d}X, \tag{4.112}$$

where C depends on λ and Λ only.

By (4.107)–(4.112) we have

$$\frac{1}{2\delta^2} \int \left| D^2 u \right|^2 \sigma^{1-2\alpha} \mathrm{e}^{-\frac{|x|^2}{4t}} \mathrm{d}X \leq \frac{C}{\delta^2} \left(\mathrm{e}^{C_0} \gamma \right)^{2\alpha + \frac{3}{2}} \int t \left| \nabla_g u \right|_g^2 \mathrm{d}X$$
$$+ C \int t\theta\left(\gamma t\right) \sigma^{-2-2\alpha} \left| \nabla_g u \right|_g^2 \mathrm{e}^{-\frac{|x|^2}{4t}} \mathrm{d}X$$
$$+ C \int t^2 \sigma^{-2-2\alpha} \left| P_1 u \right|^2 \mathrm{e}^{-\frac{|x|^2}{4t}} \mathrm{d}X, \tag{4.113}$$

for every $\alpha \geq 1$, $u \in C_0^\infty\left(\mathbb{R}^n \times [a, +\infty)\right)$ such that $\operatorname{supp} u \subset B_1 \times \left[a, \frac{1}{2\gamma}\right)$, where C depends on λ and Λ only.

Now, by the inequality

$$|P_1u| \le |Pu| + C \left|\nabla_g u\right|_g + C\sqrt{t}\left|D^2u\right|,$$

where C depends on λ and Λ only and by (4.105) and (4.113) we obtain

$$\begin{aligned}
&\mathcal{B}_1 + \frac{1}{C\delta^2}\int \left|D^2u\right|^2 \sigma^{1-2\alpha} e^{-\frac{|x|^2}{4t}} dX \\
&\quad + \frac{1}{C}\int t\theta(\gamma t)\sigma^{-2-2\alpha}\left|\nabla_g u\right|_g^2 e^{-\frac{|x|^2}{4t}} dX \\
&\quad + \frac{(\alpha+1)}{C}\int \theta(\gamma t)\sigma^{-2-2\alpha}u^2 e^{-\frac{|x|^2}{4t}} dX \\
&\quad \le \frac{C}{\delta^2}\left(e^{C_0}\gamma\right)^{2\alpha+\frac{5}{2}}\int\left(u^2 + t\left|\nabla_g u\right|_g^2\right) dX \\
&\quad + C\int t^2\sigma^{-2-2\alpha}|Pu|^2 e^{-\frac{|x|^2}{4t}} dX \\
&\quad + C\int \left|D^2u\right|^2\sigma^{1-2\alpha} e^{-\frac{|x|^2}{4t}} dX + C\int t^2\sigma^{-2-2\alpha}\left|\nabla_g u\right|_g^2 e^{-\frac{|x|^2}{4t}} dX, \quad (4.114)
\end{aligned}$$

for every $\alpha \ge 1$, $u \in C_0^\infty(\mathbb{R}^n \times [a, +\infty))$ such that $\operatorname{supp} u \subset B_1 \times \left[a, \frac{1}{2\gamma}\right)$, where C, $C \ge 1$, depends on λ and Λ only.
Observing that $\theta(\gamma t) \ge \frac{\gamma t}{C}$, whenever $0 < \gamma t \le \frac{1}{2}$, we obtain that, if δ is small enough and $\alpha \ge 1$, the third and the fourth terms on the right-hand side of (4.114) are absorbed by the second and the third terms on the left-hand side of (4.114). Therefore there exists $\delta_1 \in (0, \delta_0]$ such that for every $\delta \in (0, \delta_1]$, $\alpha \ge 1$ and $u \in C_0^\infty(\mathbb{R}^n \times [a, +\infty))$ such that $\operatorname{supp} u \subset B_1 \times \left[a, \frac{1}{2\gamma}\right)$, the following inequality holds true

$$\begin{aligned}
&\frac{1}{C}\int t\theta(\gamma t)\sigma^{-2-2\alpha}\left|\nabla_g u\right|_g^2 e^{-\frac{|x|^2}{4t}} dX \\
&\quad + \frac{(\alpha+1)}{C}\int \theta(\gamma t)\sigma^{-2-2\alpha}u^2 e^{-\frac{|x|^2}{4t}} dX \\
&\quad \le \frac{C}{\delta^2}\left(e^{C_0}\gamma\right)^{2\alpha+2}\int\left(u^2 + t\left|\nabla_g u\right|_g^2\right) dX \\
&\quad + C\int t^2\sigma^{-2-2\alpha}|Pu|^2 e^{-\frac{|x|^2}{4t}} dX - \mathcal{B}_1, \qquad (4.115)
\end{aligned}$$

where $C \ge 1$ depends on λ and Λ only.

Now for a fixed $\delta \in (0, \delta_1]$ and $a \in \left(0, \frac{1}{4\gamma}\right]$ we estimate from above the term $-\mathcal{B}_1$ on the right-hand side of (4.115).

We have

$$-\mathcal{B}_1 = -a^2 \int_{\mathbb{R}^n} \left|\left(\nabla_g u\right)(x,a) + \left(\nabla_g \phi\right)(x,a)\, u\,(x,a)\right|^2_{g(.,a)} \mathrm{e}^{2\phi(x,a)} \mathrm{d}x$$
$$+ a^2 \int_{\mathbb{R}^n} \left(\left|\left(\nabla_g \phi\right)(x,a)\right|^2_{g(.,a)} - \partial_t \phi\,(x,a) - \frac{1}{2a}\right) u^2\,(x,a)\, \mathrm{e}^{2\phi(x,a)} \mathrm{d}x.$$

Let $\varepsilon \in (0,1)$ be a number that we shall choose later. We get

$$-\mathcal{B}_1 \leq -\varepsilon a^2 \int_{\mathbb{R}^n} \left|\left(\nabla_g u\right)(x,a)\right|^2_{g(.,a)} \mathrm{e}^{2\phi(x,a)} \mathrm{d}x$$
$$+ a^2 \int_{\mathbb{R}^n} \left(\frac{1}{1-\varepsilon} \left|\left(\nabla_g \phi\right)(x,a)\right|^2_{g(.,a)} - \partial_t \phi\,(x,a) - \frac{1}{2a}\right)$$
$$\times u^2\,(x,a)\, \mathrm{e}^{2\phi(x,a)} \mathrm{d}x.$$

Denote by $\eta_0 = \min\left\{1, \frac{1}{2\Lambda}\right\}$. Let us choose $\varepsilon = \frac{1}{4}$, we have, for every $x \in B_{\eta_0}$,

$$\frac{1}{1-\varepsilon} \left|\left(\nabla_g \phi\right)(x,a)\right|^2_{g(.,a)} - \partial_t \phi\,(x,a) - \frac{1}{2a}$$
$$\leq \frac{1}{3}\left(2\Lambda\,|x| - 1\right) \frac{|x|^2}{8a^2} + \frac{(\alpha+1)\,\sigma'(a)}{\sigma(a)} \leq \frac{(\alpha+1)}{\mathrm{e}^{C_0}\sigma(a)}.$$

Therefore

$$-\mathcal{B}_1 \leq -\frac{a^2}{4} \int_{\mathbb{R}^n} \left|\left(\nabla_g\right)(x,a)\right|^2_{g(.,a)} \mathrm{e}^{2\phi(x,a)} \mathrm{d}x$$
$$+ \frac{(\alpha+1)\, a^2}{\mathrm{e}^{C_0}\sigma(a)} \int_{\mathbb{R}^n} u^2\,(x,a)\, \mathrm{e}^{2\phi(x,a)} \mathrm{d}x$$

for every $\delta \in (0, \delta_1]$, $a \in \left(0, \frac{1}{4\gamma}\right]$, $\alpha \geq 1$ and $u \in C_0^\infty\left(\mathbb{R}^n \times [a, +\infty)\right)$ such that $supp\, u \subset B_{\eta_0} \times \left[a, \frac{1}{2\gamma}\right)$. This inequality and (4.115) give (4.71). □

In order to prove the next lemma we need some properties of the fundamental solution $\Gamma\,(x,t;\,y,\tau)$ of the adjoint operator $P^* = \partial_i\left(g^{ij}\,(x,t)\,\partial_j u\right) - \partial_t u$ of operator P appearing in (3.30). We refer to [12] for the definition and the proofs of the properties of function $\Gamma\,(x,t;\,y,\tau)$. In what follows we recall some properties of $\Gamma\,(x,t;\,y,\tau)$ that we shall use later on.

(i) For every $(y,\tau) \in \mathbb{R}^{n+1}$, the function $\Gamma\,(.,.;\,y,\tau)$ is a solution of the equation

$$P^*\left(\Gamma\,(.,.;\,y,\tau)\right) = 0, \quad \text{in } \mathbb{R}^n \times (\tau, +\infty)\,. \tag{4.116}$$

(ii) For every $(x,t)\,,(y,\tau) \in \mathbb{R}^{n+1}$ such that $(x,t) \neq (y,\tau)$ we have

$$\Gamma\,(x,t;\,y,\tau) \geq 0 \tag{4.117}$$

and

$$\Gamma(x,t;y,\tau)=0, \quad \text{for every } t<\tau. \tag{4.118}$$

(iii) There exists a constant $C>1$ depending on λ (and n) only such that, for every $(x,t),(y,\tau)\in\mathbb{R}^{n+1}$, $t>\tau$ we have

$$\frac{C^{-1}}{(t-\tau)^{n/2}}\mathrm{e}^{-\frac{C|x-y|^2}{(t-\tau)}}\le\Gamma(xt;y,\tau)\le\frac{C}{(t-\tau)^{n/2}}\mathrm{e}^{-\frac{|x-y|^2}{C(t-\tau)}}. \tag{4.119}$$

(iv) If u_0 is a function on $\mathbb{R}^n$ continuous at a point $y\in\mathbb{R}^n$ which satisfies $\mathrm{e}^{-\mu|x|^2}u_0\in L^2(\mathbb{R}^2)$, for a positive number μ, then

$$\lim_{t\to0^+}\int_{\mathbb{R}^n}\Gamma(x,t;y,0)\,u_0(x)\,\mathrm{d}x=u_0(y). \tag{4.120}$$

Moreover, by standard regularity results [64], [72] we have that, by (3.31) and (4.62), for every $(y,\tau)\in\mathbb{R}^{n+1}$ the function $\Gamma(.,.;y,\tau)$ belongs to $H^{2,1}_{loc}\left(\mathbb{R}^{n+1}\setminus\{(y,\tau)\}\right)$.

LEMMA 4.2.4. *Let P be the operator (3.30) whose coefficients satisfy (3.31)* and *(4.62). Assume that $u\in H^{2,1}(B_1\times(0,1))$ satisfies the inequality*

$$|Pu|\le\Lambda\left(\left|\nabla_g u\right|_g+|u|\right), \quad \text{in } B_1\times[0,1). \tag{4.121}$$

Then there exists a constant $C>1$, depending on λ and Λ only, such that for every $\rho_0,\rho_1,\rho_2,T\in\left(0,\frac{1}{2}\right]$ satisfying $\rho_0<\rho_1<\rho_2$ the following inequality holds true

$$\int_{B_{\rho_2}}u^2(x,t)\,\mathrm{d}x\ge\frac{1}{C}\int_{B_{\rho_0}}u^2(x,0)\,\mathrm{d}x, \quad \text{for every } t\in[0,s_0], \tag{4.122}$$

where

$$s_0=\min\left\{T,\frac{(\rho_1-\rho_0)^2}{C}\left[\left(\log\left(\widetilde{C}N(u)\right)\right)_+\right]^{-1}\right\}, \tag{4.123}$$

(here $x_+=\max\{x,0\}$),

$$N(u)=\frac{\int_{B_{\rho_2}\times(0,2T)}u^2\mathrm{d}X}{\int_{B_{\rho_0}}u^2(x,0)\,\mathrm{d}x}, \tag{4.124}$$

and

$$\widetilde{C}=\frac{C\left(\rho_2^n-\rho_1^n\right)\rho_0^n}{T(\rho_1-\rho_0)^{n-1}(\rho_2-\rho_1)^2}. \tag{4.125}$$

PROOF. Let ρ_1, ρ_2 satisfy $0 < \rho_1 < \rho_2 \le 1$ and let φ be a function belonging to $C_0^2(\mathbb{R}^n)$ and satisfying $0 \le \varphi \le 1$ in $\mathbb{R}^n$, $\varphi = 1$ in B_{ρ_1}, $\varphi = 0$ in $\mathbb{R}^n \setminus B_{\rho_2}$ and

$$(\rho_2 - \rho_1)|\nabla\varphi| + (\rho_2 - \rho_1)^2 \left|D^2\varphi\right| \le C, \quad \text{in } B_{\rho_2} \setminus B_{\rho_1}, \tag{4.126}$$

where C is an absolute constant.

