NTNU – Trondheim
Norwegian University of
Science and Technology

Relaxation Systems with Applications to Two-Phase Flow

Susanne Solem

Master of Science in Mathematics
Submission date: May 2014
Supervisor: Helge Holden, MATH
Co-supervisor: Peder Kristian Aursand, IMF
Tore Flåtten, SINTEF Energi

Norwegian University of Science and Technology
Department of Mathematical Sciences

Abstract

Relaxation systems are widely studied and much used to describe nonequilibrium phenomena, occurring in, for example, two-phase flow. In this thesis we will therefore consider relaxation systems in one space dimension and get some insight into their applications to two-phase flow. Two main topics will be considered: hyperbolic constant-coefficient relaxation systems and a specific two-phase model. Entropy conditions are also studied. All three topics are connected with relaxation processes.

The first part consists of a study of the transitional wave-dynamics of strictly hyperbolic constant-coefficient relaxation systems with stable rank one relaxation matrices. By realizing that the eigenvalue polynomial of such a system can be written as a convex sum of the two eigenvalue polynomials of the corresponding formal limiting systems, we show that the system is stable if and only if it fulfills the interlacing property known as the subcharacteristic condition. Further, if the system is stable, it is shown that the transitional wave-velocities can never exceed the velocities of the corresponding homogeneous system. The results are applied to a two-phase model.

Mathematical entropy is studied in connection with conservative and nonconservative relaxation systems. Beneficial properties, such as symmetry and the fulfillment of the subcharacteristic condition, follow directly from the existence of a convex entropy for conservative systems. This does not hold in general for nonconservative systems.

A two-phase model with a well-reservoir interaction term and a viscous term is studied. The estimates that exist for the full model are relaxation parameter dependent. An existence result for the reduced model, the formal limit model as the relaxation time tends to 0, does therefore not follow directly from the existence result for the full model. A new existence result for the reduced model is therefore achieved in a similar way to that of the full model. It relies on the assumption that specific parameters and initial conditions are small enough. The result also ensures that both phases will exist at any spatial point for any finite time.

Abstract

Relaksasjonssystemer blir mye studert og brukes ofte for å beskrive ulike ikke-likevektsfenomen, som oppstår i, for eksempel, to-fasestrømning. I denne tesen vil derfor relaksasjonssystemer i en romlig dimensjon, og bruk av slike systemer i sammenheng med to-fase strømning, studeres. Tesen er konsentrert rundt to hovedtemaer: hyperbolske konstant-koeffisient relaksasjonssystemer og en spesifikk to-fasemodell. Entropibetingelser blir også studert. Disse tre temaene er alle knyttet til relaksasjonsprosesser.

Den første delen omhandler overgangsbølgedynamikken for strengt hyperbolske relaksasjonssystemer med konstante koeffisienter og med en stabil relaksasjonsmatrise av rang en. Oppdagelsen av at det er mulig å skrive egenverdipolynomet til et slikt system som en konveks sum av egenverdipolynomene til de to formelle grensesystemene, gjør det mulig å vise at systemet er stabilt hvis og bare hvis en flettingsbetingelse, kjent som den subkarakteristiske betingelsen, er oppfylt. Det blir også vist at dersom systemet er stabilt, så vil overgangsbølgehastighetene aldri overstige hastighetene til det tilhørende homogene systemet. Disse resultatene blir anvendt på en to-fasemodell.

Matematisk entropi blir studert i sammenheng med ikke-konservative og konservative relaksasjonssystem. Fordelaktige egenskaper som symmetri og garantien for at den subkarakteristiske betingelsen er oppfylt, følger direkte fra eksistensen av en konveks entropi for konservative system. Dette holder ikke generelt for ikke-konservative system.

En to-fasemodell med et brønn-reservoar interaksjonsledd og et viskøst ledd blir studert. De eksisterende estimatene for den fulle modellen avhenger av relaksasjonsparameteren, og dermed følger ikke eksistens for den reduserte modellen, den formelle grensemodellen når relaksasjonsparameteren går mot 0, direkte fra eksistensresultatene for den fulle modellen. Et eget eksistensresultat for den reduserte modellen blir derfor funnet på en lignende måte som for den fulle modellen. Det avhenger av at enkelte parametre og initial data er små nok. Resultatet viser også at begge faser vil eksistere i ethvert punkt for enhver endelig tid.

Preface

This master's thesis contains the fruits of my labour at the Institute of Mathematical Sciences at NTNU during the academic year of 2013/2014.

Though the thesis is slightly divided into three parts, I am content with the results and satisfied with the knowledge that I've obtained. In particular, I am glad to have had the opportunity to learn certain tools in nonlinear analysis, which I knew very little about in advance. The opportunity to pursue topics of interest during the last year has been invaluable, especially each time my patience was challenged. I believe this has helped to cause some interesting, and maybe new, results.

A great thanks to my co-supervisors Tore Flåtten and Peder Aursand. This thesis would not have existed without their great interest and continuous support. Also, a great thanks to Steinar Evje for suggesting to work with the well-reservoir model and for voluntarily acting as a co-supervisor and giving valuable feedback. I would like to thank the Institute of Mathematical Sciences at NTNU for giving me the opportunity to present parts of my thesis at the SIAM PDE13 conference in Florida, December 2013. Also, a thanks to my supervisor Professor Helge Holden for always having an open door. And lastly, a special thanks to Bjørnar Skaug Karlsen for moral support throughout the year.

Susanne Solem
Trondheim, June 2014

Contents

Abstract i

Abstrakt iii

Preface v

1 Introduction 1
- 1.1 Relaxation systems . . . 1
- 1.2 Two-phase flow models . . . 3
- 1.3 Objective . . . 4
- 1.4 Main Results . . . 5
- 1.5 Outline . . . 6

2 Basic concepts 7
- 2.1 Borel measure . . . 7
- 2.2 Properties for functions . . . 8
- 2.3 Various spaces . . . 9
- 2.4 Inequalities . . . 10
- 2.5 The Legendre transform . . . 11
- 2.6 Mollification . . . 12

3 Hyperbolic relaxation systems 13
- 3.1 The limits . . . 14
- 3.2 The subcharacteristic condition . . . 15
- 3.3 Linear analysis . . . 16
  - 3.3.1 Linearization . . . 16
  - 3.3.2 Plane-wave solutions . . . 16

3.4 Strictly nonlinear hyperbolic conservation laws 18
3.4.1 Weak solutions 18
3.4.2 Non-uniqueness 19
3.4.3 Entropy 20
3.4.4 Vanishing viscosity 21
3.5 Summary 22

4 Wave-dynamics for linear hyperbolic relaxation systems 25
4.1 The characteristic equation as a convex sum 26
4.1.1 Structure of the relaxation matrix 26
4.1.2 The equilibrium system 27
4.1.3 A convexity lemma 28
4.2 Stability and the subcharacteristic condition 30
4.3 A maximum principle 39
4.4 Summary 42

5 Examples: Two linearized 3 $\times$ 3 relaxation systems 43
5.1 Linear stability and interlacing 44
5.2 A critical region 47
5.3 A solution 47
5.4 Summary 52

6 Entropies, conservative and nonconservative systems 53
6.1 Constant-coefficient systems 55
6.2 Nonlinear conservative systems 59
6.2.1 Symmetric form 59
6.2.2 The vanishing viscosity approach 61
6.2.3 The subcharacteristic condition 62
6.3 Nonconservative systems 64
6.3.1 Defining a weak solution 65
6.3.2 Modeling hyperbolic relaxation systems 66
6.3.3 Symmetric form 67
6.3.4 Nonconservative entropy conditions 68
6.3.5 The subcharacteristic condition 69
6.4 Summary 70

7 A two-phase model with well-reservoir interaction 73
7.1 The model 74
7.1.1 Existence of solutions 75

7.1.2 The equilibrium limit 76
7.2 A relaxation system without viscosity 77
7.3 Analysis of the reduced system 79
7.3.1 Variable transformations 82
7.3.2 The estimates 84
7.3.3 Closing the a priori assumptions 94
7.4 A discussion concerning the estimates 108
7.4.1 Existence 110
7.4.2 The a priori assumption 113
7.4.3 Difficult terms 113
7.4.4 Restrictions 114
7.5 Summary 115

8 Conclusions 117
8.1 Linear relaxation systems of rank one 117
8.2 Entropy conditions 118
8.3 The well-reservoir model 119
8.4 Suggestions for further work 119
8.4.1 Linear relaxation systems 120
8.4.2 The well-reservoir model 120

A Article: Wave dynamics of linear hyperbolic relaxation systems 129

1 Introduction

Relaxation systems, in the most general way presented in this thesis, are systems in one space dimension taking the form

$$\partial_t \boldsymbol{U} + \boldsymbol{P}(\boldsymbol{U})\partial_x \boldsymbol{U} = \frac{1}{\varepsilon}\boldsymbol{Q}(\boldsymbol{U}) + \boldsymbol{S}(\partial_{xx}\boldsymbol{U}, \partial_{xxx}\boldsymbol{U}, ...), \tag{1.1}$$

where $\boldsymbol{U}$ is a solution vector in some convex state space $G \in \mathbb{R}^N$ and where $\boldsymbol{P}(\boldsymbol{U})$ is a real $N \times N$ matrix and $\boldsymbol{Q}(\boldsymbol{U})$ a real N-vector. Both are smooth in $\boldsymbol{U}$. The parameter $\varepsilon \in (0, \infty)$ is some characteristic relaxation time for the system. The variable $\boldsymbol{S}$ consists of terms of higher order derivatives. Such systems are widely used in the applied sciences to model different kinds of nonequilibrium phenomena. These phenomena can occur in, for example, two-phase flow [49, 36, 46, 18, 19], traffic flow [3] and elastoplastic materials [24].

1.1 Relaxation systems

For the most part, we will consider first-order relaxation systems in one space dimension, i.e. systems in the form

$$\partial_t \boldsymbol{U} + \boldsymbol{P}(\boldsymbol{U})\partial_x \boldsymbol{U} = \frac{1}{\varepsilon}\boldsymbol{Q}(\boldsymbol{U}). \tag{1.2}$$

In many cases $\boldsymbol{A}(\boldsymbol{U})$ is equal to the Jacobian of some flux term. The system is then conservative. If not, the system is nonconservative. The formal limits of (1.2) as $\varepsilon \to \infty$ or $\varepsilon \to 0$, are, respectively, the homogeneous system with $\boldsymbol{U} \in \mathbb{R}^N$,

$$\partial_t \boldsymbol{U} + \boldsymbol{P}(\boldsymbol{U})\partial_x \boldsymbol{U} = 0, \tag{1.3}$$

and the equilibrium system for some reduced variable $\boldsymbol{u} \in \mathbb{R}^n$, where $n < N$,

$$\partial_t \boldsymbol{u} + \boldsymbol{p}(\boldsymbol{u})\partial_x \boldsymbol{u} = 0. \tag{1.4}$$

More precise definitions will be presented later. The formal limits of relaxation systems in the form (1.2) have been widely studied. Well-posedness in the equilibrium limit, $\varepsilon \to 0$, has especially been a subject of interest. Chen [11] gives an overview of the existing literature concerning these relaxation limits. We mention some of the results below.

The well-posedness of hyperbolic conservative relaxation systems as the relaxation time $\varepsilon \to 0$ depends on the stability of the systems. Bouchut [8] gives a short overview of the relations between existing stability conditions for nonlinear relaxation systems. Chen et al. [22] propose a convex entropy. This convex entropy ensures that the equilibrium limit is hyperbolic and also endowed with a convex entropy. It also implies that the wave-velocities of the equilibrium limit are interlaced between the wave-velocities of the corresponding homogeneous system. This interlacing property is known as the *subcharacteristic condition*. Bouchut [7] proposes a reduced stability condition, weaker than the entropy condition, which also ensures that the subcharacteristic condition is fulfilled.

Well-posedness in the limit $\varepsilon \to 0$ for constant-coefficient hyperbolic relaxation systems is shown to be equivalent to the stability of the system, when assuming that the relaxation matrix satisfies a nonoscillation condition, by Lorenz and Schroll [34]. Stability conditions for constant-coefficient systems are also studied by Yong [54, 53]. These constant-coefficient conditions, amongst some stronger conditions, are necessary for the solutions of both conservative and nonconservative nonlinear relaxation systems to have well-behaved limits.

The connection between entropy and well-posedness for constant-coefficient systems is also studied by Lorenz and Schroll [35]. It is shown that the existence of a convex entropy is sufficient, but not necessary, for well-posedness of these systems.

Vanishing viscosity is closely related to the convex entropy condition for conservative systems. Much of the established theory concerning the vanishing viscosity approach for general homogeneous hyperbolic systems is summarized by Bressan [9]. The method is not restricted to homogeneous systems. For example, Christoforou [14] proves global existence for systems of hyperbolic balance laws, where the source terms satisfy a suitable dissipative property, by using the vanishing viscosity approach.

Hyperbolic relaxation systems are also used to approximate systems of hyperbolic conservation laws. Tzavaras [50] builds a framework for this purpose. This framework is further developed to suit systems of balance laws by Miroshnikov and Trivisa [39].

The 2×2 relaxation systems are well-studied as they contain some information about relaxation systems in general, yet they are simpler to handle than larger systems. The typical 2×2 relaxation system was first introduced by Jin and Xin [28]. Liu [33] studied the stability of waves for the 2×2 nonlinear system. Chen et al. [22] proved strong convergence of solutions as the relaxation time $\varepsilon \to 0$. They also showed that the strict subcharacteristic condition, i.e. strict interlacing of the wave-velocities, together with the existence of a convex entropy for the limiting conservation law in the limit $\varepsilon \to 0$, is equivalent to the existence of a convex entropy for the full 2×2-system. Chen and Liu [13] proved convergence of solutions for the 2×2-system by a compensated compactness argument.

Aursand and Flåtten [2] studied the constant-coefficient 2×2 relaxation system. They gave a complete description of the wave-dynamics for this system. They identified a critical point. It was also noted that the stability of the system is equivalent to the subcharacteristic condition being satisfied.

Although some of the results above are applicable to nonconservative systems as well, these systems are in general quite different from conservative ones. Even the definition of weak solutions, proposed by Dal Maso et al. [37], for nonconservative systems differs from the one for conservative systems.

1.2 Two-phase flow models

We have already noted that relaxation systems are used to model two-phase flows. One example is the two-phase flow model with phase-transition in Solem et al. [46]. The model has a rank one relaxation term. The wave-dynamics of this model is studied with the help of linear analysis in the same way as the 2×2-system in [2] is. One of the main results is the finding of a critical region for some relaxation times, corresponding to the critical point for the 2×2-system. In this region the linearized relaxation system has zero wave-velocity. For this specific system it is also proven that the stability of the linearized system is equivalent to the subcharacteristic condition. Further, it is shown that stability of the linear system implies that the system satisfies the subcharacteristic condition.

Similar results for general $N \times N$-systems seem to be non-existent.

Another example is the nonlinear well-reservoir model studied by Evje [18]. This model is of the form (1.1). It is a relaxation system with a rank one relaxation term and a viscosity term. A similar model is considered by Evje and Karlsen [19]. The well-reservoir model in [18] describes the interaction between a well and a reservoir containing gas with mass n and liquid with mass m. By obtaining pointwise control on the sizes of n and m, such that there exists a finite amount of both gas and liquid at any spatial point for any finite time, existence of solutions is shown for this model.

Unfortunately, the estimates for the full well-reservoir model are highly dependent on the relaxation parameter of the model, meaning that the estimates blow up as the relaxation parameter goes to 0. Thus, we cannot directly derive estimates for the reduced model, the model in the formal limit $\varepsilon \to 0$, from the estimates for the full model. It is also difficult to say anything certain about the relaxation process of the system.