Denote

$$v(x,t) = u(x,t)\,\varphi(x).$$

By (3.31), (4.62), (4.121) and (4.126) we have

$$|Pv| \le C\left(\left|\nabla_g v\right|_g + |v| + K\chi_{B_{\rho_2}\setminus B_{\rho_1}}\right), \tag{4.127}$$

where

$$K = (\rho_2 - \rho_1)^{-2}\|u\|_{L^\infty(B_{\rho_2}\times(0,T))} + (\rho_2 - \rho_1)^{-1}\|\nabla u\|_{L^\infty(B_{\rho_2}\times(0,T))}, \tag{4.128}$$

where $\chi_{B_{\rho_2}\setminus B_{\rho_1}}$ is the characteristic function of $B_{\rho_2} \setminus B_{\rho_1}$ and C depends on λ and Λ only.

Now, let $\rho_0 \in (0, \rho_1)$ and let y be a fixed point of B_{ρ_0} and denote

$$H(t) = \int_{\mathbb{R}^n} v^2(x,t)\,\Gamma(x,t;y,0)\,\mathrm{d}x, \quad s > 0. \tag{4.129}$$

By differentiating H we get

$$\begin{aligned}\frac{\mathrm{d}H(t)}{\mathrm{d}t} &= 2\int_{\mathbb{R}^n} \partial_t v(x,t)\,v(x,t)\,\Gamma(x,t;y,0)\,\mathrm{d}x \\ &\quad + \int_{\mathbb{R}^n} v^2(x,t)\,\partial_t\Gamma(x,t;y,0)\,\mathrm{d}x \\ &= 2\int_{\mathbb{R}^n} (Pv(x,t))\,v(x,t)\,\Gamma(x,t;y,0)\,\mathrm{d}x \\ &\quad + \int_{\mathbb{R}^n} v^2(x,t)\,\partial_t\Gamma(x,s;y,0)\,\mathrm{d}x \\ &\quad - 2\int_{\mathbb{R}^n} \left(\Delta_g v(x,t)\right) v(x,t)\,\Gamma(x,t;y,0)\,\mathrm{d}x.\end{aligned} \tag{4.130}$$

By the identity $2\left(\Delta_g v\right)v = \Delta_g\left(v^2\right) - 2\left|\nabla_g v\right|_g^2$ and integrating by parts we obtain

$$\begin{aligned}&2\int_{\mathbb{R}^n} \left(\Delta_g v(x,t)\right) v(x,t)\,\Gamma(x,t;y,0)\,\mathrm{d}x \\ &\quad = \int_{\mathbb{R}^n} v^2(x,t)\,\Delta_g\Gamma(x,t;y,0)\,\mathrm{d}x - 2\int_{\mathbb{R}^n} \left|\nabla_g v(x,t)\right|_g^2 \Gamma(x,t;y,0)\,\mathrm{d}x.\end{aligned}$$

By plugging the identity just obtained in the right-hand side of (4.130) and by using (4.116) we have

$$\frac{\mathrm{d}H(t)}{\mathrm{d}t} = 2\int_{\mathbb{R}^n} (Pv(x,t))\, v(x,t)\,\Gamma(x,t;y,0)\,\mathrm{d}x + 2\int_{\mathbb{R}^n} \left|\nabla_g v(x,t)\right|_g^2 \Gamma(x,t;y,0)\,\mathrm{d}x. \tag{4.131}$$

By using (4.127) in the identity (4.131) we obtain

$$\begin{aligned}\frac{\mathrm{d}H(t)}{\mathrm{d}t} &\geq 2\int_{\mathbb{R}^n} \left|\nabla_g v(x,t)\right|_g^2 \Gamma(x,t;y,0)\,\mathrm{d}x \\ &\quad - C\int_{\mathbb{R}^n} |v(x,t)| \left(\left|\nabla_g v(x,t)\right|_g + |v(x,t)| + K\chi_{B_{\rho_2}\setminus B_{\rho_1}}\right) \\ &\quad \times \Gamma(x,t;y,0)\,\mathrm{d}x,\end{aligned} \tag{4.132}$$

where C depends on λ and Λ only.

By using the inequality $2ab \leq \frac{a^2}{\varepsilon} + \varepsilon b^2$ to estimate from below the right-hand side of (4.132), we have

$$\frac{\mathrm{d}H(t)}{\mathrm{d}t} \geq -CH(t) - K^2\int_{\mathbb{R}^n} \chi_{B_{\rho_2}\setminus B_{\rho_1}} \Gamma(x,t;y,0)\,\mathrm{d}x, \tag{4.133}$$

where C depends on λ and Λ only.

Since $y \in B_{\rho_0}$, by (4.119) and (4.133) we have

$$\frac{\mathrm{d}H(t)}{\mathrm{d}t} \geq -CH(t) - C_1K^2\left(\rho_2^n - \rho_1^n\right)\frac{\mathrm{e}^{-\frac{(\rho_1-\rho_0)^2}{4C_1 t}}}{t^{n/2}}, \tag{4.134}$$

where C depends on λ and Λ only and C_1 depends on λ only.

Now let us multiply both sides of (4.134) by e^{Ct} and integrate the inequality thus obtained over $(0,T)$. By using (4.120) we get

$$\begin{aligned}H(t)\,\mathrm{e}^{Ct} &\geq u^2(y,0) - C_1K^2\left(\rho_2^n - \rho_1^n\right)\int_0^t \frac{\mathrm{e}^{C\tau-\frac{(\rho_1-\rho_0)^2}{4C_1\tau}}}{\tau^{n/2}}\mathrm{d}\tau \\ &\geq u^2(y,0) - C_1K^2\frac{\left(\rho_2^n - \rho_1^n\right)}{(\rho_1-\rho_0)^{n-1}}\mathrm{e}^{-\frac{(\rho_1-\rho_0)^2}{8C_1 t}}, \\ &\qquad \text{for every } t \in (0, T_0],\end{aligned} \tag{4.135}$$

where $T_0 = \min\left\{\frac{(\rho_1-\rho_0)^2}{16C_1}, T\right\}$, and C_1 depends on λ only.

Now, integrating the inequality (4.135) over B_{ρ_0}, we obtain

$$
\begin{aligned}
&\int_{B_{\rho_0}} \mathrm{d}y \int_{\mathbb{R}^n} v^2(x,t)\,\Gamma(x,t;y,0)\,\mathrm{d}x \\
&\quad \geq \frac{1}{C}\left(\int_{B_{\rho_0}} u^2(y,0)\,\mathrm{d}y - CE^2\frac{\left(\rho_2^n-\rho_1^n\right)\rho_0^n}{(\rho_1-\rho_0)^{n-1}}\mathrm{e}^{-\frac{(\rho_1-\rho_0)^2}{Cs}}\right), \qquad (4.136)
\end{aligned}
$$

where $C > 1$, depends on λ and Λ only.

In addition, by (4.119) we have

$$
\begin{aligned}
&\int_{B_{\rho_0}} \mathrm{d}y \int_{\mathbb{R}^n} v^2(x,t)\,\Gamma(x,t;y,0)\,\mathrm{d}x \\
&\quad \leq \int_{\mathbb{R}^n}\left(v^2(x,t)\int_{\mathbb{R}^n}\frac{C\mathrm{e}^{-\frac{|x-y|^2}{4Ct}}}{t^{n/2}}\mathrm{d}y\right)\mathrm{d}x \leq C'\int_{B_{\rho_2}} u^2(x,t)\,\mathrm{d}x \qquad (4.137)
\end{aligned}
$$

and by standard regularity estimates we have

$$
K^2 \leq \frac{C}{(\rho_2-\rho_1)^2}\frac{1}{\rho_2 T}\int_{B_{\rho_2}\times(0,2T)} u^2\mathrm{d}X, \qquad (4.138)
$$

where C depends on λ and Λ only.

By (4.136)–(4.138) we obtain, for every $t \in (0, T_0]$,

$$
\begin{aligned}
\int_{B_{\rho_2}} u^2(x,t)\,\mathrm{d}x \geq{}& \frac{1}{C}\int_{B_{\rho_0}} u^2(x,0)\,\mathrm{d}x \\
&- C_3\mathrm{e}^{-\frac{(\rho_1-\rho_0)^2}{Ct}}\int_{B_{\rho_2}\times(0,2T)} u^2\mathrm{d}X, \qquad (4.139)
\end{aligned}
$$

where

$$
C_3 = \frac{C\left(\rho_2^n-\rho_1^n\right)\rho_0^n}{T(\rho_1-\rho_0)^{n-1}(\rho_2-\rho_1)^2}
$$

and $C > 1$, depends on λ and Λ only. By (4.139) we obtain (4.122). □

Now let us introduce a notation which we shall use in Lemma 4.2.5 and in Theorem 4.2.6 below.

Let $\alpha \geq 1$, $a \geq 0$, $k \geq 0$, $\rho > 0$ be given numbers. Denote

$$
D_\rho^{(a)} = \left\{(x,t) \in \mathbb{R}^n \times (0,+\infty) : \frac{|x|^2}{4\alpha(t+a)} + \log(t+a) \leq \log\frac{\rho^2}{4\alpha}\right\} \qquad (4.140)
$$

and, for $\rho > (4\alpha a)^{1/2}$,

$$
L_\rho^{(a)} = \sup\left\{(t+a)^{-(\alpha+k)}\,\mathrm{e}^{-\frac{|x|^2}{4(t+a)}} : (x,t) \in D_{2\rho}^{(a)} \setminus D_\rho^{(a)}\right\}. \qquad (4.141)
$$

LEMMA 4.2.5. *For any positive numbers α, a, k and $\rho > 0$ such that $\alpha \geq 1$ and $\rho > (4\alpha a)^{1/2}$, we have*

$$L_{\rho}^{(a)} \leq 4^{\alpha} c_k \left(\frac{\alpha}{\rho^2}\right)^{\alpha+k}, \tag{4.142}$$

where c_k depends on k only.

PROOF. First notice that $L_{\rho}^{(a)} \leq L_{\rho}^{(0)}$ for every $a \geq 0$, hence it is enough to prove (4.142) for $a = 0$.

Now, for any fixed $x \in \mathbb{R}^n$, the function $t \to t^{-\alpha-k}e^{-\frac{|x|^2}{4t}}$ attains its maximum when $t = \frac{|x|^2}{4(\alpha+k)}$. Therefore, in order to estimate $L_{\rho}^{(0)}$ from above, we need to consider the intersection Γ of the paraboloid $\left\{(x,t) \in \mathbb{R}^{n+1} : t = \frac{|x|^2}{4(\alpha+k)}\right\}$ with the set $D_{2\rho}^{(a)} \setminus D_{\rho}^{(a)}$. It is easy to check that the projection of Γ on the t-axis is equal to the interval $I = \left(\frac{\rho^2}{4\alpha e^{1+(k/\alpha)}}, \frac{(2\rho)^2}{4\alpha e^{1+(k/\alpha)}}\right]$. Hence we have

$$\begin{aligned} L_{\rho}^{(0)} &= \sup\left\{t^{-\alpha-k}e^{-\alpha-k} : t \in I\right\} \\ &\leq e^{k(k+1)}\left(4\alpha\rho^{-2}\right)^{\alpha+k}. \end{aligned}$$

□

THEOREM 4.2.6 (*two-sphere one-cylinder inequality in the interior*). *Let λ, Λ and R be positive numbers, with $\lambda \in (0,1]$ and $t_0 \in \mathbb{R}$. Let P be the parabolic operator*

$$P = \partial_i\left(g^{ij}(x,t)\,\partial_j\right) + \partial_t,$$

where $\left\{g^{ij}(x,t)\right\}_{i,j=1}^n$ is a real symmetric $n \times n$ matrix. When $\xi \in \mathbb{R}^n$, $(x,t), (y,\tau) \in \mathbb{R}^{n+1}$ assume that

$$\lambda|\xi|^2 \leq \sum_{i,j=1}^n g^{ij}(x,t)\,\xi_i\xi_j \leq \lambda^{-1}|\xi|^2 \tag{4.143}$$

and

$$\left(\sum_{i,j=1}^n\left(g^{ij}(x,t) - g^{ij}(y,\tau)\right)^2\right)^{1/2} \leq \frac{\Lambda}{R}\left(|x-y|^2 + |t-\tau|\right)^{1/2}. \tag{4.144}$$

Let u be a function belonging to $H^{2,1}\left(B_R \times \left(0, R^2\right)\right)$ which satisfies the inequality

$$|Pu| \leq \Lambda\left(\frac{\left|\nabla_g u\right|_g}{R} + \frac{|u|}{R^2}\right), \quad \textit{in } B_R \times \left[0, R^2\right). \tag{4.145}$$