1.3 Objective

The goal of this thesis is to study relaxation systems in general and get some insight into their applications to two-phase flow. This will be done by looking at both linear and nonlinear systems. The focus will be on two main subjects: constant-coefficient relaxation systems with rank one relaxation matrices and the already mentioned two-phase well-reservoir model. As a side-project, we will study entropy conditions for relaxation systems.

As general results seems to be lacking, we wish to generalize the stability results in [2] and [46] to all hyperbolic constant-coefficient relaxation systems with rank one relaxation matrices. We will then apply the generalized results to the specific rank one relaxation system in [46] to verify their validity.

Mathematical entropy is a widely used concept for hyperbolic systems. We will therefore study mathematical entropy in the literature to get a better understanding of the concept. The connection between the subcharacteristic condition and entropy will specifically be studied. And, as conservative and nonconservative hyperbolic systems seem to differ a lot, we will look at the differences concerning mathematical entropy for these systems.

We also aim to get familiar with some of the tools used in the analysis of nonlinear relaxation systems. This will be done by analyzing a specific relaxation model, the two-phase well-reservoir model in [18]. As the estimates for the full model in [18] are relaxation parameter dependent, new estimates will be developed for the reduced model to ensure existence of solutions. We wish to derive an existence result with the same pointwise control as for the full model. We will also try to extract some results concerning the stability of the relaxation process for the well-reservoir model.

1.4 Main Results

The main results in this thesis concerns constant-coefficients relaxation systems with rank one relaxation matrices and the well-reservoir model in [18]. In their completeness, these results are, to the author's knowledge, new contributions. We shortly summarize the results below.

Assuming that the relaxation matrix is stable and of rank one, we have proved the following for strictly hyperbolic $N \times N$ constant-coefficient relaxation systems:

- The eigenvalue polynomial of the full relaxation system can be written as a convex sum of the eigenvalue polynomials of the homogeneous system and the equilibrium system.
- The relaxation system is stable if and only if an interlacing condition, the subcharacteristic condition, is satisfied.
- If the relaxation system is stable, the wave-velocities of the relaxation system for any relaxation time ε will never exceed the wave-velocities of the corresponding homogeneous system.

For precise statements, we refer to Lemma 4.3, Proposition 4.14 and Proposition 4.16.

An existence result is obtained for the reduced well-reservoir model. Under some suitable smallness assumptions on the initial data, we prove that there exists a sufficiently small $M > 0$ such that the following hold:

- The size of the spatial first derivatives of the variables in the system are controlled by M.
- The solutions exist in suitable Sobolev spaces.

- Pointwise upper and lower bounds on the gas mass n and the liquid mass m exist, ensuring that there will exist both gas and liquid at any spatial point x for any finite time t.

The precisely formulated results are found in Theorem 7.2.

1.5 Outline

In the upcoming chapter, Chapter 2, we introduce some basic mathematical concepts that will be used in the thesis.

Chapter 3 concerns hyperbolic relaxation systems and we introduce some useful concepts and methods that will be used later in the thesis. The few concepts mentioned in the introduction above are more precisely defined in this chapter.

In Chapter 4 we prove the results concerning strictly hyperbolic constant-coefficient relaxation systems with rank one relaxation matrices.

We look at a two-phase model and a specific 3×3-system in Chapter 5 to illustrate the results from the previous chapter.

In Chapter 6 we take a look at mathematical entropies and study the properties of conservative and nonconservative systems with such entropies.

The two-phase well-reservoir model is studied in Chapter 7. This model is not endowed with an entropy that is globally strictly convex. We develop estimates that ensure existence of solutions for the corresponding equilibrium model. The relaxation process of the model is also briefly studied in this chapter.

We sum up all the results, conclude and suggest some topics for further work in Chapter 8.

Lastly, the content of Chapter 4 is presented as a pre-print article in Appendix A.

2 Basic concepts

In this short chapter, we present some basic mathematical concepts that will be used, or mentioned, later in the thesis. This chapter serves only as a look-up chapter for the reader if needed.

First, we mention that vectors and matrices are denoted in bold letters, for example $\boldsymbol{x}$. Also, $D\boldsymbol{f}(\boldsymbol{x})$ is the Jacobian of $\boldsymbol{f}(\boldsymbol{x})$ w.r.t. $\boldsymbol{x}$. The notations $\partial_{\boldsymbol{x}}\boldsymbol{f}(\boldsymbol{x})$ and $\boldsymbol{f}(\boldsymbol{x})_{\boldsymbol{x}}$ for the Jacobian are also used when they are more suitable.

2.1 Borel measure

To define the Borel measure, we first need to define σ-algebras and measures.

Definition 2.1 (σ-algebra) *A collection $\mathcal{B}$ of subsets of a set Ω is a σ-algebra of sets if the following three conditions hold.*

i) $\emptyset \in \mathcal{B}$.

ii) If $A \in \mathcal{B}$ then $A^c \in \mathcal{B}$.

iii) Let $\{A_n\}_{n=1}^{\infty}$ be a countable number of sets in $\mathcal{B}$ then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{B}$.

Definition 2.2 (Measure) *A function $\mu : \mathcal{B} \to [0, \infty]$ is a measure if it fulfills the following conditions.*

i) For all $B \in \mathcal{B}$ $\mu(B) \geq 0$.

ii) $\mu(\emptyset) = 0$.

iii) Let $\{A_n\}_{n=1}^{\infty}$ *be a countable number of disjoint sets, i.e.* $A_i \bigcap A_j = 0$ *when* $i \neq j$. *Then* $\mu\left(\bigcup\limits_{n=1}^{\infty} A_n\right) = \sum\limits_{n=1}^{\infty} \mu(A_n)$.

The Borel σ-algebra for the set Ω is the smallest σ-algebra generated by all open sets of Ω. A Borel measure is then a measure defined on that Borel σ-algebra.

2.2 Properties for functions

We state definitions of different properties related to functions.

Definition 2.3 (Compact support) *Functions of compact support are functions that are zero outside of some compact set.*

Theorem 2.4 (The Heine–Borel theorem) *Any subset of* $\mathbb{R}^n$ *is compact if and only if it is closed and bounded.*

Hence, from the Heine–Borel theorem, a function that is defined on $\mathbb{R}^n$ has compact support if and only if it is nonzero on a closed and bounded set $\mathcal{B}$ such that $\mathcal{B} \subset \mathbb{R}^n$. See Munkres [41, p. 173] for a proof of the Heine–Borel theorem.

Definition 2.5 (Bounded variation) *A function* $f(x) : [a, b] \to \mathbb{R}$ *is of bounded variation if for any partition* $a = x_0 < x_1 < ... < x_{n-1} < x_n = b$ *for an interval* $[a, b]$ *there exists an* M *such that*

$$\sum_{i=1}^{n} |f(x_i) - f(x_{i-1})| \leq M. \tag{2.1}$$

A function $\boldsymbol{f}(x) : [a, b] \to \mathbb{R}^n$ is of bounded variation if each component $f_i(x)$, for $i = 1, \ldots, n$, is of bounded variation.

Definition 2.6 (Measurable functions) *Let* $\mathcal{M}$ *and* $\mathcal{N}$ *be* σ*-algebras generated by subsets of* X *and* Y *respectively. A function* $f(x) : X \to Y$ *is then measurable w.r.t* $\mathcal{M}$ *and* $\mathcal{N}$ *if for any* $E \in \mathcal{N}$

$$f^{-1}(E) := \{x \in X : f(x) \in E\} \in \mathcal{M}. \tag{2.2}$$

Equicontinuity is essential in the Arzelà-Ascoli theorem [16, Ch. 19].

Definition 2.7 (Equicontinuity) *Let d be a distance function. A family of functions is equicontinuous at a point $x_0 \in \Omega$ if for every $\varepsilon > 0$, there exists a $\delta > 0$ such that $d(f(x_0), f(x)) < \varepsilon$ for all functions f in the family and for all x such that $d(x_0, x) < \delta$. The family of functions is equicontinuous if it is equicontinuous at each point of Ω.*

2.3 Various spaces

Solutions of partial differential equations often find themselves in Sobolev spaces. Let us first define the Lebesgue spaces, which are needed to define the Sobolev spaces. Let $\Omega \subset \mathbb{R}$.

Definition 2.8 (Lebesgue space) *A Lebesgue space $L^n(\Omega)$ is a space consisting of functions $f(x)$, defined on Ω, such that*

$$\|f\|_{L^n(\Omega)} = \sqrt[n]{\int_\Omega |f(x)|^n dx} < \infty, \tag{2.3}$$

where the left hand side is the L^n-norm w.r.t. Ω.

We can now define the Sobolev spaces.

Definition 2.9 (Sobolev space) *A Sobolev space $W^{k,n}(\Omega)$ is the space of all functions $f(x)$ in $L^n(\Omega)$ such that all the weak partial derivatives up to order k of $f(x)$ are also in $L^n(\Omega)$:*

$$W^{k,n}(\Omega) = \left[f \in L^n(\Omega) : \quad f^{(k)} \in L^n(\Omega) \right]. \tag{2.4}$$

As a special case, $W^{k,2}$ is usually denoted as H^k. The Sobolev-norm corresponding to the space $W^{k,n}(\Omega)$ is

$$\|f\|_{W^{k,n}(\Omega)} = \sum_{i=0}^{k} \|f^{(k)}\|_{L^n(\Omega)}. \tag{2.5}$$

We also need to define the Hölder spaces. Functions which are in some Hölder space, are continuous, making these functions nice to work with.

Definition 2.10 (Hölder space) *The Hölder space $C^\alpha(\Omega)$, with $0 < \alpha \leq 1$, is the space of all Hölder continuous functions $f(x)$. A function is Hölder continuous in a point x if it satisfies*

$$|f(x) - f(y)| \leq C|x - y|^\alpha \tag{2.6}$$

for all $y \in \Omega$ s.t. $y \neq x$ and for some constant C.

From (2.6), we can easily see that Hölder continuous functions are continuous. If (2.6) is true for all $x, y \in \Omega$, where Ω is a space, we say that $f(x)$ is uniformly Hölder continuous. The Hölder norm is

$$\|f\|_{C^\alpha(\Omega)} = \sup_{x\in\bar{\Omega}} |f(x)| + \sup_{x,y\in\Omega} \frac{|f(x) - f(y)|}{|x - y|^\alpha}. \tag{2.7}$$

We further denote $L^n(K, B)$ as the space of all measurable functions $f(x, t)$ from the space K to the space B. Let $D_T \subset \Omega \times [0, T]$. Then $C^{\alpha,\alpha/2}(D_T)$ is the space of functions, defined on the domain D_T, which are uniformly Hölder continuous with exponent α in x and exponent $\alpha/2$ in t.

2.4 Inequalities

We state some important inequalities which are used in the thesis to derive various estimates. The inequalities are well-known and, hence, the proofs are easily obtainable from most textbooks on inequalities used in mathematical analysis. See for example Garling [23] or Evans [17].

Hölder's inequality

For all measurable functions $f(x)$ and $g(x)$ defined on $\mathbb{R}^n$ for some n we have

$$\|fg\|_1 \leq \|f\|_p \|g\|_q, \quad \text{where} \quad \frac{1}{p} + \frac{1}{q} = 1. \tag{2.8}$$

A special case of this inequality is the Cauchy inequality where $p = q = 2$.

Grönwall's inequality

We state the standard Grönwall inequality. Let $f(x)$ be a differentiable function defined on an interval I. Let $g(x)$ be a real-valued continuous function. Then if

$$f'(x) \leq g(x)f(x),$$

we have

$$f(x) \leq f(x_0) \exp\left(\int\limits_{x_0}^{x} g(s)ds \right). \tag{2.9}$$

A slightly different version of Grönwall's inequality is as follows. If $f(x)$ satisfies

$$f(x) \leq K + \int_0^x g(x)f(x)dx,$$

where $K > 0$, then

$$f(x) \leq K \exp\left(\int_0^x g(s)ds \right). \tag{2.10}$$

Jensen's inequality
For a real convex function $f(x)$, we have

$$f\left(\frac{\sum_i a_i x_i}{\sum_i a_i} \right) \leq \frac{\sum_i a_i f(x_i)}{\sum a_i}. \tag{2.11}$$

A Sobolev inequality
On a space $\Omega \in \mathbb{R}^n$, we have

$$\sup_{x\in\Omega} |f(x)| \leq \|f\|_{W^{1,1}(\Omega)}, \tag{2.12}$$

where the Sobolev norm $\|\cdot\|_{W^{1,1}(\Omega)}$ is as defined above.

2.5 The Legendre transform

The Legendre transform is useful when dealing with smooth convex functions. Let $\Omega \in \mathbb{R}^k$ and let $f(\boldsymbol{x}) : \Omega \to \mathbb{R}$ be a strictly convex smooth function. Then the Legendre transform, or the Legendre dual, g is defined as

$$g(\boldsymbol{s}) = \boldsymbol{s} \cdot \boldsymbol{x}(\boldsymbol{s}) - f(\boldsymbol{x}(\boldsymbol{s})), \tag{2.13}$$

where $\boldsymbol{s}(\boldsymbol{x}) = Df(\boldsymbol{x})$.

With $f(\boldsymbol{x})$ strictly convex, $\boldsymbol{s}(\boldsymbol{x})$ is strictly increasing. Since $f(\boldsymbol{x})$ is smooth, $D^2 f(\boldsymbol{x})$ exists at any point $\boldsymbol{x}$ such that there is a one-to-one relation between $\boldsymbol{s}$ and $\boldsymbol{x}$. So, for any $\boldsymbol{s}$ we can find $\boldsymbol{x}(\boldsymbol{s})$. We can then insert $\boldsymbol{x}(\boldsymbol{s})$ into $f(\boldsymbol{x})$ to get f as a function of $\boldsymbol{s}$. With the representation (2.13) we also have $Dg(\boldsymbol{s}) = \boldsymbol{x}$.

A nice property of the Legendre transform of $f(x)$ is that it is also strictly convex when $f(x)$ is:

$$D^2 g(\boldsymbol{s}) = D_{\boldsymbol{s}} \boldsymbol{x}(\boldsymbol{s}) = (D^2 f(\boldsymbol{x}))^{-1} > 0.$$

To learn more about the Legendre transform, see for example Zia et al. [59].

2.6 Mollification

Mollification makes it possible to approximate functions in Sobolev spaces by continuous functions. We state the definition and mention a few useful properties.

Definition 2.11 (The Friedrichs mollifier) *A Friedrichs mollifier is a function $\phi(x)$ in $C_0^\infty(\mathbb{R}^N)$ such that $\phi(x) \in [0, 1]$,*

$$\int_{\mathbb{R}^N} \phi(x) dx = 1,$$

and supp $\phi(x) \subset [-1, 1]^N$.

Let $\phi_\delta(x) = 1/(\delta)^N \phi(x/\delta)$. The mollifier is used to mollify a function $f(x)$:

$$f^\delta(x) = (f * \phi_\delta)(x) = \int_{\mathbb{R}^N} f(x - y)\phi_\delta(y) dy. \tag{2.14}$$

It is well known that $f^\delta(x) \in C^\infty(R^N)$. Let $\Omega \subset R^N$. If $f \in L^q(\Omega)$, $f^\delta(x) \to f(x)$ in $L^q(\Omega)$ for $1 \leq q < \infty$. Also, if $f(x)$ is continuous and bounded, the convergence is uniform. The Friedrich's mollifier is described in almost any textbook in nonlinear analysis, see for example [16, 17].

3 Hyperbolic relaxation systems

In this chapter, we introduce some useful concepts for hyperbolic relaxation systems and conservation laws. We restrict ourselves to the theory that is of importance to the completeness of this thesis.

A conservative relaxation system in one space dimension can be written in the general form [22]:

$$\partial_t \boldsymbol{U} + \partial_x \boldsymbol{F}(\boldsymbol{U}) = \frac{1}{\varepsilon}\boldsymbol{Q}(\boldsymbol{U}), \tag{3.1}$$

where we have the solution vector $\boldsymbol{U} = \boldsymbol{U}(x,t) \in G \subseteq \mathbb{R}^N$ for some state space G. The flux of the system is denoted $\boldsymbol{F}(\boldsymbol{U})$ and $\boldsymbol{Q}(\boldsymbol{U})$ is the source, or relaxation, term. Both the flux term and the relaxation term are assumed to be real and smooth enough in $\boldsymbol{U}$. The parameter $\varepsilon \in (0, \infty)$ is a characteristic relaxation time for the system. Throughout the thesis, we will always assume that the flux terms and relaxation terms are smooth enough in $\boldsymbol{U}$.

The relaxation system is said to be hyperbolic if it satisfies the following definition.

Definition 3.1 (Hyperbolicity) *The system* (3.1) *is hyperbolic if the Jacobian of the flux, $D_{\boldsymbol{U}}\boldsymbol{F}(\boldsymbol{U})$, has real eigenvalues and is diagonalizable. The system is strictly hyperbolic if all the eigenvalues are real and distinct.*

If we let $\boldsymbol{P}(\boldsymbol{U}) = D_{\boldsymbol{U}}\boldsymbol{F}(\boldsymbol{U})$, the hyperbolic relaxation systems takes the following quasi-linear form:

$$\partial_t \boldsymbol{U} + \boldsymbol{P}(\boldsymbol{U})\partial_x \boldsymbol{U} = \frac{1}{\varepsilon}\boldsymbol{Q}(\boldsymbol{U}). \tag{3.2}$$

There does exist hyperbolic relaxation systems as (3.2), where it is not possible to write the system in the conservative form (3.1). This happens when $\boldsymbol{P}(\boldsymbol{U})$

cannot be expressed as the Jacobian of some flux-term. In these cases the system (3.2) is *nonconservative.*

For the most part, we will study conservative relaxation systems. The concepts in this chapter are therefore introduced for these systems. Note that most of the concepts makes sense for, and are applicable to, nonconservative systems as well.

An example of a relaxation system on conservation form is the much studied nonlinear 2×2 system introduced by Jin and Xin [28], mentioned in Chapter 1, for approximating the hyperbolic conservation law $u_t + f(u)_x = 0$,

$$\begin{aligned} u_t + v_x &= 0, \\ v_t + a^2 u_x &= \frac{1}{\varepsilon}(f(u) - v). \end{aligned} \tag{3.3}$$

3.1 The limits

We precisely define the formal limits of the conservative hyperbolic relaxation system (3.1). Formally, when $\varepsilon \to \infty$, we have the corresponding *homogeneous system*:

$$\partial_t \boldsymbol{U} + \partial_x \boldsymbol{F}(\boldsymbol{U}) = 0. \tag{3.4}$$

We assume that the relaxation term in (3.1) is endowed with a $n \times N$ matrix $\mathcal{Q}$, where $n < N$ such that $\mathcal{Q}\boldsymbol{R}(\boldsymbol{U}) = 0$ for all $\boldsymbol{U}$. This gives us the n-vector $\boldsymbol{u} = \mathcal{Q}\boldsymbol{U}$. We assume that $\boldsymbol{u}$ uniquely determines a local equilibrium value $\boldsymbol{U} = h(\boldsymbol{u})$ such that we have $\mathcal{Q}h(\boldsymbol{u}) = \boldsymbol{u}$ for all $\boldsymbol{u}$. Associated with $\mathcal{Q}$ are then the n conservation laws

$$\partial_t(\mathcal{Q}\boldsymbol{U}) + \partial_x(\mathcal{Q}\boldsymbol{F}(\boldsymbol{U})) = 0, \tag{3.5}$$

which can be rewritten as the local equilibrium approximation

$$\partial_t \boldsymbol{u} + \partial_x \boldsymbol{f}(\boldsymbol{u}) = 0, \tag{3.6}$$

where $\boldsymbol{f}(\boldsymbol{u}) = \mathcal{Q}\boldsymbol{F}(h(\boldsymbol{U}))$. The equation (3.6) will be referred to as the *equilibrium system.* Formally, this is the reduced system we obtain when $\varepsilon \to 0$.

For the 2×2-system (3.3), the formal limits are the homogeneous system,

$$u_t + v_x = 0, \tag{3.7a}$$

$$v_t + a^2 u_x = 0, \tag{3.7b}$$

and the scalar conservation law

$$u_t + f(u)_x = 0, \tag{3.8}$$

as the corresponding equilibrium system.

3.2 The subcharacteristic condition

The *subcharacteristic condition* is an important property regarding the wave-dynamics of hyperbolic relaxation systems. This condition states that the characteristic wave-velocities of the full system (3.4) interlace those of the equilibrium system (3.6). It was first introduced in the linear case by Whitham [52] and in the nonlinear case by Liu [33]. It is defined as follows.

Definition 3.2 (The subcharacteristic condition) *Let Λ_k and λ_j be the real eigenvalues of*

$$\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{U}} \qquad \text{and} \qquad \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{u}} \tag{3.9}$$

from (3.4) *and* (3.6) *respectively. The full system has N eigenvalues that satisfy*

$$\Lambda_1 \leq \cdots \leq \Lambda_k \leq \Lambda_{k+1} \leq \cdots \leq \Lambda_N, \tag{3.10}$$

where $\Lambda_k = \Lambda_k(h(\boldsymbol{u})$, and the n eigenvalues of the reduced system satisfy

$$\lambda_1 \leq \cdots \leq \lambda_j \leq \lambda_{j+1} \leq \cdots \leq \lambda_n, \tag{3.11}$$

with $\lambda_j = \lambda_j(\boldsymbol{u})$. Then the hyperbolic relaxation system (3.2) *fulfills the subcharacteristic condition if for each $j = 1, \ldots, n$,*

$$\lambda_j \in [\Lambda_j, \Lambda_{j+N-n}] \,. \tag{3.12}$$

The subcharacteristic condition is necessary for the stability of hyperbolic relaxation systems, as we will see in both Chapter 4 and Chapter 6. In Chapter 4 we prove that the subcharacteristic condition is both necessary and sufficient for stability of linear constant-coefficient systems with a rank one relaxation matrix.

3.3 Linear analysis

In this section, we linearize a general relaxation system in one space dimension around an equilibrium value. By doing this we can write the solution of the linearized system around an equilibrium value as plane-wave-like solutions. This makes it possible to study how each wave-component of the solution of the linearized system changes as ε varies. By studying each wave-component, we are able to establish equivalence between the subcharacteristic condition in Definition 3.2 and stability for linearized systems satisfying certain criteria in Chapter 4. In other words, by studying the plane-wave-like solutions we can find an equivalence between the subcharacteristic condition and linear stability around an equilibrium for the nonlinear relaxation system (3.1).

3.3.1 Linearization

We assume that the *equilibrium manifold* [22] $\xi = \{\boldsymbol{U} \in G : \boldsymbol{Q}(\boldsymbol{U}) = 0\}$ is non-empty and let $\hat{\boldsymbol{U}} \in \xi$. Linearizing the relaxation system (3.1) around the equilibrium state $\hat{\boldsymbol{U}}$, we have the linear system

$$\partial_t \boldsymbol{V} + \boldsymbol{A}\partial_x \boldsymbol{V} = \frac{1}{\varepsilon}\boldsymbol{R}\boldsymbol{V}, \tag{3.13}$$

$$\boldsymbol{V} = \boldsymbol{U} - \hat{\boldsymbol{U}}, \tag{3.14}$$

and where

$$\boldsymbol{A} = \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{U}} \qquad \text{and} \qquad \boldsymbol{R} = \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{U}} \tag{3.15}$$

are constant matrices evaluated at $\hat{\boldsymbol{U}}$.

3.3.2 Plane-wave solutions

For an initial condition $\boldsymbol{U}(x, 0) \in L^2(\mathbb{R})$, there exists a unique solution to (3.13) for all $\varepsilon > 0$ [29, Ch. 2], see Theorem 2.7.1 for the periodic case. If $\boldsymbol{U}(x, 0) \in L^2([a, b])$ were $[a, b]$ is an interval of length $M < \infty$, the solution can be written as a sum of its Fourier components:

$$\boldsymbol{U}(x,t) = \sum_k \boldsymbol{U}_k(x,t) = \sum_k \exp(\boldsymbol{H}(k)t)\exp(ikx)\hat{\boldsymbol{U}}(k), \tag{3.16}$$

where k is the wave number and $\boldsymbol{H}(k)$ is a wave-number dependent matrix given by

$$\boldsymbol{H}(k) = \frac{1}{\varepsilon}\boldsymbol{R} - ik\boldsymbol{A}. \tag{3.17}$$

We now write $\boldsymbol{H}$ on its Jordan form:

$$\boldsymbol{H}(k) = \boldsymbol{P}(k)\boldsymbol{J}(k)\boldsymbol{P}(k)^{-1}, \tag{3.18}$$

where $\boldsymbol{P}(k)$ is the corresponding matrix of generalized eigenvectors and $\boldsymbol{J}(k)$ the corresponding Jordan matrix. We may then write the general solution (3.16) as a combination of elementary waves:

$$\boldsymbol{V}(x,t) = \sum_k \sum_{j=1}^{N} \bar{V}_j(k,t)\exp(ikx + \lambda_j t), \tag{3.19}$$

for some amplitudes $\bar{V}_j(k,t)$, which are polynomials in t. To each eigenvalue λ_j of $\boldsymbol{H}(k)$ we can associate a dispersion relation,

$$v_j(k) = -\frac{1}{k}\mathrm{Im}(\lambda_j), \tag{3.20}$$

and amplification factor,

$$f_j(k) = \mathrm{Re}(\lambda_j), \tag{3.21}$$

as can be seen by writing (3.19) as

$$\boldsymbol{V}(x,t) = \sum_k \sum_{j=1}^{N} \bar{V}_j(k,t)\exp(f_j t)\exp\left(ik(x - v_j t)\right). \tag{3.22}$$

If $\boldsymbol{H}(k)$ is diagonalizable, $\boldsymbol{J}(k)$ reduces to a diagonal matrix consisting of the eigenvalues of $\boldsymbol{H}(k)$ and $\bar{V}_j(k,t) = \bar{V}_j(k)$. Then we can see from (3.22) that there is a plane-wave solution associated with each eigenvalue. It now follows from (3.17) that $\boldsymbol{H}$ satisfies the symmetry

$$\boldsymbol{H}(k) = \overline{\boldsymbol{H}(-k)}. \tag{3.23}$$

Hence

$$\lambda_j(k) = \overline{\lambda_j(-k)} \tag{3.24}$$

and consequently

$$f_j(k) = f_j(-k), \tag{3.25a}$$

$$v_j(k) = v_j(-k). \tag{3.25b}$$

Therefore, we may with no loss of generality study the wave-dynamics of linear relaxation systems for only non-negative wave numbers:

$$k \in [0, \infty). \tag{3.26}$$

3.4 Strictly nonlinear hyperbolic conservation laws

Conservative hyperbolic relaxation systems are hyperbolic conservation laws with source terms, i.e. balance laws, as long as ε is finite. The formal limits (3.4) and (3.6) will, in many cases, both be hyperbolic conservation laws. Understanding the theory for hyperbolic conservation laws is therefore important to better understand the theory for hyperbolic relaxation systems. So, for simplicity in this section, we will introduce some aspects concerning the solutions of homogeneous hyperbolic conservation laws.

In one dimension space, the hyperbolic conservation law has the following form:

$$\partial_t \boldsymbol{U} + \partial_x \boldsymbol{F}(\boldsymbol{U}) = 0. \tag{3.27}$$

We still assume that the flux function $\boldsymbol{F}$ is smooth enough in $\boldsymbol{U}$. The theory presented below, except the example in Section 3.4.1, is mainly from Bressan [9].

3.4.1 Weak solutions

A classical solution of the system does in general not exist, not even when we have smooth initial conditions. The variable $\partial_x \boldsymbol{U}$ may blow up in finite time. We show this by an example. Let us look at the scalar Burgers' equation,

$$u_t + \left(\frac{1}{2}u^2\right)_x = 0, \tag{3.28}$$

with smooth initial condition $u_0(x)$. By the method of characteristics, the solution can be written in the implicit form

$$x = x_0 + u_0(x_0)(t - t_0), \quad u = u_0(x_0). \tag{3.29}$$

The derivative of u w.r.t. x is

$$u_x = \frac{\partial u/\partial x_0}{\partial x/\partial x_0} = \frac{u_0'(x_0)}{1 + u_0'(x_0)(t - t_0)}, \tag{3.30}$$

where the denominator is equal to zero when $t = t_0 - 1/u_0'(x_0)$. Thus, for $t_0 \geq 1/u_0'(x_0)$,

$$\lim_{t \to t_0^-} u_x = -\infty. \tag{3.31}$$

Due to the development of discontinuities of the solutions of a hyperbolic conservation law in finite time, we have to define a global solution in a class of functions with discontinuities. Thus, we introduce the weak solutions in the sense of distributions. We multiply (3.4) with a smooth function $\phi(x,t)$ of compact support in $[0,\infty) \times \mathbb{R}$, i.e. $f(x) \in C_0^\infty\left([0,\infty) \times \mathbb{R}\right)$. We then integrate over $[0,\infty) \times \mathbb{R}$. By partial integration we are able to move the derivatives from $\boldsymbol{U}(x,t)$ to the smooth function $\phi(x,t)$. We end up with the following definition of a weak solution.

Definition 3.3 (A weak solution) *A weak solution $\boldsymbol{U}$ of* (3.4) *satisfies*

$$\int_{[0,\infty)} \int_{\mathbb{R}} \phi_t \boldsymbol{U} + \phi_x \boldsymbol{F}(\boldsymbol{U}) dx dt + \int_{\mathbb{R}} \phi(x,0) \boldsymbol{U}(x,0) dx = 0 \tag{3.32}$$

for any function $\phi(x,t) \in C_0^\infty\left(\mathbb{R} \times (0,\infty)\right)$.

3.4.2 Non-uniqueness

In general, weak solutions for hyperbolic conservation laws with Cauchy initial data are not unique. To illustrate this, we again look at the Burgers' equation (3.28), now with initial data

$$u(x,0) = \begin{cases} 1 & x \geq 0, \\ 0 & x < 0. \end{cases} \tag{3.33}$$

A weak solution of (3.28) with (3.33) has to satisfy

$$\int_0^\infty \int_{\mathbb{R}} \phi_t u + \phi_x \frac{u^2}{2} dx dt + \int_0^\infty \phi(x,0) dx = 0. \tag{3.34}$$

It is possible to show that for every α s.t. $0 < \alpha < 1$,

$$u_\alpha(x,t) = \begin{cases} 0 & \text{if} \quad x < \alpha t/2, \\ \alpha & \text{if} \quad \alpha t/2 \leq x < (1+\alpha)t/2, \\ 1 & \text{if} \quad (1+\alpha)t/2 \leq x, \end{cases} \tag{3.35}$$

is a solution satisfying (3.34). Thus, there are infinitely many weak solutions of (3.28). We need some additional conditions to find the unique, or the physically relevant, solution. In general, such conditions are called entropy conditions. We will look at two strongly related conditions: an additional conservation law for (3.4), an entropy conservation law, and vanishing viscosity. Both arise from physical considerations.