Then there exist constants $\eta_1 \in (0,1)$ *and* $C \geq 1$, *depending on* λ *and* Λ *only, such that for every r and* ρ *such that* $0 < r \leq \rho \leq \eta_1 R$, *we have*

$$\int_{B_\rho} u^2(x,0)\,\mathrm{d}x \leq \frac{CR}{\rho}\left(R^{-2}\int_{B_R\times(0,R^2)} u^2\mathrm{d}X\right)^{1-\theta_1}\left(\int_{B_r} u^2(x,0)\,\mathrm{d}x\right)^{\theta_1}, \tag{4.146}$$

where

$$\theta_1 = \frac{1}{C\log\frac{R}{r}}. \tag{4.147}$$

PROOF. By scaling we have that the inequality (4.146) is equivalent to

$$\int_{B_\rho} u^2(x,0)\,\mathrm{d}x \leq \frac{C}{\rho}\left(\int_{B_1\times(0,1)} u^2\mathrm{d}X\right)^{1-\theta_1}\left(\int_{B_r} u^2(x,0)\,\mathrm{d}x\right)^{\theta_1},$$

where $\theta_1 = \frac{1}{C\log\frac{1}{r}}$. First we consider the case $g^{ij}(0,0) = \delta^{ij}$ and we write the Carleman estimate ((4.71) in a slightly different form which is more suitable for our purposes. It is easy to check that, by (4.67) and by $C\delta^2\frac{\theta(\gamma t)}{t} \geq \alpha$, whenever $0 < \gamma t \leq \frac{1}{2}$ and by (4.71), we have

$$\begin{aligned}
&\alpha^2\int_{\mathbb{R}^{n+1}_+}\sigma_a^{-\alpha}u^2 e^{-\frac{|x|^2}{4(t+a)}}\mathrm{d}X + \alpha\int_{\mathbb{R}^{n+1}_+}\sigma_a^{-\alpha}\left|\nabla_g u\right|_g^2 e^{-\frac{|x|^2}{4(t+a)}}\mathrm{d}X\\
&\leq C\int_{\mathbb{R}^{n+1}_+}(t+a)^2\sigma_a^{1-\alpha}|Pu|^2 e^{-\frac{|x|^2}{4(t+a)}}\mathrm{d}X\\
&\quad + C^\alpha\alpha^\alpha\int_{\mathbb{R}^{n+1}_+}\left(u^2+(t+a)\left|\nabla_g u\right|_g^2\right)\mathrm{d}X\\
&\quad + C\alpha\sigma^{-\alpha}(a)\int_{\mathbb{R}^n}u^2(x,0)\,e^{-\frac{|x|^2}{4a}}\mathrm{d}x,
\end{aligned} \tag{4.148}$$

for every $\alpha \geq 2$, $0 < a \leq \frac{T_1}{\alpha}$ and $u \in C_0^\infty(\mathbb{R}^n\times[0,+\infty))$ such that $\operatorname{supp} u \subset B_{\eta_0}\times\left[0,\frac{3T_1}{\alpha}\right)$, where $T_1 = \frac{\delta_1^2}{3}$, and $\eta_0 \in (0,1)$, $C > 1$, depend on λ and Λ only (recall that η_0 is defined in Theorem 4.2.3).

By using Friedrichs density theorem we can apply the Carleman estimate (4.148) to the function $v = u\varphi$, where φ is a function belonging to $C_0^\infty\left(B_{\eta_0}\times\left[0,\frac{3T_1}{\alpha}\right)\right)$ that we are going to define.

Let $\overline{R}_0 = \min\left\{\sqrt{T_1},\sqrt{e}\eta_0\right\}$, $R_1 \in \left(0,\overline{R}_0\right]$. For any $\alpha \geq 2$ denote

$$d_1 = \log\frac{(R_1/2)^2}{4\alpha}, \qquad d_2 = \log\frac{R_1^2}{4\alpha}.$$

Let ψ_1 be the function which is equal to 0 in $\mathbb{R} \setminus [d_1, d_2]$, such that $\psi_1(\tau) = \exp \frac{(d_2-d_1)^2}{(d_1-\tau)(\tau-d_2)}$, if $\tau \in (d_1, d_2)$. Denote by ψ_2 the function such that

$$\psi_2(\tau) = \frac{\int_\tau^{d_2} \psi_1(\xi)\,\mathrm{d}\xi}{\int_{d_1}^{d_2} \psi_1(\xi)\,\mathrm{d}\xi}, \quad \text{if } \tau \in \mathbb{R}.$$

It is easy to check that $\psi_2(\tau) = 1$, for every $\tau \in (-\infty, d_1)$, $\psi_2(\tau) = 0$, for every $\tau \in (d_2, +\infty)$ and

$$\left|\psi_2'(\tau)\right| \le \frac{C}{d_2 - d_1}, \qquad \left|\psi_2''(\tau)\right| \le \frac{C}{(d_2-d_1)^2}, \quad \text{if } \tau \in [d_1, d_2], \tag{4.149}$$

where C is an absolute constant. Notice that the right-hand side of the inequalities do not depend on α and R_1.

Now, let us define

$$\varphi(x,t) = \psi_2\left(\frac{|x|^2}{4\alpha(t+a)} + \log(t+a)\right), \tag{4.150}$$

for every $a \in \left(0, \frac{T_1}{\alpha}\right]$ and $\alpha \ge 2$.

It is easy to check that $\varphi \in C_0^\infty\left(B_{\eta_0} \times \left[0, \frac{3T_1}{\alpha}\right)\right)$, $\varphi = 1$ for every $(x,t) \in D_{R_1/2}^{(a)}$ and $\varphi = 0$ for every $(x,t) \in \left(B_{\eta_0} \times \left[0, \frac{3T_1}{\alpha}\right)\right) \setminus D_{R_1}^{(a)}$. Moreover by (4.145) we have that the function $v = u\varphi$ satisfies the inequality

$$\begin{aligned} |Pv| &\le \Lambda\left(\left|\nabla_g u\right|_g + |u|\right)\chi_{D_{R_1}^{(a)}} \\ &\quad + C\left(\frac{|x|\left|\nabla_g u\right|_g}{\alpha(t+a)} + \frac{|u|}{(t+a)^2}\right)\chi_{D_{R_1}^{(a)} \setminus D_{R_1/2}^{(a)}}, \end{aligned} \tag{4.151}$$

where C depends on λ and Λ only.

Now we apply inequality (4.148) to the function v. By (4.151) we obtain, for every $\alpha \ge 2$,

$$\begin{aligned} &\alpha^2 \int_{D_{R_1/2}^{(a)}} \sigma_a^{-\alpha} u^2 \mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X + \alpha \int_{D_{R_1/2}^{(a)}} \sigma_a^{-\alpha}\left|\nabla_g u\right|_g^2 \mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X \\ &\quad \le C \int_{D_{R_1/2}^{(a)}} \sigma_a^{1-\alpha} u^2 \mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X + C \int_{D_{R_1/2}^{(a)}} \sigma_a^{1-\alpha}\left|\nabla_g u\right|_g^2 \mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X \\ &\qquad + C \int_{D_{R_1}^{(a)} \setminus D_{R_1/2}^{(a)}} \sigma_a^{1-\alpha}\left(\frac{|x|^2\left|\nabla_g u\right|_g^2}{\alpha^2(t+a)^2} + \frac{u^2}{(t+a)^4}\right)\mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\,\mathrm{d}X + I, \end{aligned} \tag{4.152}$$

where

$$I = C^{\alpha}\alpha^{\alpha}\int_{D^{(a)}_{R_1/2}}\left\{\left(\frac{|x|^2}{\alpha^2(t+a)^2}+1\right)u^2+\left|\nabla_g u\right|_g^2\right\}\mathrm{d}X$$
$$+ C\alpha\left(\sigma(a)\right)^{-\alpha}\int_{B_1} v^2(x,0)\,\mathrm{e}^{-\frac{|x|^2}{4a}}\mathrm{d}x \tag{4.153}$$

and $C > 1$, depends on λ and Λ only.

Now, for α large enough, the first and the second integral on the right-hand side of (4.152) obtained above, can be absorbed by the left-hand side of (4.152) and we have

$$\alpha^2\int_{D^{(a)}_{R_1/2}}\sigma_a^{-\alpha}u^2\mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\mathrm{d}X$$
$$\leq C\int_{D^{(a)}_{R_1}\smallsetminus D^{(a)}_{R_1/2}}\sigma_a^{1-\alpha}\left(\frac{|x|^2\left|\nabla_g u\right|_g^2}{\alpha^2(t+a)^2}+\frac{u^2}{(t+a)^4}\right)\mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\mathrm{d}X + I, \tag{4.154}$$

for every $\alpha \geq C$, where C depends on λ and Λ only.

In order to estimate from above the first integral at the right-hand side of (4.154) we use (4.67) and Lemma 4.2.5, so we get

$$\int_{D^{(a)}_{R_1}\smallsetminus D^{(a)}_{R_1/2}}\sigma_a^{1-\alpha}\left(\frac{|x|^2\left|\nabla_g u\right|_g^2}{\alpha^2(t+a)^2}+\frac{u^2}{(t+a)^4}\right)\mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\mathrm{d}X$$
$$\leq C^{\alpha}\int_{D^{(a)}_{R_1}\smallsetminus D^{(a)}_{R_1/2}}(t+a)^{-\alpha-3}\left(\left|\nabla_g u\right|_g^2+u^2\right)\mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\mathrm{d}X$$
$$\leq \frac{C'^{\alpha}\alpha^{\alpha}}{R_1^{2(\alpha+3)}}\int_{D^{(a)}_{R_1}\smallsetminus D^{(a)}_{R_1/2}}\left(\left|\nabla_g u\right|_g^2+u^2\right)\mathrm{d}X, \tag{4.155}$$

for every $\alpha \geq C$ and every $a \in \left(0, \frac{R_1^2}{16\alpha}\right)$, where C $C > 1$, and $C' > 1$, depend on λ and Λ only.

By the inequality obtained in (4.155) and by (4.154) we have

$$\alpha^2\int_{D^{(a)}_{R_1/2}}\sigma_a^{-\alpha}u^2\mathrm{e}^{-\frac{|x|^2}{4(t+a)}}\mathrm{d}X$$
$$\leq \frac{C^{\alpha}\alpha^{\alpha}}{R_1^{2(\alpha+3)}}\int_{D^{(a)}_{R_1}\smallsetminus D^{(a)}_{R_1/2}}\left(\left|\nabla_g u\right|_g^2+u^2\right)\mathrm{d}X + I, \tag{4.156}$$

for every $\alpha \geq C$ and every $a \in \left(0, \frac{R_1^2}{16\alpha}\right)$, where $C > 1$ depends on λ and Λ only.

Now we estimate from above the term I (defined in (4.153)) on the right-hand side of (4.156). Concerning the first integral at the right-hand side of formula (4.153) it is simple to check that

$$\int_{D_{R_1}^{(a)}} \left\{ \left(\frac{|x|^2}{\alpha^2 (t+a)^2} + 1 \right) u^2 + \left| \nabla_g u \right|_g^2 \right\} \mathrm{d}X$$

$$\le C \, \|u\|^2_{L^\infty\left(D_{R_1}^{(a)}\right)} + \int_{D_{R_1}^{(a)}} \left| \nabla_g u \right|_g^2 \mathrm{d}X, \tag{4.157}$$

for every $\alpha \ge 1$, where C is an absolute constant. Concerning the second integral at the right-hand side of (4.153) we have

$$\int_{B_1} v^2 (x, 0) \, \mathrm{e}^{-\frac{|x|^2}{4a}} \mathrm{d}x \le \int_{B_{r(a)}} u^2 (x, 0) \, \mathrm{d}x, \tag{4.158}$$

where

$$r(a) = \left(4\alpha a \log \frac{R_1^2}{4\alpha a} \right)^{1/2}.$$

By (4.156)–(4.158) we have

$$\alpha^2 \int_{D_{R_1/2}^{(a)}} \sigma_a^{-\alpha} u^2 \mathrm{e}^{-\frac{|x|^2}{4(t+a)}} \mathrm{d}X \le \frac{C^\alpha \alpha^\alpha}{R_1^{2(\alpha+3)}} \left(\|u\|^2_{L^\infty\left(D_{R_1}^{(a)}\right)} + \int_{D_{R_1}^{(a)}} \left| \nabla_g u \right|_g^2 \mathrm{d}X \right) + \frac{C^\alpha}{a^\alpha} \int_{B_{r(a)}} u^2 (x, 0) \, \mathrm{d}x, \tag{4.159}$$

for every $\alpha \ge C$ and every $a \in \left(0, \frac{R_1^2}{16\alpha} \right)$, where $C > 1$, depends on λ and Λ only.