3.4.3 Entropy

The smooth scalar functions $\Phi(\boldsymbol{U})$ and $\Psi(\boldsymbol{U})$ are an entropy-entropy flux pair for the system (3.4) if any smooth solution $\boldsymbol{U}$ of (3.4) also satisfies the scalar conservation law

$$\Phi(\boldsymbol{U})_t + \Psi(\boldsymbol{U})_x = 0. \tag{3.36}$$

Herein, $\Phi(\boldsymbol{U})$ is called the entropy and $\Psi(\boldsymbol{U})$ the entropy flux. Observe that $\Psi(\boldsymbol{U})$ cannot be chosen independently of $\Phi(\boldsymbol{U})$. Carrying out the differentiation in (3.36), we get

$$D\Phi(\boldsymbol{U})\boldsymbol{U}_t + D\Psi(\boldsymbol{U})\boldsymbol{U}_x = 0. \tag{3.37}$$

We multiply (3.4) with $D\Phi(\boldsymbol{U})$:

$$D\Phi(\boldsymbol{U})\boldsymbol{U}_t + D\Phi(\boldsymbol{U})D\boldsymbol{F}(\boldsymbol{U})\boldsymbol{U}_x = 0, \tag{3.38}$$

and see that the entropy flux has to satisfy

$$D\Psi(\boldsymbol{U}) = D\Phi(\boldsymbol{U})D\boldsymbol{F}(\boldsymbol{U}) \tag{3.39}$$

for $\Phi(\boldsymbol{U})$ and $\Psi(\boldsymbol{U})$ to be an entropy pair for the system (3.4). If we demand that the entropy is convex, i.e. that it satisfies $D^2\Phi(\boldsymbol{U}) \geq 0$ as a quadratic form, it is strongly related to the vanishing viscosity condition.

3.4.4 Vanishing viscosity

One approach used to single out the unique solution is the vanishing viscosity approach. We add an artificial viscosity to the system in the following way:

$$\boldsymbol{U}_t^\varepsilon + \boldsymbol{F}(\boldsymbol{U}^\varepsilon)_x = \varepsilon \boldsymbol{U}_{xx}^\varepsilon. \tag{3.40}$$

We let $\varepsilon \to 0$ and demand that the unique weak solution of (3.4) is the limit of the solutions $\boldsymbol{U}^\varepsilon$. This approach is called *viscous regularization* and is motivated by the fact that most physical systems have some sort of viscosity or diffusion. We say that the solution $\boldsymbol{U}$ of the conservation law (3.4) is admissible in the vanishing viscosity sense if there does exist a sequence of solutions $\{\boldsymbol{U}^\varepsilon\}$ that converge to $\boldsymbol{U}$ in L^1_{loc} as $\varepsilon \to 0$.

We see how this method relates to the entropy condition (3.36) by first multiplying (3.40) with $D\Phi(\boldsymbol{U}^\varepsilon)$:

$$\Phi(\boldsymbol{U}^\varepsilon)_t + \Psi(\boldsymbol{U}^\varepsilon)_x = \varepsilon D\Phi(\boldsymbol{U}^\varepsilon)\boldsymbol{U}_{xx}^\varepsilon. \tag{3.41}$$

The term on the right hand side is equal to

$$D\Phi(\boldsymbol{U}^\varepsilon)\boldsymbol{U}_{xx}^\varepsilon = \Phi(\boldsymbol{U}^\varepsilon)_{xx} - D^2\Phi(\boldsymbol{U}^\varepsilon)\boldsymbol{U}_x^\varepsilon \otimes \boldsymbol{U}_x^\varepsilon. \tag{3.42}$$

We multiply (3.41) with a nonnegative test function ϕ of compact support in $\mathbb{R} \times [0, \infty)$ and integrate:

$$\int_0^\infty \int_{\mathbb{R}} \Phi(\boldsymbol{U}^\varepsilon)_t \phi + \Psi(\boldsymbol{U}^\varepsilon)_x \phi \; dxdt = \varepsilon \int_0^\infty \int_{\mathbb{R}} D\Phi(\boldsymbol{U}^\varepsilon)\boldsymbol{U}_{xx}^\varepsilon \phi \; dxdt. \tag{3.43}$$

If we assume that the entropy is convex, i.e. $D^2\Phi(\boldsymbol{U}) \geq 0$, we get

$$\int_0^\infty \int_{\mathbb{R}} \Phi(\boldsymbol{U}^\varepsilon)_t \phi + \Psi(\boldsymbol{U}^\varepsilon)_x \phi \; dxdt \geq \varepsilon \int_0^\infty \int_{\mathbb{R}} \Phi(\boldsymbol{U}^\varepsilon)_{xx} \phi \; dxdt. \tag{3.44}$$

We integrate by parts:

$$\int_0^\infty \int_{\mathbb{R}} \Phi(\boldsymbol{U}^\varepsilon)\phi_t + \Psi(\boldsymbol{U}^\varepsilon)\phi_x dxdt \geq -\varepsilon \int_0^\infty \int_{\mathbb{R}} \Phi(\boldsymbol{U}^\varepsilon)\phi_{xx} dxdt. \tag{3.45}$$

Finally, we let $\varepsilon \to 0$ and end up with

$$\int_0^\infty \int_{\mathbb{R}} \Phi(\boldsymbol{U}^\varepsilon)\phi_t + \Psi(\boldsymbol{U}^\varepsilon)\phi_x dxdt \geq 0, \tag{3.46}$$

such that Φ satisfies

$$\Phi(\boldsymbol{U})_t + \Psi(\boldsymbol{U})_x \leq 0 \tag{3.47}$$

for any weak solution $\boldsymbol{U}$. The inequality (3.47) is often referred to as *the entropy condition* and we can see from this inequality that the entropy has to be dissipative for the entropy condition and the vanishing viscosity approach to coincide for a class of suitable solutions.

As we will see in Chapter 6, the two entropy conditions, as they are defined here, cannot straightforwardly be used on nonconservative systems. A version of the entropy condition for conservative relaxation systems was introduced by Chen et al. [22]. It is closely related to the one presented here for hyperbolic conservation laws. The entropy condition for relaxation systems will be introduced and discussed in Chapter 6.

3.5 Summary

In this chapter, we have introduced some important concepts for hyperbolic relaxation systems which will be useful in the chapters to come.

First, we defined the formal ε-limits for general relaxation systems. A necessary condition for the relaxation system to be stable is the interlacing property known as the subcharacteristic condition.

We linearized general hyperbolic relaxation systems and wrote the solution of the corresponding linear system up as plane-wave-like solutions. By studying these solutions, we will show in Chapter 4 that the subcharacteristic condition is not merely necessary but also sufficient for stability for a special class of constant-coefficient systems.

Further, we briefly studied systems of nonlinear hyperbolic conservation laws to obtain some knowledge about the solutions of consrvative hyperbolic relaxation systems. This study serves as a short introduction to Chapter 6. We showed by an example that classical solutions to nonlinear hyperbolic conservation laws in general do not exist, not even in finite time. This called for a new type of solutions, namely the weak solutions in Definition 3.3. The weak solutions are in general not unique, as we showed by an example. On can obtain a unique weak solution for the system by imposing an entropy condition. The entropy conditions for relaxation systems, systems of conservation laws and nonconservative systems will be further studied in Chapter 6.

4 Wave-dynamics for linear hyperbolic relaxation systems

In this chapter, we will look at a special type of hyperbolic relaxation systems, namely one dimensional constant-coefficient systems with a relaxation matrix of rank one, given by

$$\partial_t \boldsymbol{V} + \boldsymbol{A}\partial_x \boldsymbol{V} = \frac{1}{\varepsilon}\boldsymbol{R}\boldsymbol{V}. \tag{4.1}$$

In the above $\boldsymbol{A}$ and $\boldsymbol{R}$ are constant real matrices and $\boldsymbol{R}$ is of rank one. Such systems may be the linearization of genuinely nonlinear hyperbolic relaxation systems, as seen in Section 3.3.1. We assume that the initial condition $V(x,0)$ is in $L^2([a,b])$ such that, following the approach in Section 3.3.2, (4.1) can be written as a sum of its Fourier components. First, we define the two ε-limits of the linear system:

Definition 4.1 (The homogeneous and equilibrium systems) *Let* (4.1) *be a linear constant-coefficient relaxation system with a rank one relaxation matrix. Then, when $\varepsilon \to \infty$, the relaxation system has the form*

$$\partial_t \boldsymbol{V} + \boldsymbol{A}\partial_x \boldsymbol{V} = 0 \tag{4.2}$$

with a solution vector $\boldsymbol{V}$ of dimension N. The system (4.2) *is known as the homogeneous system. When $\varepsilon \to 0$, we formally have the equilibrium system*

$$\partial_t \boldsymbol{v} + \boldsymbol{B}\partial_x \boldsymbol{v} = 0 \tag{4.3}$$

for some reduced variable $\boldsymbol{v}$ of length $N-1$.

We will consider the case where (4.1) is strictly hyperbolic, i.e. the roots of $\boldsymbol{A}$ are real and distinct.

By noticing that we can write the eigenvalue polynomial of (4.1) as the convex sum of the two eigenvalue polynomials of (4.2) and (4.3), we can prove equivalence between stability and the *subcharacteristic condition*. Further, for stable systems, a maximum principle for the velocities will be proved. These properties are already shown for the general linear 2×2-system in [2] and for a specific two-phase flow model in [46]. We will now prove these properties more generally.

Some of the proofs rely on the use of complex reactance functions from control theory. A definition of these functions will be stated. The material in this chapter is also presented in an article submitted for publishing in Appendix A. The article is a more compact version of this chapter and the use of reactance functions is omitted.

4.1 The characteristic equation as a convex sum

By matrix manipulations, we show that the characteristic equation for the linearized relaxation system (4.1) can be written as a convex sum depending only on the wave-number k, the relaxation time ε and the characteristic equation for (4.2) and (4.3). We will consider systems where the rank one relaxation matrix is stable.

4.1.1 Structure of the relaxation matrix

Let $\boldsymbol{R} \in \mathbb{R}^{N \times N}$ be the rank one relaxation matrix in (4.1). Then $\boldsymbol{R}$ satisfies

$$\boldsymbol{T}^{-1}\boldsymbol{R}\boldsymbol{T} = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ r_{N1} & \cdots & r_{NN} \end{pmatrix}. \tag{4.4}$$

where $\boldsymbol{T}$ is a similarity transform. From now on, we assume that this transformation already has been done in (4.1).

Remark 4.2 *The similarity transform $\boldsymbol{T}$ does not change the eigenvalue polynomial of $\boldsymbol{A}$* [47]*, such that the transformed system is strictly hyperbolic as well.*

We assume that the relaxation matrix is stable, i.e. the real parts of the eigenvalues of (4.4), $\Re(\lambda)$, have to be less than or equal to zero. Further, $r_{NN} < 0$ to ensure that the relaxation matrix is not defective. Now, we let the value of r_{NN} be absorbed into r_{Nk} for $k = 1, \ldots, N-1$ and into the relaxation parameter ε such that we get the following solution matrix for each wave-number k:

$$\boldsymbol{H}(k) = \frac{1}{\varepsilon}\boldsymbol{R} - ik\boldsymbol{A} = \frac{1}{\varepsilon}\begin{pmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ r_{N1} & \cdots & -1 \end{pmatrix} - ik\begin{pmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & & \vdots \\ a_{N1} & \cdots & a_{NN} \end{pmatrix}. \tag{4.5}$$

4.1.2 The equilibrium system

The matrix $\boldsymbol{B}$ for the equilibrium system (4.3) can be found by looking at solutions $\boldsymbol{V}$ satisfying $\boldsymbol{RV} = 0$. Let $\boldsymbol{R}$ be stable and $\boldsymbol{V} = [V_1, V_2, \ldots, V_N]^T$. Then we have

$$\sum_{k=1}^{N-1} r_{Nk}V_k - V_N = 0, \tag{4.6}$$

such that the equilibrium system (4.3) with $\boldsymbol{v} = [V_1, \ldots, V_{N-1}]$ is

$$\partial_t\boldsymbol{v} + \left[\begin{pmatrix} a_{11} & \cdots & a_{1,N-1} \\ \vdots & & \vdots \\ a_{N-1,1} & \cdots & a_{N-1,N-1} \end{pmatrix} + \begin{pmatrix} a_{1,N} \\ \vdots \\ a_{N-1,N} \end{pmatrix}\begin{pmatrix} r_{N,1} \cdots r_{N,N-1} \end{pmatrix}\right]\partial_x\boldsymbol{v} = 0. \tag{4.7}$$

Thus, the matrix $\boldsymbol{B}$ of the equilibrium system is equivalent to

$$\boldsymbol{B} = \boldsymbol{C}\boldsymbol{D}^T + \tilde{\boldsymbol{A}}, \tag{4.8}$$

where $\boldsymbol{C}$, $\boldsymbol{D}$ and $\tilde{\boldsymbol{A}}$ are

$$\boldsymbol{D} = \begin{pmatrix} r_{N,1} \\ \vdots \\ r_{N,N-1} \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} a_{1,N} \\ \vdots \\ a_{N-1,N} \end{pmatrix},$$
$$\tilde{\boldsymbol{A}} = \begin{pmatrix} a_{11} & \cdots & a_{1,N-1} \\ \vdots & & \vdots \\ a_{N-1,1} & \cdots & a_{N-1,N-1} \end{pmatrix}. \tag{4.9}$$

4.1.3 A convexity lemma

We now state and prove the convexity lemma.

Lemma 4.3 *Assume that the relaxation matrix is stable and of rank one. Then the characteristic polynomial for* (4.5) *can be written as the sum of two polynomials:*

$$\Psi(z) = \chi P_h(z) + (1-\chi)P_e(z) = 0, \tag{4.10}$$

where

$$\chi = \frac{\varphi}{\varphi+1} \in [0,1], \quad \varphi = k\varepsilon, \tag{4.11}$$

and where

$$P_h(z) = det\left(-i\boldsymbol{A} - z\boldsymbol{I}\right), \tag{4.12a}$$

$$P_e(z) = -det(-i\boldsymbol{C}\boldsymbol{D}^T - i\tilde{\boldsymbol{A}} - z\boldsymbol{I}). \tag{4.12b}$$

The matrices $\boldsymbol{C}$, $\boldsymbol{D}$ *and* $\boldsymbol{A}_{N,N}$ *are as defined in* (4.9), *and* $z = \lambda/k$. *Further,* $P_h(z) = 0$ *is equivalent to the characteristic equation for the homogeneous system* (4.2) *and* $P_e(z) = 0$ *is equivalent to the characteristic equation for the equilibrium system* (4.3).