Let r be a number in the interval $\left(0, \mathrm{e}^{-1/2} R_1\right)$ and, for every $\alpha \ge C$, let $\overline{a}$ belong to $\left(0, \mathrm{e}^{-1} \frac{R_1^2}{4\alpha} \right)$, such that

$$r(\overline{a}) = \left(4\alpha \overline{a} \log \frac{R_1^2}{4\alpha \overline{a}} \right)^{1/2} = r. \tag{4.160}$$

By asymptotic estimates of $\overline{a}$ we have

$$\frac{1}{4e\alpha} r^2 \left(\log \frac{R_1^2}{r^2} \right)^{-1} \le \overline{a} \le \frac{1}{4\alpha} r^2 \left(\log \frac{R_1^2}{r^2} \right)^{-1}. \tag{4.161}$$

Furthermore, using standard regularity estimates [72], we estimate from above the first and second integrals on the right-hand side of (4.159) and we get

$$\alpha^2 \int_{D_{R_1/2}^{(\overline{a})}} \sigma_{\overline{a}}^{-\alpha} u^2 \mathrm{e}^{-\frac{|x|^2}{4(t+\overline{a})}} \mathrm{d}X$$
$$\leq \frac{C^\alpha \alpha^\alpha}{R_1^{2(\alpha+5)}} \int_{B_{R_1}\times\left(0,R_1^2\right)} u^2 \mathrm{d}X + \frac{C^\alpha}{\overline{a}^\alpha} \int_{B_r} u^2(x,0)\,\mathrm{d}x, \tag{4.162}$$

for every $\alpha \geq C$, where C, $C > 1$, depends on λ and Λ only.

Now we estimate from below the left-hand side of (4.162). Let $\rho \in \left(0, \frac{R_1}{2}\right)$, by (4.67) we have

$$\int_{D_{R_1/2}^{(\overline{a})}} \sigma_{\overline{a}}^{-\alpha} u^2 \mathrm{e}^{-\frac{|x|^2}{4(t+\overline{a})}} \mathrm{d}X$$
$$\geq \int_{D_\rho^{(\overline{a})}} \sigma_{\overline{a}}^{-\alpha} u^2 \mathrm{e}^{-\frac{|x|^2}{4(t+\overline{a})}} \mathrm{d}X \geq \frac{1}{C^\alpha}\left(\frac{\rho^2}{4\alpha}\right)^{-\alpha} \int_{D_\rho^{(\overline{a})}} u^2 \mathrm{d}X, \tag{4.163}$$

where $C > 1$, depends on λ and Λ only. By the inequalities obtained in (4.161)–(4.163) we have

$$\alpha^2 \int_{D_\rho^{(\overline{a})}} u^2 \mathrm{d}X \leq \frac{\left(C\rho^2\right)^\alpha}{R_1^{2(\alpha+5)}} \int_{B_{R_1}\times\left(0,R_1^2\right)} u^2 \mathrm{d}X$$
$$+ C^\alpha \rho^{2\alpha} \left(r^2\left(\log\frac{R_1^2}{r^2}\right)^{-1}\right)^{-\alpha} \int_{B_r} u^2(x,0)\,\mathrm{d}x, \tag{4.164}$$

for every $\alpha \geq C$, where $C > 1$, depends on λ and Λ only.

Denote

$$t_1 = \frac{\rho^2 \mathrm{e}^{-1}}{8\alpha} - \overline{a},\ t_2 = \frac{\rho^2 \mathrm{e}^{-1}}{8\alpha}\left(1+\mathrm{e}^{-1}\right) - \overline{a},$$
$$Q = B_{\frac{\rho}{2\mathrm{e}^{1/2}}} \times (t_1, t_2)\,,$$
$$M = \left(\int_{B_{R_1}\times\left(0,R_1^2\right)} u^2 \mathrm{d}X\right)^{1/2}, \qquad \varepsilon = \left(\int_{B_r} u^2(x,0)\,\mathrm{d}x\right)^{1/2}.$$

Observing that for every $\rho \in \left(\frac{r}{\mathrm{e}^{1/2}}, \frac{R_1}{2}\right)$ we have $t_2 > t_1 > 0$ and $Q \subset D_\rho^{(\overline{a})}$, inequality (4.164) gives

$$\alpha^2 \int_Q u^2 \mathrm{d}X \leq \frac{(C\rho)^{2\alpha}}{R_1^{2(\alpha+5)}} M^2 + \left(\frac{C\rho}{r\left(\log R_1^2/r^2\right)^{-1/2}}\right)^{2\alpha} \varepsilon^2, \tag{4.165}$$

for every $\alpha \geq C$, where $C > 1$, depends on λ and Λ only.

In order to estimate from below the left-hand side of inequality (4.165) we apply

Lemma 4.2.4. In doing so we choose $\rho_0 = \frac{e^{-1/2}\rho}{4}$, $\rho_1 = \frac{3e^{-1}\rho}{8}$, $\rho_2 = \frac{e^{-1/2}\rho}{2}$ and $T = \rho_2^2$. Let us denote

$$K = \left(\int_{B_{\rho_0}} u^2 (x, 0)\, dx \right)^{1/2}.$$

We have that there exists a constant $C_0 > 1$, depending on λ and Λ only such that if

$$\alpha \geq \alpha_0 := C_0 \max \left\{ 1, \log \left(\frac{\rho^{n-4} M^2}{K^2} \right) \right\} \tag{4.166}$$

then

$$\begin{aligned} \int_Q u^2 dX = \int_{t_1}^{t_2} dt \int_{B_{\rho_2}} u^2 (x, t)\, dx \geq \frac{t_2 - t_1}{C} \int_{B_{\rho_0}} u^2 (x, 0)\, dx \\ = \frac{\rho^2}{8\alpha C} K^2. \end{aligned} \tag{4.167}$$

By inequality (4.165) and (4.167) we get

$$\rho^2 K^2 \leq \frac{(C\rho)^{2\alpha}}{R_1^{2(\alpha+5)}} M^2 + \left(\frac{C\rho}{r \left(\log R_1^2/r^2 \right)^{-1/2}} \right)^{2\alpha} \varepsilon^2, \tag{4.168}$$

for every $\alpha \geq \alpha_0$ and every r, R_1, ρ satisfying the relations $R_1 \in \left(0, \overline{R}_0\right]$, $r \in \left(0, e^{-1/2} R_1\right]$, $\rho \in \left(e^{1/2} r, \frac{R_1}{2}\right]$, where $C > 1$ depends on λ and Λ only.

Denote by

$$\alpha_1 = \frac{\log \left(M^2 \varepsilon^{-2} \right)}{\log \left(R_1^2 r^{-2} \log \left(R_1^2 r^{-2} \right) \right)}.$$

The following cases occur: (i) $\alpha_1 \geq \alpha_0$, (ii) $\alpha_1 < \alpha_0$. If case (i) occurs then we choose in (4.168) $\alpha = \alpha_1$ and we have

$$\rho^2 K^2 \leq \frac{2}{R_1^{10}} M^{2(1-\theta_0)} \varepsilon^{2\theta_0}, \tag{4.169}$$

where

$$\theta_0 = \frac{\log \left(M^2 \rho^{-2}/C \right)}{\log \left(R_1^2 r^{-2} \log \left(R^2 r^{-2} \right) \right)}, \tag{4.170}$$

where $C > 1$, depends on λ and Λ only.

Now consider case (ii). We have either

$$\alpha_0 = C_0 \log\left(\frac{\rho^{n-4}M^2}{K^2}\right) \geq C_0, \tag{4.171}$$

or

$$\alpha_0 = C_0 \leq C_0 \log\left(\frac{\rho^{n-4}M^2}{K^2}\right), \tag{4.172}$$

where C_0 is the same constant which appears in (4.166). Let us introduce the notation

$$\widetilde{\theta}_0 = \frac{1}{C_0 \log\left(R_1^2 r^{-2} \log\left(R_1^2 r^{-2}\right)\right)}, \qquad \theta_1 = \frac{1}{C_0 \log \frac{1}{r}}.$$

If (4.171) occurs then, recalling that $\alpha_1 < \alpha$, we have trivially

$$\theta_1 \log\left(\frac{M^2}{\varepsilon^2}\right) \leq \log\left(\frac{\rho^{n-4}M^2}{K^2}\right),$$

hence

$$K^2 \leq M^{2\left(1-\widetilde{\theta}_0\right)} \varepsilon^{2\widetilde{\theta}_0}. \tag{4.173}$$

If (4.172) occurs then we have trivially

$$\widetilde{\theta}_0 \log\left(\frac{M^2}{\varepsilon^2}\right) \leq 1,$$

hence

$$M^{2\theta_1} \leq e\varepsilon^{2\theta_1}. \tag{4.174}$$

Also, by standard regularity estimates for parabolic equations we have

$$K^2 \leq \frac{C_1}{R_1^2} M^2, \tag{4.175}$$

where C_1 depends on λ only. Therefore (4.174) and (4.175) yield

$$K^2 \leq \frac{C_1}{R_1^2} M^2 = \frac{C_1}{R_1^2} M^{2\left(1-\widetilde{\theta}_0\right)} M^{2\widetilde{\theta}_0} \leq \frac{C_1}{R_1^2} M^{2\left(1-\widetilde{\theta}_0\right)} \varepsilon^{2\widetilde{\theta}_0}. \tag{4.176}$$

Now there exists a constant $C > 1$, depending on λ and Λ only, such that if $\rho^2 \leq \frac{R_1^2}{\mathrm{e}^{1/C} C}$ then we have $\widetilde{\theta}_0 \leq \theta_0$. Therefore by (4.169), (4.173), (4.175) and (4.176), we get

$$K^2 \leq \frac{C_2}{R_1^5 \rho} M^{2\left(1-\widetilde{\theta}_0\right)} \varepsilon^{2\widetilde{\theta}_0}, \tag{4.177}$$

for every $R_1 \in \left(0, \overline{R}_0\right]$, $r \in \left(0, \mathrm{e}^{-1/2} R_1\right]$, $\rho \in \left(\mathrm{e}^{1/2} r, \frac{R_1}{C \mathrm{e}^{1/(2C)}}\right]$, where $C_2 > 1$, depends on λ only.

In the case $g^{ij}(0,0) \neq \delta^{ij}$ we can consider a linear transformation $S : \mathbb{R}^n \to \mathbb{R}^n$, $Sx = \left\{ S^i_j x_j \right\}_{i=1}^n$, such that, denoting $\widetilde{g}^{ij}(y,t) = \frac{1}{|\det S|} S^i_k g^{kl}\left(S^{-1} y, t\right) S^j_l$, we have $\widetilde{g}^{ij}(0,0) = \delta^{ij}$ and

$$B_{\varrho/\sqrt{\lambda}} \subset S\left(B_\varrho\right) \subset B_{\sqrt{\lambda}\varrho}, \quad \text{for every } \varrho > 0.$$

Therefore, by a change of variables, by using the inequality (4.177) and the trivial inequality $r^2 \left(\log \frac{1}{r^2}\right)^{-1} > r^3$, for every $r \in (0,1)$ the thesis follows. □

COROLLARY 4.2.7 (*Spacelike strong unique continuation*). *Let $u \in H^{2,1}\left(B_R \times \left(0, R^2\right)\right)$ satisfy inequality* (4.145).

If for every $k \in \mathbb{N}$ we have

$$\int_{B_r} u^2(x,0)\,\mathrm{d}x = O\left(r^{2k}\right), \quad \textit{as } r \to 0, \tag{4.178}$$

then

$$u(.,0) = 0, \quad \textit{in } B_R.$$

PROOF. Is the same as the proof of Corollary 4.1.7. □

In order to state the theorem below we need to introduce some new notations. Let E and R be positive numbers, $t_0 \in \mathbb{R}$ and let $\varphi : \mathbb{R}^{n-1} \times \mathbb{R} \to \mathbb{R}$ be a function satisfying the conditions

$$\varphi(0,0) = |\nabla_{x'} \varphi(0,0)| = 0, \tag{4.179}$$

$$\|\varphi\|_{L^\infty\left(Q_R^{t_0\prime}\right)} + R\,\|\nabla_{x'}\varphi\|_{L^\infty\left(Q_R^{t_0\prime}\right)} + R^2 \left\|D^2_{x'}\varphi\right\|_{L^\infty\left(Q_R^{t_0\prime}\right)} + R^2\,\|\partial_t \varphi\|_{L^\infty\left(Q_R^{t_0\prime}\right)} \leq ER, \tag{4.180}$$

where $Q'_R = B'_R \times \left[0, R^2\right)$.