Proof. We explicitly write out the expression for the solution matrix (4.5):

$$\boldsymbol{H} - \lambda\boldsymbol{I} = \frac{1}{\varepsilon}\boldsymbol{R} - ik\boldsymbol{A} - \lambda\boldsymbol{I} = \begin{pmatrix} -ika_{11} - \lambda & \cdots & -ika_{1N} \\ \vdots & & \vdots \\ \frac{r_{N1}}{\varepsilon} - ika_{N1} & \cdots & \frac{-1}{\varepsilon} - ika_{NN} - \lambda \end{pmatrix}. \tag{4.13}$$

Multiplying the characteristic equation of $\boldsymbol{H}$ with k^N, we get

$$\det\left(\frac{1}{\varphi}\boldsymbol{R} - i\boldsymbol{A} - z\boldsymbol{I}\right) = \det\begin{pmatrix} -ia_{11} - z & \cdots & -ia_{1N} \\ \vdots & & \vdots \\ \frac{r_{N1}}{\varphi} - ia_{N1} & \cdots & \frac{-1}{\varphi} - ia_{NN} - z \end{pmatrix} = 0, \tag{4.14}$$

where $\varphi = k\varepsilon$ and $z = \lambda/k$. Introducing $\boldsymbol{A}_{n,k}$ as the sub-matrix of $-i\boldsymbol{A} - z\boldsymbol{I}$ where the nth row and the kth column is removed, we have the characteristic equation in the following form:

$$\Psi(z) = \sum_{k=1}^{N-1} (-1)^{k-1} r_{Nk} \cdot \det(\boldsymbol{A}_{n,k}) - \det(\boldsymbol{A}_{N,N}) + \varphi \cdot \det(-i\boldsymbol{A} - z\boldsymbol{I}) = 0, \tag{4.15}$$

when expanding along the bottom row of (4.14). We now have $\Psi(z)$ as the sum of the two polynomials

$$\tilde{P}_h(z) = \det(-i\boldsymbol{A} - z\boldsymbol{I}), \tag{4.16a}$$

$$\tilde{P}_e(z) = \sum_{k=1}^{n} (-1)^{k-1} r_{nk} \cdot \det(\boldsymbol{A}_{n,k}) - \det(\boldsymbol{A}_{n,n}). \tag{4.16b}$$

Now we can easily see that $P_h(z) = \tilde{P}_h(z)$. We rewrite (4.16b):

$$\begin{aligned}
&\tilde{P}_e(z) \\
&= \det \begin{pmatrix} -ia_{11} - z & \cdots & -ia_{1,N-1} & -ia_{1N} \\ \vdots & \ddots & & \vdots \\ -ia_{N-1,1} & \cdots & -ia_{N-1,N-1} - z & -ia_{N-1,N} \\ r_{N,1} & \cdots & r_{N,N-1} & -1 \end{pmatrix} \qquad (4.17) \\
&= -\det \begin{pmatrix} ia_{11} + ia_{1N} r_{N,1} + z & \cdots & ia_{1,N-1} + ia_{1N} r_{N,N-1} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ ia_{N-1,1} + ia_{N-1,N} r_{N,1} & \cdots & ia_{N-1,N-1} + ia_{N-1,N} r_{N,N-1} + z & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \\
&= -\det(-i\boldsymbol{C}\boldsymbol{D}^T - i\tilde{\boldsymbol{A}} - z\boldsymbol{I}), \\
&= P_e(z)
\end{aligned}$$

by first adding $-ia_{in}$ multiplied with the last row to the ith row of (4.17) and then adding $r_{n,i}$ multiplied with the last column to the ith column for $i = 1, \ldots, n-1$. This does not change the determinant [47]. We now have

$$\Psi(z) = \varphi P_h(z) + P_e(z) = 0. \tag{4.18}$$

Substituting χ in (4.11) into (4.18), we get the characteristic polynomial (4.10). □

4.2 Stability and the subcharacteristic condition

In this section we prove that the linear relaxation system (4.1) is stable if and only if the roots of the two limiting polynomials interlace on the imaginary axis. This interlacing property is equivalent to the subcharacteristic condition from Chapter 3.

We prove the theorem by using complex reactance functions, which arise from system theory. We refer the reader to [45, 44, 6] to learn more about reactance functions. Here, we simply use the definition where it is convenient.

Definition 4.4 (Complex reactance function) *With α_k in the right half plane, the general monic reactance function is* [45]

$$Z(z) = \frac{\prod_{k=1}^{N}(z+\overline{\alpha_k}) - \prod_{k=1}^{N}(z-\alpha_k)}{\prod_{k=1}^{N}(z+\overline{\alpha_k}) + \prod_{k=1}^{N}(z-\alpha_k)}. \tag{4.19}$$

The reciprocal of $Z(z)$ is also a complex reactance function.

For clarity, we also define stability and strict stability for polynomials.

Definition 4.5 (Stability of polynomials) *Let $\{\lambda_i\}_{i=1}^N$ be the N roots of a polynomial of Nth order. The polynomial is stable if λ_i lie in the left half plane for all $i = 1, \ldots, N$. It is strictly stable if all eigenvalues lie in the open left half plane.*

We further say that the relaxation system (4.1) is stable if

$$|\exp(\boldsymbol{H}(k)t)| \leq C \quad \forall k \in \mathbb{R}, \quad \forall t \in [0, \infty), \tag{4.20}$$

where C is some positive constant and $|\cdot|$ denotes the L^2 norm for matrices. By the variable transformations

$$\eta = \frac{t}{\varepsilon}, \qquad \xi = -kt \tag{4.21}$$

the definition of stability of the linear relaxation system (4.1) is equivalent to the following definition:

Definition 4.6 *Consider the relaxation system* (4.1). *Assume that there is a* $C > 0$ *such that*

$$|\exp(\eta R + i\xi A)| \leq C \tag{4.22}$$

for all $\eta \geq 0$ *and* $\xi \in \mathbb{R}$. *Then the system is stable. If in addition all eigenvalues of the matrix* $\boldsymbol{H}(k)$ *have a real part* $\Re(\lambda_j) < 0$ *for all* k*, the system is said to be strictly stable.*

We prove the following lemma, which is from Kreiss [29].

Lemma 4.7 *Stability in the sense of Definition* 4.6 *is equivalent to the following two conditions.*

- *All eigenvalues* λ_j *of the matrix* $\boldsymbol{H}(k)$ *have a real part* $\mathrm{Re}(\lambda_j) \leq 0$.
- *If* J_r *is a Jordan block of the Jordan matrix* $\boldsymbol{J} = \boldsymbol{P}^{-1}\boldsymbol{H}\boldsymbol{P}$ *which corresponds to an eigenvalue* λ_j *with* $\Re(\lambda_j) = 0$*, then* J_r *has dimension* 1×1.

Proof. Observe that

$$\exp(\boldsymbol{H}(k)t) = \exp(\boldsymbol{P}\boldsymbol{J}\boldsymbol{P}^{-1}t) = \boldsymbol{P}\exp(\boldsymbol{J}t)\boldsymbol{P}^{-1}, \tag{4.23}$$

where $\boldsymbol{J}$ is the Jordan matrix. If J_r is a Jordan block with eigenvalue λ in $\boldsymbol{J}$, $\exp(J_r t)$ is bounded for all times $t \geq 0$ if and only if $\Re(\lambda) < 0$ or J_r has dimension 1×1. □

From Lemma 4.7, we can easily see that if the system (4.1) is strictly hyperbolic, the eigenvalue polynomial of (4.1) is stable if and only if the system is stable in sense of Definition 4.6. Thus, the two definitions 4.5 and 4.6 are equivalent for strictly hyperbolic systems.

Let

$$\Psi(z) = \chi P_h(z) + (1-\chi)P_e(z) = 0 \tag{4.24}$$

be the eigenvalue polynomial for the $N \times N$ linear hyperbolic relaxation system (4.1) where $\boldsymbol{R}$ is of rank one. Further, let $P_h(z)$ and $P_e(z)$ be as in Lemma 4.3. Then $\boldsymbol{A}$ is a $N \times N$ real matrix and $\boldsymbol{B}$ a $(N-1) \times (N-1)$ real matrix, making the coefficients of $P_h(z)$ and $P_e(z)$ alter between being purely real and purely imaginary:

$$P_h(z) = z^n + ib_{n-1}z^{n-1} + b_{n-2}z^{n-2} + \ldots, \tag{4.25a}$$

$$P_e(z) = z^{n-1} + ic_{n-2}z^{n-2} + c_{n-3}z^{n-3} + \dots, \tag{4.25b}$$

such that

$$\begin{aligned}\Psi(z) =& \chi P_h(z) + (1-\chi)P_e(z) \\ =& \chi(z^n + ib_{n-1}z^{n-1} + b_{n-2}z^{n-2} + \dots) \\ &+ (1-\chi)(z^{n-1} + ic_{n-2}z^{n-2} + c_{n-3}z^{n-3} + \dots).\end{aligned} \tag{4.26}$$

Before we continue with the analysis of the relaxation polynomial (4.24), we take a closer look at the general complex polynomial

$$P(z) = \sum_{k=0}^{n}(a_k + ib_k)z^k, \quad a_n + ib_n \neq 0. \tag{4.27}$$

This polynomial has two axially complementary polynomials

$$\begin{aligned} m(z) &= \frac{1}{2}\left[P(z) + \overline{P(-\overline{z})}\right], \\ p(z) &= \frac{1}{2}\left[P(z) - \overline{P(-\overline{z})}\right], \end{aligned} \tag{4.28}$$

which we assume have no roots in common. We can now prove the following relation between reactance functions and stable polynomials. This is also stated by Bose and Shi [6].

Lemma 4.8 *The polynomial* (4.27) *is strictly stable if and only if* $m(z)/p(z)$ *generates a complex reactance function.*

Proof. The fact that a complex reactance function generates a stable polynomial $P(z) = m(z) + p(z)$ can easily be seen by adding the numerator of (4.19) to the denominator:

$$\prod_{k=1}^{N}(z+\overline{\alpha_k}) - \prod_{k=1}^{N}(z-\alpha_k) + \prod_{k=1}^{N}(z+\overline{\alpha_k}) + \prod_{k=1}^{N}(z-\alpha_k) = 2\prod_{k=1}^{n}(z+\overline{\alpha_k}), \tag{4.29}$$

and since α_k by definition lies in the right half plane, $\lambda_k = -\overline{\alpha_k}$ will lie in the left half plane, thus $\Re(\lambda_k) < 0$. For the other direction, let us assume that we have a stable polynomial as in (4.27). Assume, for simplicity, and without loss of generality, that $P(z)$ is monic. We can rewrite (4.27) to

$$P(z) = \prod_{k=1}^{N} (z - z_k), \tag{4.30}$$

where $z_1, \dots, z_k$ are the roots with $\Re(z_k) < 0$. We have

$$\overline{P(-\overline{z})} = \prod_{k=1}^{N} \overline{(-\overline{z} - z_k)} \tag{4.31}$$

$$= \prod_{k=1}^{N} (-z - \overline{z_k}). \tag{4.32}$$

Let $\alpha_k = -\overline{z_k}$, then α_k will lie in the right half plane and

$$\frac{m(z)}{p(z)} = \frac{\prod_{k=1}^{N} (z + \overline{\alpha_k}) + (-1)^N \prod_{k=1}^{N} (z - \alpha_k)}{\prod_{k=1}^{N} (z + \overline{\alpha_k}) - (-1)^N \prod_{k=1}^{N} (z - \alpha_k)}, \tag{4.33}$$

which is a complex reactance function as in (4.19). From Definition 4.4 we know that the reciprocal of a complex reactance function is also a reactance function, such that $m(z)/p(z)$ has the same form as in (4.19) regardless of if N is even or odd. $\square$

To prove the next lemma, we need to define interlacing.

Definition 4.9 (Interlacing) *Let $\{\lambda_1, \lambda_2, \dots, \lambda_N\}$ be the roots of a polynomial $m(z)$ of order N and $\{\alpha_1, \alpha_2, \dots, \alpha_{N-1}\}$ the roots of $p(z)$ of order $N-1$. Then the roots of $m(z)$ interlace the roots of $p(z)$ on the real axis or on the imaginary axis if*

$$h(\lambda_1) \le h(\alpha_1) \le h(\lambda_2) \le \dots \le h(\alpha_{n-1}) \le h(\lambda_N) \tag{4.34}$$

where h denotes, respectively, the real parts or the imaginary parts of the roots. When $m(z)$ and $p(z)$ have no roots with real parts or imaginary parts in common, the respective interlacing is said to be strict.

A useful property for complex reactance functions is introduced in the following lemma:

Lemma 4.10 *The roots of $m(z)$ and $p(z)$ for a complex reactance function are simple, purely imaginary and interlace on the imaginary axis.*

Proof. We look at the case when N is even. For N odd, we can use the reciprocal of the reactance function (4.19) and then the proof is carried out in the same way.

By a fractional expansion it is possible to show [44, 57, 56] that the complex reactance function is equivalent to

$$Z(z) = \sum_{k=1}^{N} \frac{A_k}{2z + \overline{\alpha_k} - \alpha_k}, \quad A_k > 0, \tag{4.35}$$

where $A_k = \Re(\alpha_k)/2^{N-1}$ andα_k is as in Definition 4.4. Let $\beta_k = \Im(\alpha_k)$ be in increasing order and let $z = \sigma + j\omega$. Then (4.35) is equivalent to

$$\begin{aligned} Z(z) = \frac{p(z)}{m(z)} &= \sum_{k=1}^{N} \frac{2A_k\sigma}{2\sigma^2 + (2\omega - 2\beta_k)^2} + j \sum_{k=1}^{N} - \frac{A_k(2\omega - 2\beta_k)}{2\sigma^2 - (2\omega + 2\beta_k)^2} \\ &:= R(z) + jX(z). \end{aligned} \tag{4.36}$$

We see that $Z(z) = 0$ if $\sigma = 0$. Thus, the zeroes of $Z(z)$, hence the roots of $p(z)$, lie on the imaginary axis. We see from (4.37) that $\sigma = 0$ also for the poles, i.e the roots of $m(z)$. As $Z(z)$ is analytic in the right half plane, we can find the derivatives:

$$\frac{\partial R}{\partial \sigma} = \frac{\partial X}{\partial \omega} = \sum_{k=1}^{n} 2A_k \frac{(2\omega - 2\beta_k)^2 - 2\sigma^2}{(2\sigma^2 + (2\omega - 2\beta_k)^2)^2}. \tag{4.37}$$

We set $\sigma = 0$:

$$\sum_{k=1}^{n} 2A_k \frac{1}{(2\omega - 2\beta_k)^2} > 0. \tag{4.38}$$

From this expression we can see that X is strictly increasing on the imaginary axis from $-\infty$ to ∞ in the interval $(\Im(\alpha_{k-1}), \Im(\alpha_k))$. Therefore, the roots and poles of $Z(z)$ will alternate on the imaginary axis, i.e. the roots of $m(z)$ interlaces the roots of $p(z)$ on the imaginary axis. This gives us the desired conclusion. □

Similar proofs of Lemma 4.10 can be are made by Bose and Shi [6], Reza [44] and Zhareddine [57].

Reactance functions always fulfill the following lemma.

Lemma 4.11 *For a strictly stable polynomial of the nth order with complex coefficients written as in* (4.27)*, we have*

$$a_N a_{N-1} + b_N b_{N-1} > 0. \tag{4.39}$$

Proof. Notice that the coefficient for z^{N-1} is equal to minus the sum of all the roots and that the real part of the sum of the roots have to be less than zero. □

Remark 4.12 *As mentioned, reactance functions always fulfill Lemma* 4.11*. For general polynomials, we need to ensure that this condition is satisfied as well. One can easily find two interlacing polynomials $m(z)$ and $p(z)$ with distinct roots which generate a polynomial that is not strictly stable:*

$$\frac{m(z)}{p(z)} = \frac{(z-2i)(z+3i)}{-z+i}, \tag{4.40}$$

which generates

$$m(z) + p(z) = z^2 + (-1+i)z + 6 + i. \tag{4.41}$$

The polynomial (4.41) *cannot be stable since the sum of the roots is equal to $1-i$ such that the polynomial has to have at least one root in the right half plane.*

The following lemma can now be established for general complex-coefficient polynomials (4.27).

Lemma 4.13 *A general polynomial as in* (4.27) *is strictly stable if and only if the following holds:*

- *The axially complementary polynomials $m(z)$ and $p(z)$ have no roots in common.*
- *Their roots are distinct.*
- *Their roots interlace on the imaginary axis.*
- *The inequality* (4.39) *holds.*

Proof. Combine Lemma 4.11 with Lemma 4.8 and Lemma 4.10. □

Now, assuming strict hyperbolicity, we can formulate and prove a stability proposition for the system (4.1) when the relaxation matrix is stable and of rank one.

Proposition 4.14 *Assume that $\boldsymbol{R}$ is a stable rank one relaxation matrix for the linear relaxation system* (4.1)*. Let $\Psi(z)$ in* (4.10) *be its characteristic polynomial. Further, assume that the system is strictly hyperbolic.*

Then the system (4.1) *is stable for all $\chi \in [0, 1]$ if and only if the roots of $P_e(z)$ are purely imaginary and interlace the roots of $P_h(z)$ on the imaginary axis, i.e. the subcharacteristic condition is satisfied.*

Further, the subcharacteristic condition is strictly satisfied if and only if the system is strictly stable for all $\chi \in (0, 1)$.

Proof. We assume that all roots $P_h(z)$ and $P_e(z)$ have in common have been factored out such that we are left with the reduced polynomial $\Psi_r(z)$. All the roots that P_e and P_h have in common satisfy $\mathrm{Re}(\lambda) = 0$. Since the system is assumed to be strictly hyperbolic, all the roots that $P_e(z)$ and $P_h(z)$ have in common are distinct. Thus, the Jordan blocks corresponding to these eigenvalues will have dimension 1×1 and, according to Lemma 4.7, they will not cause any linear instability in the sense of Definition 4.6.

We now split the remaining polynomial $\Psi_r(z)$ into two axially complementary polynomials as in (4.28):

$$m(z) = \frac{1}{2}\left[\Psi_r(z) + \overline{\Psi_r(-\overline{z})}\right], \tag{4.42a}$$

$$p(z) = \frac{1}{2}\left[\Psi_r(z) - \overline{\Psi_r(-\overline{z})}\right]. \tag{4.42b}$$

From (4.26), observe that $m(z) = \chi P_{h,r}(z)$ and $p(z) = (1-\chi)P_{e,r}(z)$ if N is even and $p(z) = \chi P_{h,r}(z)$ and $m(z) = (1-\chi)P_{e,r}(z)$ if N is odd, making $P_{h,r}(z)$ and $P_{e,r}(z)$ axially complementary. We easily see that for $\chi = 1$ we have the homogeneous eigenvalue polynomial and for $\chi = 0$ we have the equilibrium eigenvalue polynomial.

From now on, we look at $\chi \in (0, 1)$. Corresponding to the coefficients for the general polynomial (4.27), $\Psi_r(z)$ has

$$a_N = \chi, \quad b_N = 0, \quad a_{N-1} = (1-\chi), \tag{4.43}$$

such that (4.39) always is fulfilled. It now follows from Lemma 4.13 that the roots $\{z_j\}$ of $\Psi_r(z)$ satisfy $\mathrm{Re}(z_j) < 0$ if and only if the roots of $P_{h,r}(z)$ interlace the roots of $P_{e,r}$ and their roots are distinct and purely imaginary.

If $P_e(z)$ and $P_h(z)$ have no roots in common, we have $\mathrm{Re}(z_j) < 0$ for all roots of $\Psi(z)$ when $\chi \in (0, 1)$, making the system strictly linearly stable.

If all roots of $P_e(z)$ are roots of $P_h(z)$, the remaining eigenvalue polynomial will have one root, $\Psi_r(z) = \varphi(z - z_k) + (\varphi - 1)$, where z_k is a root of $P_h(z)$. This root is always stable as $\varphi \leq 1$.

Remember from Chapter 3 that the imaginary parts of the eigenvalues correspond to the wave-velocities of the system. We have therefore shown that the stability of (4.1) is equivalent to the subcharacteristic condition being fulfilled. □

Though the equivalence between stability of convex combinations of polynomials and the interlacing property is well-known in system theory [6], it is less known in the context of relaxation systems. There does already exist some results concerning the relationship between stability of relaxation systems and the subcharacteristic condition. That the stability of (4.1) with a stable rank one relaxation matrix implies the subcharacteristic condition is proven, by contradiction, by Yong [55, 54]. This proof does, however, not make use of the convexity of the eigenvalue polynomial. The opposite direction, to our knowledge, has not been shown. Lorenz and Schroll [34] focus on the well-posedness of constant-coefficient systems in the limit as $\varepsilon \to 0$, but they also prove that stability of the relaxation system implies a different version of the subcharacteristic condition. Their version is a maximum principle stating that if the system is stable, the maximum and minimum velocities of the equilibrium system will not exceed the velocities of the corresponding homogeneous system. Chen et al. [22] shows for nonlinear systems that the existence of a strictly convex entropy implies the subcharacteristic condition. We will see in Chapter 6 that the existence of a strictly convex entropy is a stronger criterion than the stability in Definition 4.6 for linear systems. The equivalence between stability and the subcharacteristic condition is already proven for general 2×2-systems by Aursand and Flåtten [2] and for a specific two-phase model with a rank one relaxation term in Solem et al. [46]. Proposition 4.14 is therefore a generalization of these results to $N \times N$-systems. From this discussion, it is likely that Proposition 4.14 is an original contribution.

From the results in Proposition 4.14 we can show that the subcharacteristic condition is sufficient for convergence of solutions in $L^2(\mathbb{R})$ with the following proposition:

Proposition 4.15 *Assume that* (4.1) *is strictly hyperbolic. Further assume that* (4.1) *satisfies the subcharacteristic condition. Then the solution* $\boldsymbol{V}$ *to* (4.1) *with initial data* $\boldsymbol{V}_0 \in L^2(\mathbb{R})$ *converges in* $L^2(\mathbb{R})$ *for each* $t > 0$ *as* $\varepsilon \to 0$.

Proof. From Proposition 4.14 we know that the subcharacterstic condition implies stability, i.e. (4.20) is fulfilled.

The rest of the proof follows the proof of Theorem 2.3 by Yong [54]. Let $\boldsymbol{V}^\varepsilon(x,t) \in L^2(\mathbb{R})$ denote the solution of (4.1). Let $\hat{\boldsymbol{V}}^\varepsilon(k,t)$ be the Fourier transform of the solution. It satisfies

$$\partial_t \hat{\boldsymbol{V}}^\varepsilon(k,t) = (\boldsymbol{H}(k))\, \hat{\boldsymbol{V}}^\varepsilon(k,t). \tag{4.44}$$

When ε is small, we have a singular perturbation problem. From singular perturbation theory [51] we get that as ε goes to zero, $\hat{\boldsymbol{V}}^\varepsilon$ will converge, locally uniformly with respect to $t \in (0, \infty)$, to the solution of

$$\partial_t \hat{\boldsymbol{v}}^\varepsilon = \boldsymbol{B}\hat{\boldsymbol{v}}_0, \tag{4.45}$$

where $\boldsymbol{B}$ is defined as in (4.8). Also, since the system is stable, we have

$$\|\hat{\boldsymbol{V}}^\varepsilon(k,t)\| \le C\|\hat{\boldsymbol{U}}_0\|. \tag{4.46}$$

By Parseval's formula we then have

$$\|\boldsymbol{V}^\varepsilon(x,t)\| = \|\hat{\boldsymbol{V}}^\varepsilon(k,t)\| \le C\|\hat{\boldsymbol{U}}_0\|, \tag{4.47}$$

such that, by the dominated convergence theorem, see for example [38, Ch. 5],

$$\begin{aligned}\lim_{\varepsilon\to 0} \|\boldsymbol{V}^\varepsilon(x,t)\| &= \lim_{\varepsilon\to 0} \|\hat{\boldsymbol{V}}^\varepsilon(k,t)\| \\ &= \|\hat{\boldsymbol{v}}(k,t)\| = \|\boldsymbol{v}(x,t)\|. \end{aligned} \tag{4.48}$$

□

Proposition 4.15 is a version, restricted to rank one systems, of a more general theorem found in [54]. The proof presented here is mainly the same as the one for general constant-coefficient systems in [54]. So, with a slightly altered version we could prove convergence for general constant-coefficient relaxation systems assuming that the relaxation system satisfies the stability criterion

(4.20) together with a *non-oscillation assumption*: none of the eigenvalues of the relaxation matrix are purely imaginary. Lorenz and Schroll [34] shows that the opposite direction of the more general version also holds.

We can also observe from Proposition 4.14 that when the linear relaxation system 4.1 is stable and strictly hyperbolic, the equilibrium system (4.3) is strictly hyperbolic.

4.3 A maximum principle

We prove that the velocities of strictly hyperbolic constant-coefficient relaxation systems with rank one relaxation matrices never exceed the velocities of the corresponding homogeneous system when the relaxation system is stable. To do this, we use some properties for polynomials from Fisk [20]. The result has some similarities with a result in Lorenz and Schroll [34], where it is shown that stability implies that the maximum and minimum velocities of the equilibrium system do not exceed the velocities of the corresponding homogeneous system. The maximum principle in this section is stronger as it shows that the velocities for *any* relaxation time ε will never exceed the velocities of the corresponding homogeneous system.

Let (4.24) be the eigenvalue polynomial for the $N \times N$ linear hyperbolic relaxation system with a stable relaxation matrix of rank one. Assume that the system is stable. Then, as shown in Proposition 4.14, the roots of $P_h(z)$ interlace the roots of $P_e(z)$ on the imaginary axis. Further, let $\Psi(z)_r(z)$ be the remaining polynomial after the roots that $P_e(z)$ and $P_h(z)$ have in common are factored out. We make a translation of the roots from the left half plane to the lower half plane:

$$\begin{aligned}\hat{\Psi}_r(z) &= i^n \Psi_r(-iz) \\ &= i^n \chi P_{h,r}(-iz) + i^n (1-\chi) P_{e,r}(-iz) \\ &= h(z) + ig(z).\end{aligned} \tag{4.49}$$

The roots of $h(z)$ and $g(z)$ in (4.50) interlace on the real axis. Further, the real roots of $h(z)$ and $g(z)$ correspond to the roots of $P_h(z)$ and $P_e(z)$ on the imaginary axis. When the roots $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ of $h(z)$ strictly interlace the roots $\{\alpha_1, \alpha_2, \ldots, \alpha_{n-1}\}$ of $g(z)$, $g(z)$ changes sign on the roots of $h(z)$. This is illustrated by an example in Figure 4.1.

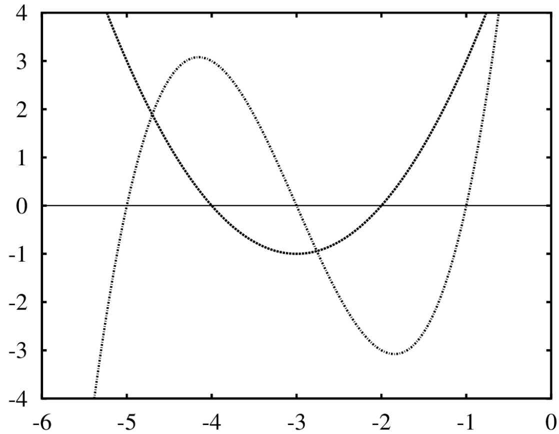


Figure 4.1: Two interlacing polynomials. $h(z) = (z+5)(z+3)(z+1)$ with a dotted line and $g(z) = (z+4)(z+2)$ with a stippled line.

The polynomial $h(z)$ is of order N and of one order more than $g(z)$. The homogeneous system is assumed to be strictly hyperbolic, making the roots of $h(z)$ distinct such that

$$\frac{h(z)}{z-\lambda_1}, \quad \frac{h(z)}{z-\lambda_2}, \dots, \frac{h(z)}{z-\lambda_n} \tag{4.50}$$

is a basis for all real polynomials with real roots of order $N-1$. We can therefore express $g(z)$ with this basis:

$$g(z) = \sum_{k=1}^{n} c_k \frac{h(z)}{z-\lambda_k}. \tag{4.51}$$

The c_ks have the same sign when the remaining eigenvalue polynomial $\Psi_r(z)$ is strictly stable. For a root λ_k of $h(z)$, we have

$$g(\lambda_k) = c_k(\lambda_k - \lambda_1)\dots(\lambda_k - \lambda_{k-1})(\lambda_k - \lambda_{k+1})\dots(\lambda_k - \lambda_n), \tag{4.52}$$

such that

$$\mathrm{sgn}(g(\lambda_k)) = \mathrm{sgn}(c_k)(-1)^{k+n} \tag{4.53}$$

and we can see that $g(z)$ changes sign on the roots of $h(z)$ if all the c_ks have the same sign.

The polynomial $P_e(z)$ in (4.24) has a positive leading coefficient, which makes all the c_k's positive. The c_ks also have to be strictly greater than zero. If not, λ_k would also be a root of $g(z)$, which contradicts the fact that $h(z)$ and $g(z)$ have no roots in common.

We can now prove the following theorem:

Proposition 4.16 *Let* (4.1) *be a strictly hyperbolic relaxation system with a stable rank one relaxation matrix. Let the system be stable. Then the imaginary parts of the roots, $\Im(z_k)$, for $k = 1, \dots, N$, of* (4.24) *satisfy*

$$\min_i\{\lambda_i)\} \leq \Im(z_k) \leq \max_i\{\lambda_i\}, \tag{4.54}$$

where $i\lambda_k$ are the roots of $P_h(z)$, for all $\chi \in [0, 1]$.

Proof. Since the system is stable, the remaining polynomial $\Psi_r(z)$ will be strictly stable for all $\chi \in (0, 1)$, such that the roots of $\chi P_{h,r}(z)$ strictly interlace the roots of $(1 - \chi)P_{e,r}(z)$ on the imaginary axis. We now look at the translated polynomial in (4.50). We can write (4.50) as

$$\hat{\Psi}_r(z) = h(z) + i \sum_{k=1}^{N} c_k \frac{h(z)}{z - \lambda_k}. \tag{4.55}$$

For a root z_i of (4.50), we will have

$$0 = 1 + i \sum_{k=1}^{N} c_k \frac{1}{z_i - \lambda_k}. \tag{4.56}$$

Assume that z_i is a root of $\hat{\Psi}_r(z)$ with $\Re(z_i) > \lambda_k$ for all $k = 1, \dots, n$. All the c_ks are greater than zero, making the real part of the sum in (4.56) greater than zero, such that the right hand side can not be equal to zero. Therefore, there are no roots z_i of (4.50) with real part greater than all the roots of $h(z)$. The proof for $\Re(z_i) < \lambda_k$ is similar.

We conclude that (4.50) has no roots with real part greater than or smaller than the real roots of $h(z)$. Translating (4.50) back to (4.24), we observe that the real part of the roots in (4.50) correspond to the imaginary parts of the roots in (4.24). The roots that $P_e(z)$ and $P_h(z)$ have in common are constant for any $\psi \in [0, 1]$ and will not be able to exceed any maximum or minimum value. $\square$

Remark 4.17 *The converse direction of Proposition* 4.16 *does not hold. We can easily generate two polynomials $P_1(z)$ and $P_2(z)$ satisfying the maximum principle that do not interlace, making the convex combination unstable,*

$$P_1(z) = (z + i5)(z + i)(z - i2), \tag{4.57a}$$

$$P_2(z) = (z + i4)(z + i2). \tag{4.57b}$$

4.4 Summary

We looked at strictly hyperbolic constant-coefficient relaxation systems with a stable relaxation matrix of rank one and proved three properties:

i) The characteristic polynomial for the system can be written as a convex sum of the characteristic polynomial for the homogeneous system and the equilibrium system.

ii) With this property, we could further prove that the stability of the system is equivalent to the subcharacteristic condition.

iii) The stability implies a maximum principle for the wave-velocities. The wave-velocities for any wave-number k and any relaxation time ε will never exceed the wave-velocities for the homogeneous system.

Here, i) makes it possible to prove ii) and iii) is a consequence of ii). To the author's knowledge, although one direction of ii) is proven by Yong [55] and the result in Proposition 4.13 is well-known in system theory, the results, and the proofs, in their completeness are original contributions in the context of hyperbolic relaxation systems.

The results from this section are useful in the applied sciences and in the next chapter we will apply the results from this chapter to a two-phase model.