For any numbers $\rho \in (0, R]$, $\tau \in \left[0, R^2\right)$, denote by

$$Q_{\rho,\varphi} = \left\{ (x,t) \in B_\rho \times \left(0, \rho^2\right) : x_n > \varphi\left(x', t\right) \right\},$$
$$Q_{\rho,\varphi}(\tau) = \left\{ x \in B_\rho : x_n > \varphi\left(x', \tau\right) \right\},$$
$$\Gamma_{\rho,\varphi} = \left\{ (x,t) \in B_\rho \times \left(0, \rho^2\right) : x_n = \varphi\left(x', t\right) \right\}.$$

THEOREM 4.2.8 (*two-sphere one-cylinder inequality at the boundary*). *Let P be a second order parabolic operator as in Theorem 4.2.6. Let $\varphi : \mathbb{R}^{n-1} \times \mathbb{R} \to \mathbb{R}$ be a function* *satisfying* (4.179) and (4.180). *Assume that* $u \in H^{2,1}\left(Q_{R,\varphi}\right)$ *satisfies the inequality*

$$|Pu| \leq \Lambda\left(\frac{|\nabla u|}{R} + \frac{|u|}{R^2}\right), \quad \text{in } Q_{R,\varphi} \tag{4.181}$$

and

$$u\left(x,t\right) = 0, \quad \text{for every } \left(x,t\right) \in \Gamma_{R,\varphi}. \tag{4.182}$$

Then there exist constants $\eta_2 \in (0,1)$ *and* $C > 1$, *depending on* λ, Λ *and* E *only, such that for every* r, ρ, $0 < r \leq \rho \leq \eta_2 R$, *we have*

$$\begin{aligned}&\int_{Q_{\rho,\varphi}(0)} u^2\left(x,0\right)\mathrm{d}x \\ &\quad \leq \frac{CR^2}{\rho^2}\left(R^{-2}\int_{Q_{R,\varphi}} u^2 \mathrm{d}X\right)^{1-\theta_2}\left(\int_{Q_{r,\varphi}(0)} u^2\left(x,0\right)\mathrm{d}x\right)^{\theta_2},\end{aligned} \tag{4.183}$$

where

$$\theta_2 = \frac{1}{C\log\frac{R}{r}}. \tag{4.184}$$

For the sake of simplicity, here we give a proof of Theorem 4.2.8 with the following additional condition on the matrix $\left\{g^{ij}\left(x,t\right)\right\}_{i,j=1}^{n}$:

$$\left(\sum_{i,j=1}^{n}\left(g^{ij}\left(x,t\right) - g^{ij}\left(x,\tau\right)\right)^2\right)^{1/2} \leq \frac{\Lambda}{R^2}\left|t-\tau\right|, \tag{4.185}$$

for every $\left(x,t\right), \left(x,\tau\right) \in \mathbb{R}^{n+1}$.

An outline of the proof of Theorem 4.2.8 (without condition ((4.185)) is given in [33].

We need the following proposition proved in [93].

PROPOSITION 4.2.9. *Let* λ_0 *and* Λ_0 *be positive numbers, with* $\lambda_0 \in (0,1]$. *Let* $A\left(t\right)$ *be a* $n \times n$ *symmetric real matrix whose entries are Lipschitz continuous in* $\mathbb{R}$. *Assume that*

$$\lambda_0\left|\xi\right|^2 \leq A\left(t\right)\xi\cdot\xi \leq \lambda_0^{-1}\left|\xi\right|^2, \quad \text{for every } \xi \in \mathbb{R}^n \quad \text{and} \quad s \in \mathbb{R}, \tag{4.186}$$

$$\left|\frac{\mathrm{d}}{\mathrm{d}s}A\left(t\right)\right| \leq \Lambda_0, \quad \text{a.e. in } \mathbb{R}. \tag{4.187}$$

Denote by $\sqrt{A\left(t\right)}$ *the positive square root of* $A\left(t\right)$. *We have*

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\sqrt{A\left(t\right)}\right| \leq 2^{3/2}\lambda_0^{-7/2}\Lambda_0, \quad \text{a.e. in } \mathbb{R}. \tag{4.188}$$

PROOF OF THEOREM 4.2.8 WITH CONDITION (4.185). Let us introduce the following notations. Let $\psi_1 : \mathbb{R}^n \to \mathbb{R}^n$ and $\Phi_1 : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ be, respectively, the maps $\psi_1(y,\tau) = \left(y', y_n + \varphi\left(y',\tau\right)\right)$ and $\Phi_1(y,\tau) = (\psi_1(y,\tau), \tau)$. We have $\Phi_1\left(y', 0, \tau\right) = \left(y', \varphi\left(y',\tau\right), \tau\right)$, for every $y' \in \mathbb{R}^{n-1}$, $\tau \in \mathbb{R}$. Furthermore, denoting $R_1 = \frac{R}{1+2E}$, by (4.180) we have

$$\Phi_1\left(B_1^+ \times \left[0, R_1^2\right)\right) \subset Q_{R,\varphi}. \tag{4.189}$$

Denote

$$g_1^{-1}(y,\tau) = \left(\frac{\partial \psi_1(y,\tau)}{\partial y}\right)^{-1} g^{-1}(\Phi_1(y,\tau)) \left(\left(\frac{\partial \psi_1(y,\tau)}{\partial y}\right)^{-1}\right)^*,$$

(here $\frac{\partial \psi_1(y,\tau)}{\partial y}$ is the Jacobian matrix of ψ_1),

$$u_1(y,\tau) = u(\Phi_1(y,\tau))$$

and

$$(P_1 u_1)(y,\tau) = \operatorname{div}\left(g_1^{-1}(y,\tau)\nabla_y u_1\right) + \partial_\tau u_1.$$

We have

$$(Pu)(\Phi_1(y,\tau)) = (P_1 u_1)(y,\tau) - \left(\frac{\partial \psi_1(y,\tau)}{\partial \tau}\right)^* \left(\left(\frac{\partial \psi_1(y,\tau)}{\partial y}\right)^{-1}\right)^* \nabla_y u_1.$$

By (4.143), (4.144), (4.180), (4.181), (4.185) and (4.189) we have that there exist $\lambda_1 \in (0,1]$, $\Lambda_1 > 0$, λ_1 depending on λ and E only, and Λ_1 depending on λ, Λ and E only, such that, when $(y,\tau), (z,s) \in \mathbb{R}^{n+1}$, $\xi \in \mathbb{R}^n$, we have

$$\lambda_1 |\xi|^2 \le \sum_{i,j=1}^n g_1^{ij}(y,\tau)\xi_i\xi_j \le \lambda_1^{-1}|\xi|^2, \tag{4.190}$$

$$\left(\sum_{i,j=1}^n \left(g_1^{ij}(y,\tau) - g_1^{ij}(z,s)\right)^2\right)^{1/2} \le \frac{\Lambda_1}{R_1}\left(|y-z|^2 + |\tau - s|\right)^{1/2}, \tag{4.191}$$

$$\left(\sum_{i,j=1}^n \left(g_1^{ij}(y,\tau) - g_1^{ij}(y,s)\right)^2\right)^{1/2} \le \frac{\Lambda_1}{R_1^2}|\tau - s| \tag{4.192}$$

and

$$|P_1 u_1| \le \Lambda_1\left(\frac{|\nabla u_1|}{R_1} + \frac{|u_1|}{R_1^2}\right), \quad \text{in } B_1^+ \times \left[0, R_1^2\right). \tag{4.193}$$

Furthermore

$$u_1\left(y', 0, \tau\right) = 0, \quad \text{for every } y' \in B'_{R_1}, \tau \in \left[0, R_1^2\right). \tag{4.194}$$

Now let $H(\tau)$ and $S(\tau)$ be, respectively, the positive square root of $g_1(0, \tau)$ and the Lipschitz continuous rotation of $\mathbb{R}^n$, such that $S(\tau)\, e_n = vers\,(H(\tau)\, e_n)$, for every $\tau \in \mathbb{R}$. Denote

$$\begin{aligned}
&K(\tau) = (H(\tau))^{-1} S(\tau),\\
&g_2^{-1}(\widetilde{y}, \tau) = (K(\tau))^{-1} g_1^{-1}(K(\tau)\,\widetilde{y}, \tau)\left((K(\tau))^{-1}\right)^*,\\
&u_2(\widetilde{y}, \tau) = u_1(K(\tau)\,\widetilde{y}, \tau),\\
&(P_2 u_2)(\widetilde{y}, \tau) = \operatorname{div}\left(g_2^{-1}(\widetilde{y}, \tau)\,\nabla_{\widetilde{y}} u_2\right) + \partial_\tau u_2.
\end{aligned}$$

We have

$$\begin{aligned}
&g_2^{-1}(0, \tau) = I_n, \quad \text{for every } \tau \in \mathbb{R},\\
&(P_1 u_1)(K(\tau)\,\widetilde{y}, \tau) = (P_2 u_2)(\widetilde{y}, \tau) - \left(K'(\tau)\,\widetilde{y}\right) K^{-1}(\tau)\,\nabla_{\widetilde{y}} u_2.
\end{aligned}$$

By (4.190)–(4.194) and by Proposition 4.2.9 we have that there exist $\lambda_2 \in (0, 1]$, $\Lambda_2 > 0$ and $C > 1$, λ_2 depending on λ and E only, and Λ_2, C depending on λ, Λ and E only, such that we have, setting $R_2 = \frac{R_1}{C}$, for every $(\widetilde{y}, \tau), (\widetilde{y}_*, s) \in \mathbb{R}^{n+1}$, $\xi \in \mathbb{R}^n$,

$$\lambda_2 |\xi|^2 \le \sum_{i,j=1}^{n} g_2^{ij}(\widetilde{y}, \tau)\,\xi_i \xi_j \le \lambda_2^{-1} |\xi|^2, \tag{4.195}$$

$$\left(\sum_{i,j=1}^{n} \left(g_2^{ij}(\widetilde{y}, \tau) - g_2^{ij}(\widetilde{y}_*, s)\right)^2\right)^{1/2} \le \frac{\Lambda_2}{R_2}\left(|\widetilde{y} - \widetilde{y}_*|^2 + |\tau - s|\right)^{1/2}, \tag{4.196}$$

$$\left(\sum_{i,j=1}^{n} \left(g_2^{ij}(\widetilde{y}, \tau) - g_2^{ij}(\widetilde{y}_*, s)\right)^2\right)^{1/2} \le \frac{\Lambda_2}{R_2^2} |\tau - s|, \tag{4.197}$$

$$|P_2 u_2| \le \Lambda_2 \left(\frac{|\nabla_x u_2|}{R_2} + \frac{|u_2|}{R_2^2}\right), \quad \text{in } B_2^+ \times \left[0, R_2^2\right) \tag{4.198}$$

and

$$u_2\left(\widetilde{y}', 0, \tau\right) = 0, \quad \text{for every } \widetilde{y}' \in B'_{R_2}, \tau \in \left[0, R_2^2\right). \tag{4.199}$$

For the sake of brevity, we shall omit the sign "$\sim$" over y.

Now we adapt ideas in [1] to the case of time varying coefficients. Let ζ be a function belonging to $C_0^\infty\left(\mathbb{R}^{n-1}\right)$, such that $\operatorname{supp}\zeta \subset B'_1$, $\zeta \ge 0$ and $\int_{\mathbb{R}^{n-1}} \zeta\left(y'\right) \mathrm{d}y' = 1$. Denote by $\zeta_{(n)}$ the function $y_n^{1-n}\zeta\left(\frac{y'}{y_n}\right)$. Denote

$$\begin{aligned}
&\beta\left(z',\tau\right) = -g_2^{-1}\left(z',0,\tau\right)e_n,\\
&w_j\left(z',z_n,\tau\right) = z_j - z_n\left(\zeta_{(n)} * \beta_j\left(.,\tau\right)\right)\left(z'\right), \quad j = 1,\dots,n-1,\\
&w_n\left(z',z_n,\tau\right) = -z_n\left(\zeta_{(n)} * \beta_n\left(.,\tau\right)\right)\left(z'\right).
\end{aligned}$$

With $\psi_2\left(.,\tau\right) : \mathbb{R}^n \to \mathbb{R}^n$ we denote the map whose components are defined as follows, for each $j = 1,\dots,n$

$$\left(\psi_2\right)_j\left(z',z_n,\tau\right) = \begin{cases} w_j\left(z',z_n,\tau\right), & \text{for } z_n \geq 0,\\ 4w_j\left(z',z_n,\tau\right) - 3w_j\left(z',z_n,\tau\right), & \text{for } z_n < 0.\end{cases}$$

It turns out that ψ_2 is a $C^{1,1}$ function with respect to z and it is a Lipschitz continuous function with respect to $\tau \in \left[0,R_2^2\right)$. Moreover there exist constants $C_1, C_2, C_3 \in [1,+\infty)$ depending on λ, Λ and E only, such that, setting $\rho_1 = \frac{R_2}{C_1}$, $\rho_2 = \frac{R_2}{C_2}$, we have, for every $\tau \in \left[0,R_2^2\right)$,

(i) $\psi_2\left(z,\tau\right) \in B_{\rho_1}$, for every $z \in B_{\rho_2}$,

(ii) $\psi_2\left(z',0,\tau\right) = \left(z',0\right)$, for every $z' \in B'_{\rho_2}$,

(iii) $\psi_2\left(z,\tau\right) \in B^+_{\rho_1}$, for every $z \in B^+_{\rho_2}$,

(iv) $C_3^{-1}\left|z - z_*\right| \leq \left|\psi_2\left(z,\tau\right) - \psi_2\left(z_*,\tau\right)\right| \leq C_3\left|z - z_*\right|$, for every $z, z_* \in B_{\rho_2}$,

(v) $\left|\frac{\partial^2\psi_2(z,\tau)}{\partial z_i\partial z_j}\right| \leq C_3$, for every $z \in B_{\rho_2}$,

(vi) $C_3^{-1} \leq \left|\det\frac{\partial\psi_2(z,\tau)}{\partial z}\right| \leq C_3$, for every $z \in B_{\rho_2}$.

Let us denote

$$\begin{aligned}
&u_3\left(z,\tau\right) = u_2\left(\psi_2\left(z,\tau\right),\tau\right), \; J\left(z,\tau\right) = \left|\det\frac{\partial\psi_2\left(z,\tau\right)}{\partial z}\right|,\\
&g_3^{-1}\left(z,\tau\right) = \left(\frac{\partial\psi_2\left(z,\tau\right)}{\partial y}\right)^{-1} g_2^{-1}\left(\psi_2\left(z,\tau\right),\tau\right)\left(\left(\frac{\partial\psi_2\left(z,\tau\right)}{\partial z}\right)^{-1}\right)^*.
\end{aligned}$$

We have, for every $\tau \in \left[0,R_2^2\right)$ and $z' \in B'_{\rho_2}$,

$$\begin{aligned}
&J\left(0,\tau\right)g_3^{-1}\left(0,\tau\right) = I_n,\\
&g_3^{nj}\left(z',0,\tau\right) = g_3^{jn}\left(z',0,\tau\right) = 0, \quad \text{for } j = 1,\dots,n-1.
\end{aligned}$$

Moreover by (4.195)–(4.197) we have that there exist $\lambda_3 \in (0,1]$, $\Lambda_3 > 0$ depending on λ, Λ and E only, such that for every $\left(z,\tau\right), \left(z_*,s\right) \in B^+_{\rho_2} \times \left[0,\rho_2^2\right)$ and every $\xi \in \mathbb{R}^n$ we have

$$\lambda_3\left|\xi\right|^2 \leq \sum_{i,j=1}^n g_3^{ij}\left(z,\tau\right)\xi_i\xi_j \leq \lambda_3^{-1}\left|\xi\right|^2,$$

$$\left(\sum_{i,j=1}^{n}\left(g_3^{ij}(z,\tau)-g_3^{ij}(z_*,s)\right)^2\right)^{1/2}\leq\frac{\Lambda_3}{R_3}\left(|z-z_*|^2+|\tau-s|\right)^{1/2}.$$

By (4.198) we have

$$\left|\operatorname{div}\left(g_3^{-1}\nabla_z u_3\right)+\partial_\tau u_3\right|\leq C\left(\frac{|\nabla u_3|}{\rho_2}+\frac{|u_3|}{\rho_2^2}\right),\quad \text{in } B_{\rho_2}^+\times\left[0,\rho_2^2\right),$$

where C depends on λ, Λ and E only.

For every $(z,\tau)\in B_{\rho_2}\times\left(-\rho_2^2,0\right]$, let us denote by $\overline{g}_3^{-1}(z,\tau)=\left\{\overline{g}_3^{ij}(z,\tau)\right\}_{i,j=1}^n$ the symmetric matrix whose entries are given by

$$\begin{aligned}
\overline{g}_3^{ij}\left(z',z_n,\tau\right)&=g_3^{ij}\left(z',|z_n|,\tau\right),\quad \text{if either } 1\leq i,j\leq n-1\quad \text{or } i=j=n,\\
\overline{g}_3^{nj}\left(z',z_n,\tau\right)&=\overline{g}_3^{jn}\left(z',z_n,\tau\right)\\
&=\operatorname{sgn}(z_n)\,g_3^{jn}\left(z',|z_n|,\tau\right),\quad \text{if } 1\leq j\leq n-1.
\end{aligned}$$

It turns out that $\overline{g}_3^{-1}$ satisfies the same ellipticity condition and Lipschitz condition as g_3^{-1}.

Denoting

$$v(z,\tau)=\operatorname{sgn}(z_n)\,u_3\left(z',|z_n|,\tau\right),\quad \text{for every } (z,\tau)\in B_{\rho_2}\times\left[0,\rho_2^2\right),$$

we have that v belongs to $H^{2,1}\left(B_{\rho_2}\times\left(-\rho_2^2,0\right)\right)$ and satisfies the inequality

$$\left|\operatorname{div}\left(\overline{g}_3^{-1}\nabla_z u_3\right)+\partial_\tau u_3\right|\leq C\left(\frac{|\nabla u_3|}{\rho_2}+\frac{|u_3|}{\rho_2^2}\right),\quad \text{in } B_{\rho_2}\times\left[0,\rho_2^2\right),$$

where C depends on λ, Λ and E only.

By Theorem 4.2.6 we have that there exist constants $\eta_2\in(0,1)$ and $C>1$ depending on λ, Λ and E only, such that for every $0<r\leq\rho\leq\eta_2\rho_2$, we have

$$\begin{aligned}
&\int_{B_\rho}v^2(x,0)\,\mathrm{d}x\\
&\quad\leq\frac{C\rho_2^2}{\rho^2}\left(\rho_2^{-2}\int_{B_{\rho_2}\times\left(0,\rho_2^2\right)}v^2\mathrm{d}X\right)^{1-\overline{\theta}_1}\left(\int_{B_r}v^2(x,0)\,\mathrm{d}x\right)^{\overline{\theta}_1},
\end{aligned}\tag{4.200}$$

where

$$\overline{\theta}_1=\frac{1}{C\log\frac{\rho_2}{r}}.$$

Now, denoting by $\psi_3(z,\tau) = \psi_1(\Gamma(\tau)\psi_2(z,\tau),\tau)$, we have for every $\rho \in (0,\rho_2)$ and $\tau \in \left[0,\rho_2^2\right)$

$$Q_{C^{-1}\rho,\varphi}(\tau) \subset (\psi_3(.,\tau))\left(B_\rho^+\right) \subset Q_{C\rho,\varphi}(\tau), \tag{4.201}$$

where $C > 1$, depends on λ, Λ and E only. By (4.200) and (4.201), by using the change of variables $x = \psi_3(z,\tau)$, $t = \tau$ we obtain (4.183). □

COROLLARY 4.2.10 (*Spacelike strong unique continuation at the boundary*). *Let P be a second order parabolic operator as in Theorem 4.2.6. Let $\varphi : \mathbb{R}^{n-1}\times\mathbb{R} \to \mathbb{R}$ be a function satisfying* (4.179) and (4.180). *Assume that $u \in H^{2,1}\left(Q_{R,\varphi}\right)$ satisfies the inequality*

$$|Pu| \le \Lambda\left(\frac{|\nabla u|}{R} + \frac{|u|}{R^2}\right), \quad \textit{in } Q_{R,\varphi}$$

and

$$u(x,t) = 0, \quad \textit{for every } (x,t) \in \Gamma_{R,\varphi}.$$

If for every $k \in \mathbb{N}$ we have

$$\int_{Q_{r,\varphi}(0)} u^2(x,0)\,\mathrm{d}x = O\left(r^{2k}\right), \quad \textit{as } r \to 0,$$

then

$$u(.,0) = 0, \quad \textit{in } Q_{R,\varphi}(0).$$

PROOF. It is the same as the proof of Corollary 4.2.7. □

4.3. *Stability estimates from Cauchy data*

We are given positive numbers R, T, E, λ and Λ such that $\lambda \in (0,1]$. Let us consider the following parabolic operator

$$Lu = \partial_i\left(g^{ij}(x,t)\,\partial_j u\right) - \partial_t u + b_i(x,t)\,\partial_i u + c(x,t)\,u, \tag{4.202}$$

where $\left\{g^{ij}(x,t)\right\}_{i,j=1}^n$ is a real symmetric $n\times n$ matrix. When $\xi \in \mathbb{R}^n$ $(x,t),(y,\tau) \in \mathbb{R}^{n+1}$ assume that

$$\lambda|\xi|^2 \le \sum_{i,j=1}^n g^{ij}(x,t)\,\xi_i\xi_j \le \lambda^{-1}|\xi|^2 \tag{4.203}$$

and

$$\left(\sum_{i,j=1}^{n}\left(g^{ij}(x,t)-g^{ij}(y,\tau)\right)^{2}\right)^{1/2}\leq\frac{\Lambda}{R}\left(|x-y|^{2}+|t-\tau|\right)^{1/2}. \tag{4.204}$$

Concerning $b_i(x,t)$, $i=1,\dots n$ and $c(x,t)$, we assume that

$$R\sum_{i=1}^{n}\|b_i\|_{L^{\infty}(\mathbb{R}^{n+1})}+R^{2}\|c\|_{L^{\infty}(\mathbb{R}^{n+1})}\leq\Lambda. \tag{4.205}$$

Let $\varphi:\mathbb{R}^{n-1}\times\mathbb{R}\to\mathbb{R}$ be a function satisfying the conditions

$$\varphi(0)=|\nabla_{x'}\varphi(0)|=0, \tag{4.206}$$

$$\|\varphi\|_{L^{\infty}(B'_R)}+R\|\nabla_{x'}\varphi\|_{L^{\infty}(B'_R)}+R^{2}\left\|D^{2}\varphi\right\|_{L^{\infty}(B'_R)}\leq ER. \tag{4.207}$$

For any positive numbers ρ and t_0 denote

$$D_{\rho,\varphi}=\left\{x\in B_\rho:x_n>\varphi\left(x'\right)\right\},\qquad D^{t_0}_{\rho,\varphi}=D_{\rho,\varphi}\times(0,t_0),$$
$$\Gamma_{\rho,\varphi}=\left\{x\in B_\rho:x_n=\varphi\left(x'\right)\right\},\qquad \Gamma^{t_0}_{\rho,\varphi}=\Gamma_{\rho,\varphi}\times(0,t_0).$$

If $x=\left(x',\varphi\left(x'\right)\right)$, we denote by $\nu(x)$, or simply by ν, the unit vector of $\mathbb{R}^{n-1}$

$$\nu(x)=\frac{\left(\nabla_{x'}\varphi\left(x'\right),-1\right)}{\sqrt{1+|\nabla_{x'}\varphi(x')|^{2}}}.$$

THEOREM 4.3.1. *Let L be the parabolic operator (4.202). Let $u\in H^{2,1}\left(D^{T}_{R,\varphi}\right)$ be a solution to the equation*

$$Lu=0,\quad \text{in } D^{T}_{R,\varphi}. \tag{4.208}$$

Let

$$\varepsilon:=\|u\|_{H^{\frac{3}{2},\frac{3}{4}}\left(\Gamma^{T}_{R,\varphi}\right)}+R\left\|g^{ij}\partial_j u\nu_i\right\|_{H^{\frac{1}{2},\frac{1}{4}}\left(\Gamma^{T}_{R,\varphi}\right)}. \tag{4.209}$$

There exist constants C, $\eta_3\in(0,1)$, $s_1\in(0,1)$ depending on λ, Λ and E only, such that for every $r\in\left(0,\min\left\{R,\sqrt{T}\right\}\right)$ we have

$$\|u\|_{L^{2}(D_{\eta_3 r,\varphi}\times(r^{2},T))}\leq C\left(\varepsilon^{s_1}\|u\|^{1-s_1}_{L^{2}\left(D^{T}_{R,\varphi}\right)}+\varepsilon\right). \tag{4.210}$$