5 Examples: Two linearized 3×3 relaxation systems

The two systems in this chapter serve as illustrative examples of the propositions established in Chapter 4. As we mentioned in Chapter 1, the theory in Chapter 4 has several applications as hyperbolic relaxation systems are widely used to describe different kinds of nonequilibrium phenomena.

In this chapter we will apply the theory from Chapter 4 to a specific two-phase model with applications to CO_2-transport. An interesting region with zero velocity is also identified for this system. To study this phenomenon, we solve a simpler system with the same wave-properties as the corresponding two-phase model.

The two-phase model is studied by Solem et al. [46]. The wave-dynamics for the model is considered by looking at the characteristic polynomial of the linearization of the model. The system models phase transfer and it takes the following form:

$$
\begin{aligned}
\partial_t(\alpha_{\mathrm{g}}\rho_{\mathrm{g}}) + \partial_x(\alpha_{\mathrm{g}}\rho_{\mathrm{g}}u) &= \frac{1}{\varepsilon}(\mu_\ell - \mu_{\mathrm{g}}), \\
\partial_t(\alpha_\ell\rho_\ell) + \partial_x(\alpha_\ell\rho_\ell u) &= -\frac{1}{\varepsilon}(\mu_\ell - \mu_{\mathrm{g}}), \\
\partial_t(\rho u) + \partial_x(\rho u^2 + p) &= 0, \\
\partial_t E + \partial_x(u(E+p)) &= 0.
\end{aligned}
\tag{5.1}
$$

Herein, α_i is the volume fraction of phase i, ρ_i is the density, and u is the common velocity. The variable p is the common pressure for the two phases and E is the total energy of the mixture. The mixed density is denoted as ρ and is given by

$$\rho = \rho_{\mathrm{g}}\alpha_{\mathrm{g}} + \rho_\ell \alpha_\ell, \tag{5.2}$$

where g denotes gas and l liquid. For a more thorough description of the model, see [46]. The two-phase model has a source term of rank one, and thus, assuming that the system is hyperbolic, the propositions from Chapter 4 can be applied. The characteristic polynomial for the linearized two-phase model, after noticing that the characteristic polynomial is invariant under a change of inertial reference frame and by introducing the dimensionless parameters [46]

$$\varphi = k\varepsilon\frac{\hat{c}^2}{\gamma\tilde{c}}, \quad z = \frac{\lambda}{k\tilde{c}}, \quad r = \frac{\hat{c}}{\tilde{c}}, \tag{5.3}$$

is

$$\varphi z^2(z^2+1) + z(z^2+r^2) = 0, \tag{5.4}$$

where k is the wave-number and ε the relaxation time. Also, λ is the original polynomial variable, $\tilde{c}$ is the sound velocity of the homogeneous system, $\hat{c}$ is the sound velocity of the equilibrium system and γ is a parameter that depends on the variables in the system. Corresponding to Lemma 4.3, we can now see that the homogeneous and equilibrium polynomials are, respectively,

$$P_h(z) = z^2(z^2+1), \tag{5.5a}$$
$$P_e(z) = z(z^2+r^2). \tag{5.5b}$$

Remark 5.1 *Note that the linearized two-phase model is not strictly hyperbolic as the eigenvalues of the polynomial for the homogeneous system has two zero-roots. However, by checking the eigenvectors of the system, it can be verified that the system is hyperbolic.*

5.1 Linear stability and interlacing

By using the results from Chapter 4, we can easily state stability conditions for the system (5.1). We eliminate the root in common, the zero root, and write the remaining polynomial (5.4) as a convex combination,

$$\begin{aligned}\Psi_r(z) &= \chi z(z^2+1) + (1-\chi)(z^2+r^2)\\ &= \chi P_{h,r}(z) + (1-\chi)P_{e,r}(z) = 0,\end{aligned} \tag{5.6}$$

where

$$\chi = \frac{\varphi}{1+\varphi}. \tag{5.7}$$

Remark 5.2 *Note from the stability discussions, especially from Lemma* 4.7*, from the previous chapter that a single eliminated root will not cause any instabilities for the linearized system.*

Now, the two corollaries below follow directly from the results in Chapter 4.

Corollary 5.3 *Let* $\chi \in [0, 1]$*, then the linearized two-phase model is stable if and only if* $0 \leq r \leq 1$*.*

Proof. From Proposition 4.14, the linearized two-phase model is stable if and only if the roots of the homogeneous polynomial $P_{h,r}(z)$ interlace the roots of the equilibrium polynomial $P_{e,r}(z)$ on the imaginary axis, meaning

$$-1 \leq -r \leq 0 \leq r \leq 1, \tag{5.8}$$

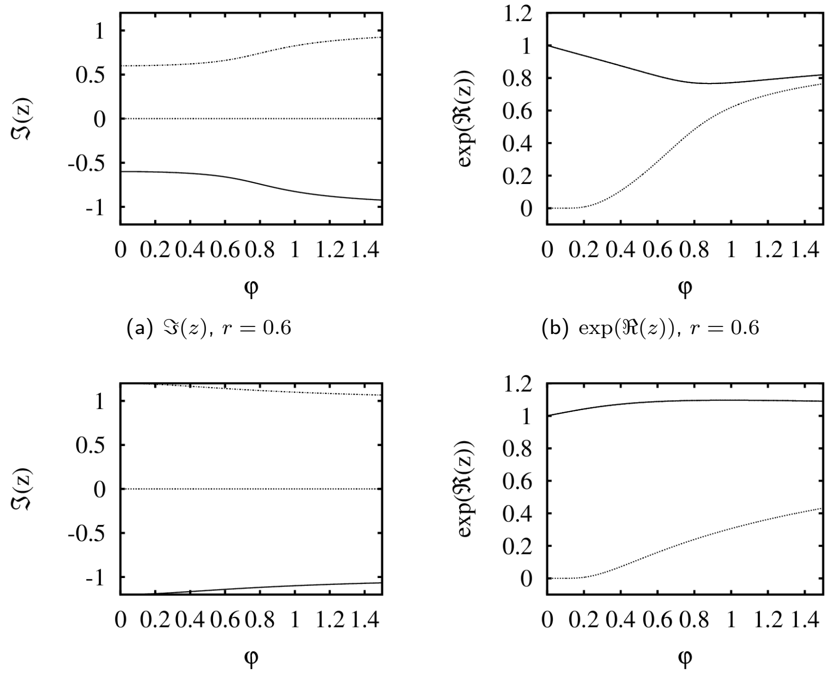
which is the same as $0 \leq r \leq 1$. □

Corollary 5.4 *Assume that the linearized two-phase model is stable, i.e. that the eigenvalue polynomial* (5.4) *is stable. Then the imaginary part of the roots of* (5.4) *for* $\chi \in [0, 1]$*,* $\Im(z_k)$ *will satisfy* $-1 \leq \Im(z_k) \leq 1$*.*

Proof. The proof follows directly from Proposition 4.16. □

Remark 5.5 *It may be worth noting that the even though the two-phase model, linearized around an equilibrium value, is stable, the full nonlinear version of the system* (5.1) *may not be stable around that equilibrium as the linearized system has one root equal to zero.*

From Chapter 3 we know that the imaginary parts and real parts of the eigenvalues are directly connected to the wave-velocities and amplifications of the system. To illustrate the two corollaries, we plot how the wave-velocities and amplifications behave in the stable and the unstable regime. The plots are shown in Figure 5.1. $\Im(z)$ are the imaginary parts of the eigenvalues and $\exp(\Re(z))$ the exponential of the real parts. Notice that we have plotted the eigenvalues w.r.t the variable φ instead of χ.



(a) $\Im(z)$, $r = 0.6$ (b) $\exp(\Re(z))$, $r = 0.6$

(c) $\Im(z)$, $r = 1.2$ (d) $\exp(\Re(z))$, $r = 1.2$

Figure 5.1: The wave-velocities $\Im(z)$ and amplification factors $\exp(\Re(z))$ in the stable region, $r = 0.6$, and in the unstable region, $r = 1.2$.

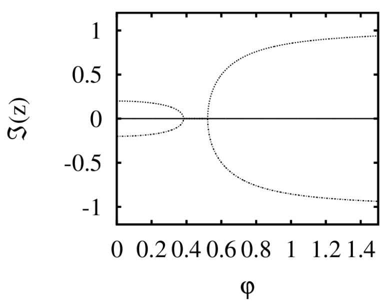
We note that φ depends on both the wave-number k and the relaxation time ε, such that there is a duality between k and ε; when $\varphi \to 0$ it could mean that $k \to 0$ or $\varepsilon \to 0$ or both.

When $r = 1.2$ the roots of the homogeneous polynomial no longer interlace the roots of the equilibrium polynomial, as one can see in Figure 5.1(c). Figure 5.1(d) shows that the corresponding amplification factor, the exponential of the real parts of the eigenvalues, $\exp(\Re(z))$, for $r = 1.2$ is larger than one. This means that the system is unstable. This corresponds to the result in Corollary 5.3.

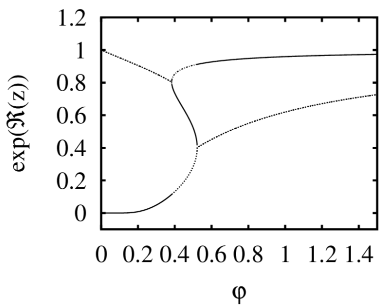
The system is stable by Corollary 5.3 for $r = 0.6$. Figures 5.1(a) and 5.1(b) show, respectively, the imaginary parts of the eigenvalues and the amplification factors for $r = 0.6$. We can see in Figure 5.1(c) that the interlacing property holds. Also, as none of the amplification factors are greater than one, the system is stable as expected.

5.2 A critical region

A property for the linearized two-phase model, first identified by Solem et al. [46], is a *critical region* for given relaxation times ε and wave-numbers k. In this region all the wave-velocities of the system are equal to zero, see Figure 5.2. The amplification factors also behave differently in this region. When this region occurs, we cannot continuously connect an amplification factor to a wave-velocity for all φ. This region emerges for the two-phase model when $|r| \leq 1/3$ and the region increases as r decreases.



(a) $\Im(z)$



(b) $\exp(\Re(z))$

Figure 5.2: Wave-velocities $\Im(z)$ and amplification factors $\exp(\Re(z))$ for the system when $r = 0.2$.

5.3 A solution

The critical region is an interesting phenomenon. We want to study what solutions in the region where this phenomena occurs, will look like compared to solutions in other regions. To do this, we construct a simpler system which has essentially the same characteristic polynomial as the two-phase model (5.1),

$$\begin{aligned}
\partial_t u_1 + \partial_x u_2 &= 0, \\
\partial_t u_2 + \partial_x u_3 &= 0, \\
\partial_t u_3 + \partial_x u_2 &= \frac{1}{\varepsilon}(r^2 u_1 - u_3),
\end{aligned} \tag{5.9}$$

where

$$\boldsymbol{H}(k) = \frac{1}{\varepsilon}\boldsymbol{R} - ik\boldsymbol{A} = \frac{1}{\varepsilon}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ r^2 & 0 & -1 \end{pmatrix} - ik\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \tag{5.10}$$

is the corresponding wave-number-dependent solution matrix. Here, $\varphi = k\varepsilon$. Observe that the zero root in (5.6) is eliminated as it is constant and will therefore not give much new insight. Further notice that since the eigenvalue polynomial of (5.9) is the same as $\Psi_r(z)$ for the two-phase model, Corollary 5.3 and Corollary 5.4 will hold for (5.9) as well.

Following the approach in Chapter 3, we find the solution of the system, with given initial conditions, as a sum of its Fourier components. We can then visualize how different values of both ε and r affect these solutions.

As noted in Section 3.3.2, for an initial condition $\boldsymbol{V}_0(x) \in L^2(\mathbb{R})$ there exists an unique solution $\boldsymbol{V}(x,t) \in L^2(\mathbb{R})$, such that $\boldsymbol{V}_0(x) = \boldsymbol{V}(x,0)$, to (5.9). If in addition $\boldsymbol{V}(x,0) \in L^2([a,b])$, where $[a,b]$ is of length M, there exists a unique solution to (5.9) in the general form

$$\boldsymbol{V}(x,t) = \sum_k \boldsymbol{V}_k(x,t) = \sum_k \exp(\boldsymbol{H}(k)t)\exp(ikx)\hat{\boldsymbol{V}}(k), \tag{5.11}$$

where we sum over all wave-numbers k. To see why (5.11) is a solution, we look at the Fourier transform of the general constant-coefficient relaxation system

$$\partial_t \boldsymbol{V} + \boldsymbol{A}\partial_x \boldsymbol{V} = \frac{1}{\varepsilon}\boldsymbol{R}\boldsymbol{V}. \tag{5.12}$$

The Fourier transform is

$$\hat{\boldsymbol{V}}_t(\xi,t) + i2\pi\xi\boldsymbol{A}\hat{\boldsymbol{V}}(\xi,t) = \frac{1}{\varepsilon}\boldsymbol{R}\hat{\boldsymbol{V}}(\xi,t), \tag{5.13}$$

with solution

$$\hat{\boldsymbol{V}}(\xi,t) = \exp\left(\frac{1}{\varepsilon}\boldsymbol{R} - i2\pi\xi\boldsymbol{A}\right)\hat{\boldsymbol{V}}_0, \tag{5.14}$$

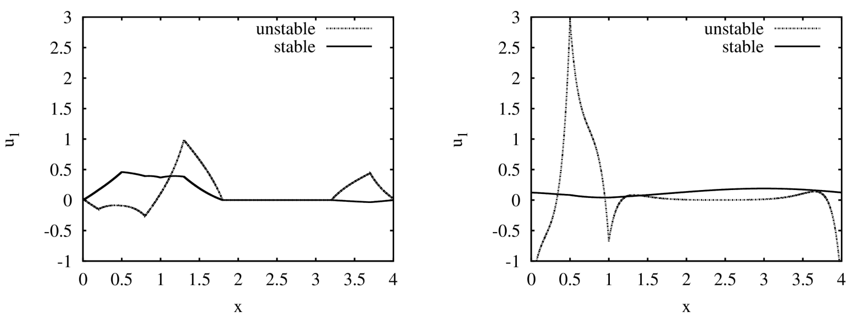
where $\hat{\boldsymbol{V}}_0$ is the Fourier transform of the initial condition. When $\boldsymbol{V}_0(x)$ is periodic, the Fourier coefficients of the solution are exactly the Fourier transform (5.14) divided by M and with $\xi = n/M$. This gives us (5.11) with $k = 2\pi\xi = 2\pi n/M$, where M is the length of the period, $n \in \mathbb{Z}$ and $\hat{\boldsymbol{V}} = \hat{\boldsymbol{V}}_0/M$.

As an initial condition we choose to use a triangle wave,

$$u_1(x,t) = \begin{cases} 1/2x & 0 \le x \le 1/2, \\ 2 - 1/2x & 1/2 < x \le 1, \\ 0 & \text{otherwise}, \end{cases} \tag{5.15}$$

for the first component. The triangle wave is simple, yet it illustrates both how the critical region affects the solution and how the relaxation term may have a smoothing effect on the solution. Since we have linearized around an equilibrium state, we set $u_3 = r^2 u_1$ and, for simplicity, we choose $u_2 = 1/2u_1$.