If, in addition, u satisfies

$$u(x,0) = 0, \quad \textit{in } D_{R,\varphi}, \tag{4.211}$$

then we have, for every $r \in (0, R]$,

$$\|u\|_{L^2\left(D^T_{\eta_3 r,\varphi}\right)} \le C\left(\varepsilon^{s_1}\|u\|^{1-s_1}_{L^2\left(D^T_{R,\varphi}\right)} + \varepsilon\right), \tag{4.212}$$

where C depends λ, Λ *and E only.*

PROOF. By using the extension theorem in Sobolev spaces, [74], there exists $v \in H^{2,1}\left(D^T_{R,\varphi}\right)$ such that

$$v = u, \qquad g^{ij}\partial_j v \nu_i = g^{ij}\partial_j u \nu_i, \quad \text{on } \Gamma^T_{R,\varphi} \tag{4.213}$$

and

$$\|v\|_{H^{2,1}\left(D^T_{R,\varphi}\right)} \le C\varepsilon, \tag{4.214}$$

where C depends on E only. Let $w = u - v$. We have that $w \in H^{2,1}\left(D^T_{R,\varphi}\right)$ and w satisfies

$$Lw = -Lv, \quad \text{in } D^T_{R,\varphi}, \tag{4.215}$$

$$w = 0, \qquad g^{ij}\partial_j w \nu_i = 0, \quad \text{on } \Gamma^T_{R,\varphi}. \tag{4.216}$$

Define

$$\overline{f} = \begin{cases} -Lv, & \text{in } D^T_{R,\varphi}, \\ 0, & \text{in } (B_R \times (0,T)) \setminus D^T_{R,\varphi} \end{cases}$$

and

$$\overline{w} = \begin{cases} w, & \text{in } D^T_{R,\varphi}, \\ 0, & \text{in } (B_R \times (0,T)) \setminus D^T_{R,\varphi}. \end{cases}$$

Since $w = g^{ij}\partial_j w \nu_i = 0$, on $\Gamma^T_{R,\varphi}$, we have $\overline{w} \in H^{2,1}(B_R \times (0,T))$ and $\overline{w}$ satisfies

$$L\overline{w} = \overline{f}, \quad \text{in } B_R \times (0,T). \tag{4.217}$$

Let $z \in H^{2,1}(B_R \times (0,T))$ be the solution to the following initial-boundary value problem

$$\begin{cases} Lz = \overline{f}, & \text{in } B_R \times (0,T), \\ z_{|\partial_p(B_R \times (0,T))} = 0, \end{cases} \tag{4.218}$$

where $\partial_p(B_R \times (0,T)) = (\partial B_R \times (0,T]) \cup (B_R \times \{0\})$ is the parabolic boundary of

$B_R \times (0, T)$. We have by (4.214) and (4.218)

$$\|z\|_{H^{2,1}(B_R\times(0,T))} \leq C \left\| \overline{f} \right\|_{L^2(B_R\times(0,T))} \leq C'\varepsilon, \tag{4.219}$$

where C and C' depend on λ, Λ and E only. Denote by w_1 the function $\overline{w} - z$, by (4.217) and (4.218) we have $w_1 \in H^{2,1}(B_R \times (0, T))$ and

$$Lw_1 = 0, \quad \text{in } B_R \times (0, T). \tag{4.220}$$

Let r be any number belonging to $\left(0, \min\left\{R, \sqrt{T}\right\}\right]$. Let us denote by $r_1 = r \min\left\{\frac{1}{2}, \frac{1}{E}\right\}$, $r_2 = \eta_1 r_1$, $r_3 = \frac{\eta_1 r_1}{4}$, where η_1 is defined as in Theorem 4.2.6. Note that, by (4.206) and (4.207) we have

$$B_{r_3}(-r_3 e_n) \subset B_R \setminus D_{R,\varphi}. \tag{4.221}$$

Since $Lw_1 = 0$, in $B_{r_1}(-r_3e_n) \times (0, T)$, by Theorem 4.2.6 we have, for every $t \in \left[r^2, T\right]$,

$$\begin{aligned} &\|w_1(.,t)\|^2_{L^2\left(B_{r_2}(-r_3e_n)\right)} \\ &\quad \leq C \|w_1(.,t)\|^{2s_1}_{L^2\left(B_{r_3}(-r_3e_n)\right)} \|w_1(.,t)\|^{2(1-s_1)}_{L^2\left(B_{r_1}(-r_3e_n)\times(0,T)\right)}, \end{aligned} \tag{4.222}$$

where C and s_1, $s_1 \in (0, 1)$, depend on λ and Λ only. By integrating both sides of the last inequality with respect to t over $\left(r^2, T\right)$ and by the Hölder inequality, we get

$$\begin{aligned} &\|w_1\|^2_{L^2\left(B_{r_2}(-r_3e_n)\times\left(r^2,T\right)\right)} \\ &\quad \leq C' \|w\|^{2s_1}_{L^2\left(B_{r_3}(-r_3e_n)\times\left(r^2,T\right)\right)} \|w_1\|^{2(1-s_1)}_{L^2\left(B_{r_1}(-r_3e_n)\times(0,T)\right)}, \end{aligned} \tag{4.223}$$

where C depends on λ and Λ only.

By (4.214) and (4.219), using the triangle inequality and recalling the definition of w we have

$$\|w_1\|_{L^2\left(B_{r_1}(-r_3e_n)\times(0,T)\right)} \leq \|u\|_{L^2\left(D^T_{R,\varphi}\right)} + C\varepsilon, \tag{4.224}$$

where C depends on λ, Λ and E only.

By (4.221) we have $w_1 = -z$ in $B_{r_3}(-r_3e_n) \times (0, T)$, so using (4.219) we have

$$\|w_1\|_{L^2\left(B_{r_2}(-r_3e_n)\times\left(r^2,T\right)\right)} \leq C\varepsilon, \tag{4.225}$$

where C depends on λ, Λ and E only.

Since $D_{3r_3,\varphi} \subset B_{r_2}(-r_3e_n) \cap D_{R,\varphi}$, using (4.214) and (4.219) and the triangle inequality we have

$$\begin{aligned} &\|w_1\|_{L^2\left(B_{r_2}(-r_3e_n)\times\left(r^2,T\right)\right)} \geq \|w_1\|_{L^2\left(D_{3r_3,\varphi}\times\left(r^2,T\right)\right)} \\ &\quad = \|u - v - z\|_{L^2\left(D_{3r_3,\varphi}\times\left(r^2,T\right)\right)} \geq \|u\|_{L^2\left(D_{3r_3,\varphi}\times\left(r^2,T\right)\right)} - C\varepsilon, \end{aligned}$$

therefore

$$\|u\|_{L^2\left(D_{3r_3,\varphi}\times\left(r^2,T\right)\right)} \leq \|w_1\|_{L^2\left(B_{r_2}(-r_3e_n)\times\left(r^2,T\right)\right)} + C\varepsilon, \qquad (4.226)$$

where C depends on λ, Λ and E only.

By (4.223)–(4.226) we obtain

$$\|u\|_{L^2\left(D_{3r_3,\varphi}\times\left(r^2,T\right)\right)} \leq C\left(\varepsilon^{s_1}\,\|u\|^{1-s_1}_{L^2\left(D^T_{R,\varphi}\right)} + \varepsilon\right),$$

where C depends on λ, Λ and E only, so (4.210) is proved.

To prove (4.212) we only need some slight variations of the previous proof. Indeed, due to (4.211) we can choose the function v in such a way that, besides (4.224) and (4.225), v satisfies $v\,(.,0) = 0$ on $D_{R,\varphi}$. Therefore $w_1 \in H^{2,1}\left(B_R \times (0,T)\right)$ and

$$w_1\,(.,0) = 0, \quad \text{in } B_R.$$

Now, considering the trivial extension $\widetilde{w}_1$ of w_1 to $B_R \times (-\infty, T)$

$$\widetilde{w}_1 = \begin{cases} w_1, & \text{in } B_R \times (0,T)\,, \\ 0, & \text{in } B_R \times (-\infty, 0)\,, \end{cases}$$

we have that $\widetilde{w}_1 \in H^{2,1}\left(B_R \times (-\infty,T)\right)$ and

$$L\widetilde{w}_1 = 0, \quad \text{in } B_R \times (-\infty, T)\,.$$

Therefore applying Theorem 4.2.6 to the function $\widetilde{w}_1$ we have that (4.222) holds true for every $t \in [0,T]$. Hence we can replace interval $\left(r^2,T\right)$ with interval $(0,T)$ in (4.223). Finally, by definition of w_1, we have

$$\|u\|_{L^2\left(D^T_{3r_3,\varphi}\right)} \leq C\left(\varepsilon^{s_1}\,\|u\|^{1-s_1}_{L^2\left(D^T_{R,\varphi}\right)} + \varepsilon\right),$$

where C depend on λ, Λ and E only, so (4.212) is proved. □

References

[1] V. Adolfsson, L. Escauriaza, $C^{1\alpha}$ *domains and unique continuation at the boundary*, Comm. Pure Appl. Math. **L** (1997) 935–969.

[2] V. Adolfsson, L. Escauriaza, C.E. Kenig, *Convex domains and unique continuation at the boundary*, Rev. Mat. Iberoamericana **11** (3) (1995) 513–525.

[3] S. Agmon, *Unicité et convexité dans les problèmes différentiels*, Les presses de l'Université de Montreal, 1966 .

[4] S. Agmon, L. Nirenberg, *Lower bounds and uniqueness theorems for solutions of differential equations in a Hilbert space*, Comm. Pure Appl. Math. **20** (1967) 207–229.

[5] G. Alessandrini, E. Beretta, E. Rosset, S. Vessella, *Optimal stability for inverse elliptic boundary value problems with unknown boundaries*, Ann. Sc. Norm. Sup. Pisa Cl. Sci. (4) **XXIX** (2000) 755–806.

[6] G. Alessandrini, S. Vessella, *Local behaviour of solutions to parabolic equations*, Comm. Part. Differential Equations **13** (1988) 1041–1058.

[7] G. Alessandrini, S. Vessella, *Remark on the strong unique continuation property for parabolic operators*, Proc. AMS **132** (2) (2004) 499–501.

[8] B.K. Amonov, *The stability of solution of Cauchy problem for a second order equation of parabolic type with data on a time-like manifold*, Funkcional. Anal. i Priložen **6** (3) (1972) 1–9.

[9] B.K. Amonov, S.P. Shishatskii, *An a priori estimate of the solution of the Cauchy problem with data on a time-like surface for a second order parabolic equation and related uniqueness theorems*, Dokl. Akad Nauk SSSR **206** (1) (1972) 11–12.

[10] N. Aronszajn, *Sur l'unicité du prolongement des solutions des équations aux dérivées partielles elliptiques du second ordre*, C. R. Acad. Sci. Paris **242** (1956) 723–725.

[11] N. Aronszajn, A. Krzywicki, J. Szarski, *A unique continuation theorem for exterior differential forms on Riemannian manifolds*, Ark. Matematik **4** (34) (1962) 417–453.

[12] D.G. Aronson, *Non-negative solutions of linear parabolic equations*, Ann. Sc. Norm. Sup. Pisa Cl. Sci. **XXII** (1968) 607–694.

[13] C. Bardos, L. Tartar, *Sur l'unicité rétrograde des équations paraboliques et quelques questions voisines*, Arch. Rat. Mech. Anal. **50** (1973) 10–25.

[14] J.R. Cannon, *A priori estimate for continuation of the solution of the heat equation in the space variable*, Ann. Mat. Pura Appl. (4) **65** (1964) 377–387.

[15] J.R. Cannon, *A Cauchy problem for the heat equation*, Ann. Mat. Pura Appl. (4) **66** (1964) 155–165.

[16] J.R. Cannon, *The One-dimensional Heat Equation*, in: Encyclopedia of Mathematics and its Applications, vol. 83, Addison-Wesley, New York, 1984.

[17] B. Canuto, E. Rosset, S. Vessella, *Quantitative estimates of unique continuation for parabolic equations and inverse initial-boundary value problems with unknown boundaries*, Trans. Amer. Math. Soc. **354** (2) (2002) 491–535.

[18] B. Canuto, E. Rosset, S. Vessella, *A stability result in the localization of cavities in a thermic conducting medium*, ESAIM: Control Optim. Calc. Var. **7** (2002) 521–565.

[19] T. Carleman, *Sur un problème d'unicité pour les systèmes d'équations aux dérivées partielles à deux variables*, Ark. Mat. Astr. Fys. **26B** (17) (1939) 1–9.

[20] X.Y. Chen, *A strong unique continuation theorem for parabolic equations*, Math. Ann. **311** (1996) 603–630.

[21] D. Colton, *The noncharacteristic Cauchy problem for parabolic equations in one space variable*, SIAM J. Math. Anal. **5** (1974) 263–272.

[22] H.O. Cordes, *Über die eindeutige Bestimmtheit der Lösungen elliptischer Differentialgleichungen durch Anfangsvorgaben*, Nachr. Akad. Wiss. Göttingen. Math.-Phys. Kl. IIa. (1956) 239–258.

[23] R. Courant, D. Hilbert, *Methods of Mathematical Physics, Vol. 1*, Wiley Interscience, New York, 1962.

[24] Dè-yuan' Li, *An inequality for the parabolic operator*, Sci. Sinica **12** (1963) 1425–1467.

[25] M. Di Cristo, L. Rondi, S. Vessella, *Stability properties of an inverse parabolic problem with unknown boundaries*, Ann. Mat. Pura Appl. (4) **185** (2) (2006) 223–255.

[26] D. Del Santo, M. Prizzi, *Backward uniqueness for parabolic operators whose coefficients are non-Lipschitz continuous in time*, J. Math. Pures Appl. **84** (2005) 471–491.

[27] D. Del Santo, M. Prizzi, *On the backward uniqueness property for a class of parabolic operators*, in: Bove, Antonio, et al. (Eds.), Phase Space Analysis of Partial Differential Equations, in: Progress in Nonlinear Differential Equations and Their Applications, vol. 69, Birkhäuser, Basel, 2006, pp. 95–105.

[28] D. Del Santo, M. Prizzi, *On the absence of rapidly decaying solutions for parabolic operators whose coefficients are non-Lipschitz continuous in time*, Proc. Amer. Math. Soc. **135** (2) (2007) 383–391.

[29] M.M. Eller, V. Isakov, *Carleman estimates with two large parameters and applications*, Contemp. Math. **268** (2000) 117–136.

[30] L. Escauriaza, F.J. Fernandez, *Unique continuation for parabolic operator*, Ark. Mat. **41** (1) (2003) 35–60.

[31] L. Escauriaza, G. Seregin, V. Šverák, *Backward uniqueness for the heat operator in a half-space*, St. Petersburg Math. J. **15** (1) (2004) 139–148.

[32] L. Escauriaza, S. Vessella, *Optimal three cylinder inequalities for solutions to parabolic equations with Lipschitz leading coefficients*, in: G. Alessandrini, G. Uhlmann (Eds.), Inverse Problems: Theory and Applications, in: Contemporary Mathematics, vol. 333, American Mathematical Society, Providence, RI, 2003, pp. 79–87.

[33] L. Escauriaza, F.J. Fernandez, S. Vessella, *Doubling properties of caloric functions*, Appl. Anal. **85** (1–3) (2006) Special issue dedicated to the memory of Carlo Pucci, editors R. Magnanini and G. Talenti.

[34] L.C. Evans, *Partial Differential Equations*, American Mathematical Society, Providence, 1998.

[35] F.J. Fernandez, *Unique continuation for parabolic operator II*, Comm. Part. Differential Equations **28** (9&10) (2003) 1597–1604.

[36] A. Friedman, *Partial Differential Equation of Parabolic Type*, Prentice Hall, Englewood Cliffs, NJ, 1964.

[37] A.V. Fursikov, O.Yu. Imanuvilov, *Controllability of Parabolic Equations*, in: Lecture Notes, vol. 34, Seoul National University, Seoul Korea, 1996.

[38] N. Garofalo, F.H. Lin, *Monotonicity properties of variational integrals A_p-weights and unique continuation*, Indiana Univ. Math. J. **35** (1986) 245–267.

[39] M. Gevrey, *Sur la nature analytique des solutions des équations aux dérivées partielles*, Primier mémoire. Ann. Sci. École Norm Sup. **35** (3) (1918) 129–190.

[40] J.-M. Ghidaglia, *Some backward uniqueness results*, Nonlinear Analysis TMA **10** (1986) 777–790.

[41] D. Gilbarg, N. Trudinger, *Elliptic Partial Differential Equations of Second Order*, 2nd ed., Springer, Berlin, 1983.

[42] R.Ja. Glagoleva, *Some properties of solutions of a linear second order parabolic equation*, Math. USSR-Sbornik **3** (1) (1967) 41–67.

[43] J. Hadamard, *Lectures on Cauchy's Problem in Linear Partial Differential Equations*, Dover, New York, 1953.

[44] D.N. Hào, *A noncharacteristic Cauchy problem for linear parabolic equations and related inverse problems. I. Solvability*, Inverse Problems **10** (2) (1994) 295–315.

[45] C.D. Hill, *Parabolic equations in one space variable and the non-characteristic Cauchy problem*, Comm. Pure Appl. Math. **20** (1967) 619–633.

[46] E. Holmgren, *Sur l'equation de la propagation de la chaleur*, Ark. Mat. Fys. **14** (1908) 1–11.

[47] L. Hörmander, *Linear Partial Differential Operators*, Springer, New York, 1963.

[48] L. Hörmander, *The Analysis of Linear Partial Differential Operators*, Springer, New York, 1985.

[49] L. Hörmander, *Uniqueness theorems for second order elliptic differential equations*, Comm. Partial Differential Equations **8** (1983) 21–64.

[50] F. John, *Continuous dependence on data for solutions of partial differential equations with a prescribed bound*, Comm. Pure Appl. Math. **XIII** (1960) 551–585.

[51] V. Isakov, *Carleman type estimate in an anisotropic Case and applications*, J. Differential Equations **105** (2) (1993) 217–238.

[52] V. Isakov, *Inverse Problems for Partial Differential Equations*, Springer, New York, 1998.

[53] V. Isakov, *Some inverse problems for the diffusion equation*, Inverse Problems **15** (1999) 3–10.

[54] V. Isakov, *Carleman type estimates and their applications*, in: K. Bingham, Ya.V. Kurylev, E. Sommersalo (Eds.), New Analytic and Geometric Methods in Inverse Problems, in: Lectures given at the EMS Summer School and Conference, Edinburgh, 2000, Springer, 2000.

[55] S. Ito, H. Yamabe, *A unique continuation theorem for solutions of a parabolic differential equation*, J. Math. Soc. Japan **10** (1958) 314–321.

[56] C.E. Kenig, *Some recent quantitative unique continuation theorems*, Semin. Équ. Dériv. Partielles, École Polytech., Palaiseau, 2006.

[57] C.E. Kenig, W. Wang, *A note on boundary unique continuation for harmonic function in non-smmooth domains*, Potential Anal. **8** (1998) 143–147.

[58] M.V. Klibanov, A. Timonov, *Carleman estimates for coefficient inverse problems and numerical applications*, in: Inverse and Ill-Posed Problems Series, VSP, Utrecht, 2004 .

[59] M.V. Klibanov, *Inverse problems and Carleman estimates*, Inverse Problems **8** (1992) 575–596.

[60] H. Koch, D. Tataru, *Carleman estimates and unique continuation for second order parabolic equations with nonsmooth coefficients*, 11 Apr 2007. arXiv: 0704.1349v1[math.AP] .

[61] P. Knabner, S. Vessella, *Stabilization of ill-posed Cauchy problems for parabolic equations*, Ann. Mat. Pura Appl. **149** (1987) 393–409.

[62] I. Kukavica, *Quantitative uniqueness for second-order elliptic operators*, Duke Math. J. **91** (1998) 225–240.

[63] I. Kukavica, N. Kaj, *Unique continuation on the boundary for Dini domains*, Proc. Amer. Math. Soc. **126** (2) (1998) 441–446.

[64] O.A. Ladyzhenskaya, V.A. Solonnikov, N.N. Ural'tseva, *Linear and Quasilinear Equations of Parabolic Type*, in: Transl. Math. Monogr., vol. 23, AMS, Providence, RI, 1968.

[65] E.M. Landis, *Some questions in the qualitative theory of elliptic and parabolic equations*, Amer. Math. Soc. Transl. **2** (20) (1962) 173–238.

[66] E.M. Landis, *A three sphere theorem*, Dokl. Akad. Nauk SSSR **148** (1963) 277–279, Engl. trans. Soviet Math. Dokl. 4 (1963) 76–78.
[67] E.M. Landis, *Some problems of the qualitative theory of second order elliptic equations (case of several independent variables)*, Uspehi Mat. Nauk **18** (1) (1963) 3–62, Engl. trans. Russian Math. Surveys 18 (1963) 1–62.
[68] E.M. Landis, O.A. Oleinik, *Generalized analyticity and some related properties of solutions of elliptic and parabolic equations*, Russian Math. Surveys **29** (1974) 195–212.
[69] M.M. Lavrentiev, V.G. Romanov, S.P. Shishatskii, *Ill-posed Problems of Mathematical Physics and Analysis*, in: Transl. Math. Monogr., vol. 64, AMS, Providence, RI, 1986.
[70] P.D. Lax, *A stability theorem for solutions of abstract differential equations and its application to the study of the local behavior of solutions of elliptic equations*, Comm. Pure Appl. Math. **9** (1956) 747–766.
[71] M. Lees, M.H. Protter, *Unique continuation for parabolic equations*, Duke Math. J. **28** (1961) 369–382.
[72] G.M. Lieberman, *Second Order Parabolic Differential Equations*, World Scientific, London, 1996.
[73] F.H. Lin, *A uniqueness theorem for parabolic equations*, Comm. Pure Appl. Math. **XLIII** (1990) 127–136.
[74] J.-L. Lions, E. Magenes, *Non-homogeneous Boundary Value Problems and Applications I, II*, Springer, New York, 1972.
[75] J.-L. Lions, B. Malgrange, *Sur l'unicité rétrograde dans les problèmes mixtes paraboliques*, Math. Scand. **8** (1960) 277–286.
[76] N. Mandache, *On a counterexample concerning unique continuation for elliptic equations in divergence form*, Math. Phys. Anal. Geom. **1** (1998) 273–292.
[77] K. Miller, *Non-unique continuation for certain ode's in Hilbert space and for uniformly parabolic and elliptic equations in divergence form*, Arch. Rat. Mech. Anal. **54** (1963) 105–117.
[78] L. Nirenberg, *Uniqueness in Cauchy problems for differential operators with constant leading coefficients*, Comm. Pure Appl. Math. **10** (1957) 89–105.
[79] L.E. Payne, *Improved stability estimates for classes of illposed Cauchy problems*, Appl. Anal. **19** (1985) 63–74.
[80] A. Pliš, *On non-uniqueness in Cauchy problem for an elliptic second order differential equation*, Bull. Acad. Polon. Sci. Math Astronom. Phys. **11** (1963) 95–100.
[81] C.C. Poon, *Unique continuation for parabolic equations*, Comm. Partial Differential Equations **21** (1996) 521–539.
[82] M.H. Protter, *Properties of solutions of parabolic equations and inequalities*, Canad. J. Math. **13** (1961) 331–345.
[83] C. Pucci, *Alcune limitazioni per le soluzioni di equazioni paraboliche*, Ann. Mat. Pura Appl. **IV** (48) (1959) 161–172.
[84] C. Pucci, *Nuove ricerche sul problema di Cauchy*, Mem. Accad. Sci. Torino. Cl. Sci. Fis. Mat. Nat. **1** (3) (1953) 45–67.
[85] J.-C. Saut, B. Scheurer, *Unique continuation for some evolution equations*, J. Differential Equations **66** (1987) 118–139.
[86] S.P. Shishatskii, A. Asanov, E.R. Atamanov, *Uniqueness Problems for Degenerating Equations and Nonclassical Problems*, in: Inverse and Ill-Posed Problems Series, VSP, Utrecht, 2001.
[87] C.D. Sogge, *A unique continuation theorem for second order parabolic differential operators*, Ark. Mat. **28** (1) (1990) 159–182.
[88] D. Tataru, *Unique continuation problems for partial differential equations*, in: Geometric methods in inverse problems and PDE conro, in: IMA Vol. Math. Appl., vol. 137, Spriner, New York, 2004, pp. 239–255.
[89] F. Trèves, *Linear Partial Differential Equations*, Gordon and Breach Science Publisher, New York, 1970.
[90] A. Tychonoff, *Théoreme d'unicité pour l'équation de la chaleur*, Math. Sbornik **42** (1935) 199–215.
[91] A.A. Varin, *Three-cylinder theorem for certain class of semilinear parabolic equations*, Mat. Zametki **51** (1) (1992) 32–41.
[92] S. Vessella, *Three cylinder inequalities and unique continuation properties for parabolic equations*, Rend. Mat. Ac. Lincei, 9 **13** (2) (2002) 107–120.
[93] S. Vessella, *Carleman estimates optimal three cylinder inequality and unique continuation properties for solutions to parabolic equations*, Comm. Partial Differential Equations **28** (3&4) (2003) 627–637.
[94] S. Vessella, *Quantitative estimates of unique continuation for parabolic equations determination of unknown time-varying boundaries and optimal stability estimates*, preprint .
[95] C. Zuily, *Uniqueness and Non-uniqueness in the Cauchy Problem*, Birkhäuser, Boston, 1983.

Author Index

Roman numbers refer to pages on which the author (or his/her work) is mentioned. Italic numbers refer to reference pages. Numbers between brackets are the reference numbers. No distinction is made between the first author and co-author(s).

Subject Index

A

B

C

F

L

M

R

S

T