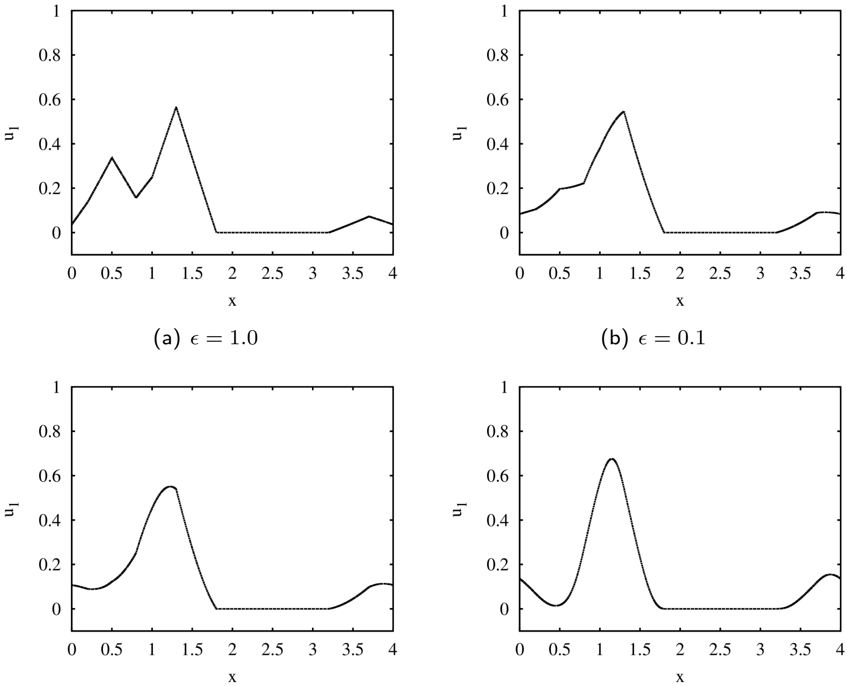
We have solved the system by using the discrete Fourier transform. We believe that the solutions depicted in Figure 5.4, Figure 5.3 and Figure 5.5 are close to the exact solutions, as increasing the number of Fourier-components with a large amount does not seem to change the appearance of the solutions in the figures.



(a) $t = 0.2$ (b) $t = 1.0$

Figure 5.3: The variable u_1 with $\varepsilon = 0.1$ at $t = 0.2$ and $t = 1.0$ in the stable, $r = 0.6$, and unstable, $r = 1.2$, regime.

We know from Corollary 5.3 that when $r > 1$ the solution of the system is unstable. This is illustrated in Figure 5.3 where the solution of the system is shown for $r = 0.6$ and $r = 1.2$ for both $t = 0.2$ and $t = 1.0$. The difference between the two solutions at $t = 1.0$ are significant. The stable solution, where



(a) $\epsilon = 1.0$ (b) $\epsilon = 0.1$

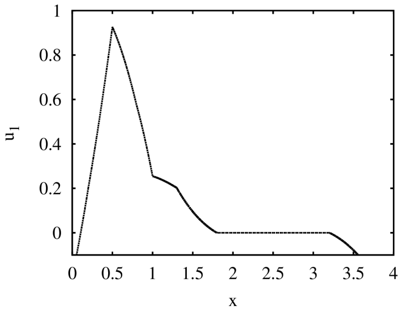
(c) $\epsilon = 0.05$ (d) $\epsilon = 0.01$

Figure 5.4: The variable u_1 for different values of ϵ at $t = 0.2$ with $r = 0.8$.

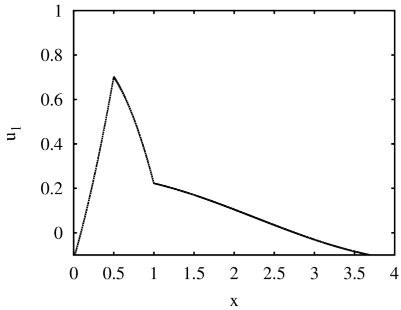
$r = 0.6$, is close to zero at $t = 1.0$, but the unstable solution, where $r = 1.2$, has grown significantly compared to the one at $t = 0.2$.

Figure 5.4 shows the solution u_1 of the system with $r = 0.8$ and $t = 0.2$. We observe how different values of the relaxation time ε affects the solution of u_1. The figure shows that the relaxation term has a smoothing effect on the solution and that this effect seems to increase as ε decreases. This smoothing effect does, however, disappear in the equilibrium limit $\varepsilon \to 0$, where (5.9) is reduced to a 2×2 homogeneous hyperbolic system with traveling wave solutions.

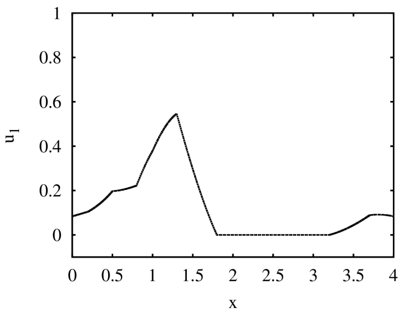
Figure 5.5 shows the difference between the solution in the critical region, where $r = 0.2$, and the solution in the noncritical region, where $r = 0.8$. The two top figures 5.5(a) and 5.5(b) display the solution for $r = 0.2$ at $t = 0.2$ and $t = 1.0$.



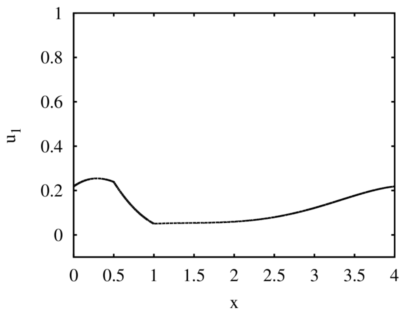
(a) $r = 0.2$, $t = 0.2$



(b) $r = 0.2$, $t = 1.0$



(c) $r = 0.8$, $t = 0.2$



(d) $r = 0.8$, $t = 1.0$

Figure 5.5: The variable u_1 with $\epsilon = 0.1$ at $t = 0.2$ to the left and $t = 1.0$ to the right, and with $r = 0.2$ at the top and $r = 0.8$ at the bottom.

The figures show that, as the time t increases, parts of the solution seem to stand still. As we know, the solution is a sum of Fourier components. As some of the Fourier components have zero wave-velocity in the critical region, these components will stand still as time increases and the other components move towards the right. If the sum of the components with zero velocity has nonzero amplitude, there will be nonzero parts of the solution standing still in the plot, as we can see in figures 5.5(a) and 5.5(b). Also, for this solution, by choosing a r smaller than 0.2, the critical region increases such that more components have zero wave-velocity and the resulting solution is more affected. In the same way, by choosing r closer to 1/3, the solution has fever zero-velocity components and the resulting solution looks more like a complete wave moving towards the right.

The figures 5.5(c) and 5.5(d) shows the solution for $r = 0.8$ at $t = 0.2$ and $t = 1.0$. As expected, these two figures show that the whole solution, each and every wave-component, is moving towards the right as the time increases. The wave has moved towards the right and reappeared at the left in the plot when $t = 1.0$ in Figure 5.5(d). This indicates that all wave-components have velocities larger than zero. We can also see that the wave has been smoothed out for $t = 1.0$ compared to for $t = 0.2$.

5.4 Summary

The propositions from Chapter 4 were applied to a linearized two-phase model such that stability properties could be derived for that model. We have shown, and also visualized, how the interlacing property is crucial for the stability of the linearized system. For this specific system, a critical region with zero wave-velocity for some wave-numbers k and relaxation times ε, was also identified.

We visualized how both instability and the critical region can affect solutions of systems having the same characteristic polynomial as the two-phase model. By finding a specific solution of a constant-coefficient system with essentially the same eigenvalue polynomial as the linearized two-phase model, we showed how the critical region can affect solutions of models with such regions. We observed that parts of the solution were left behind as the rest moved towards the right as time increased. This was not unexpected, as the solution is a sum of Fourier components. Some of these components have zero wave-velocity in the critical region. These components will, if the sum of them have nonzero amplitude, result in non-moving nonzero parts of the solution, which will be visible in the plots.

6 Entropies, conservative and nonconservative systems

Mathematical entropy is a concept, or tool, which is helpful when proving existence and uniqueness of weak solutions of hyperbolic systems, see for example Bressan and LeFloch [10]. Mathematical entropy was first introduced by Godunov [25] and then by Friedrichs and Lax [21] for hyperbolic conservation laws. Chen et al. [22] extended the notion of entropy to hyperbolic relaxation systems in conservative form by defining a dissipative convex entropy. In the same paper, it was shown that if there exists such an entropy for a hyperbolic relaxation system, the corresponding equilibrium system will also be hyperbolic and endowed with a convex entropy defined on the equilibrium manifold of the system.

For systems with at least three equations, there does not need to exist a corresponding entropy, as the system may be overdetermined [32]. But there does often exist entropies for systems which arise from physical considerations, systems from continuum physics. In many cases, these entropies are also strictly convex. It is therefore useful to look at systems with corresponding strictly convex entropies, as we will do in this chapter.

In this chapter we will study different aspects of entropies in connection with conservative and nonconservative relaxation systems. We start with constant-coefficient systems and show that the existence of a convex entropy is sufficient, but not necessary, for the stiffly well-posedness of these systems.

The existence of a convex entropy implies that the corresponding constant-coefficient system is symmetrizable. The other direction does also hold. The same is true for conservative nonlinear systems. We will also see that the subcharacteristic condition is a consequence of the existence of a strictly convex entropy for conservative relaxation systems.

Further, the notion of entropy will be studied in connection with nonconservative systems. Due to path-dependence, the entropies for these systems are not connected to the vanishing viscosity approach in the same way as the entropies for conservative systems are. We also show that nonconservative systems, satisfying a symmetry assumption, can be written in a conservative form.

We start with the definition of a strictly convex entropy for hyperbolic relaxation systems as defined by Chen et al. [22]. For the hyperbolic system

$$\partial_t \boldsymbol{U} + \partial_x \boldsymbol{F}(\boldsymbol{U}) = \frac{1}{\varepsilon}\boldsymbol{Q}(\boldsymbol{U}), \quad x \in \mathbb{R}, \tag{6.1}$$

where $\boldsymbol{U} \in G$, $G \subset \mathbb{R}^N$ is convex, and where $\boldsymbol{Q}$ and $\boldsymbol{F}$ are smooth in $\boldsymbol{U}$ and take values in $\mathbb{R}^{N\times N}$ and $\mathbb{R}^N$, the entropy is defined in the following way:

Definition 6.1 (Entropy for relaxation systems) *A strictly convex entropy $\Phi(\boldsymbol{U})$ for the system* (6.1) *is a smooth scalar function that, for any $\boldsymbol{U} \in G$, fulfills the following conditions.*

i) $D^2\Phi(\boldsymbol{U})D\boldsymbol{F}(\boldsymbol{U})$ *is symmetric.*

ii) $D\Phi(\boldsymbol{U})\boldsymbol{Q}(\boldsymbol{U}) \leq 0$.

iii) The following three statements are equivalent:

$$\boldsymbol{Q}(\boldsymbol{U}) = 0, \quad D\Phi(\boldsymbol{U})\boldsymbol{Q}(\boldsymbol{U}) = 0, \quad D\Phi(\boldsymbol{U}) = v^T\mathcal{Q} \text{ for some } v \in \mathbb{R}^N.$$

Herein, $\mathcal{Q}$ is a matrix such that $\mathcal{Q}\boldsymbol{Q}(\boldsymbol{U}) = 0$ for all $\boldsymbol{U} \in G$ and such that $\mathcal{Q}\boldsymbol{U}$ uniquely determines an equilibrium value $\boldsymbol{u}$, assuming that the equilibrium manifold is not empty.

We observe that *i)* ensures the existence of an entropy flux $\Psi(\boldsymbol{U})$ for the hyperbolic relaxation system. From Chapter 3 we know that the entropy flux has to satisfy

$$D\Psi(\boldsymbol{U}) = D\Phi(\boldsymbol{U})D\boldsymbol{F}(\boldsymbol{U}). \tag{6.2}$$

We differentiate (6.2) w.r.t $\boldsymbol{U}$:

$$D^2\Psi(\boldsymbol{U}) = D^2\Phi(\boldsymbol{U})D\boldsymbol{F}(\boldsymbol{U}) + D\Phi(\boldsymbol{U})D^2\boldsymbol{F}(\boldsymbol{U}). \tag{6.3}$$

Since G is assumed to be a convex space, we know that it is connected. The last term on the right hand side of (6.3) is symmetric. When *i)* holds, the whole expression on the right hand side is symmetric. By the Poincaré Lemma [40, Theorem 6.2.8, Remark 6.8.11], we can then integrate twice to get the entropy flux $\Psi(\boldsymbol{U})$.

As the entropy is assumed to be strictly convex, we know from Chapter 3 that zero viscosity solutions of the corresponding homogeneous system of conservation laws should satisfy

$$\Phi(\boldsymbol{U})_t + \Psi(\boldsymbol{U}) \leq 0. \tag{6.4}$$

Condition *ii)* is therefore necessary to ensure that the entropy is locally dissipated.

Notice that the definition of entropy in Definition 6.1 when letting $\boldsymbol{Q}(\boldsymbol{U}) = 0$ is reduced to the definition of entropy for homogeneous conservation laws, as defined in Chapter 3.

6.1 Constant-coefficient systems

In this section we will show that the existence of strictly convex entropy for constant-coefficient relaxation systems is sufficient, but not necessary, for stiffly well-posedness, assuming that the relaxation matrix $\boldsymbol{R}$ satisfies a nonoscillation condition. As before, the linear hyperbolic system takes the form

$$\partial_t \boldsymbol{U} + \boldsymbol{A}\partial_x(\boldsymbol{U}) = \frac{1}{\varepsilon}\boldsymbol{R}\boldsymbol{U}, \tag{6.5}$$

where $\boldsymbol{A}$ and $\boldsymbol{R}$ are constant real matrices.

Let $\boldsymbol{H}(k)$ be the solution matrix (3.17) in Chapter 3. The following two conditions hold if and only if the solutions, with initial conditions in $L^2(\mathbb{R})$, of (6.5) converge in $L^2(\mathbb{R})$ as $\varepsilon \to 0$, see Lorenz and Schroll [34].

i) $|\exp(\boldsymbol{H}(k)t)| \leq C, \forall k \in \mathbb{R}, \forall t \geq 0.$

ii) The relaxation matrix $\boldsymbol{R}$ has no purely imaginary eigenvalue different from 0.

If condition i) above is fulfilled, we say that the corresponding initial value problem to (6.5) is *stiffly well-posed*. We will, from now on, assume that the nonoscillation condition ii) is fulfilled for (6.5). Then stiffly well-posedness is equivalent to convergence of solutions in $L^2(\mathbb{R})$.

As mentioned, the existence of a strictly convex entropy is not necessary for well-posedness and is therefore not necessary for the convergence of solutions in $L^2(\mathbb{R})$. We prove that a convex entropy is not necessary by showing that there exist stiffly well-posed systems that do not satisfy both conditions in the following proposition:

Proposition 6.2 *Any hyperbolic relaxation system* (6.1) *endowed with a strictly convex entropy as defined in Definition* 6.1 *is symmetrizable, that is, there exists a positive definite symmetric matrix* $\boldsymbol{A}_0$ *s.t.*

$$\boldsymbol{A}_0(\boldsymbol{U})D\boldsymbol{F}(\boldsymbol{U}) = (D\boldsymbol{F}(\boldsymbol{U}))^T\boldsymbol{A}_0(\boldsymbol{U}). \tag{6.6}$$

Further, the relaxation term $\boldsymbol{Q}$ *satisfies*

$$\boldsymbol{A}_0(\boldsymbol{U})D\boldsymbol{Q}(\boldsymbol{U}) + (D\boldsymbol{Q}(\boldsymbol{U}))^T\boldsymbol{A}_0(\boldsymbol{U}) \leq 0 \quad \text{for} \quad \boldsymbol{U} \in \xi, \tag{6.7}$$

where $\xi \in G$ *is the equilibrium manifold.*

Proof. This proof is a slightly altered version of a proof by Yong [54].

We can easily see that $\boldsymbol{A}_0(\boldsymbol{U}) = D^2\Phi(\boldsymbol{U})$ is symmetric positive definite and that $D^2\Phi(\boldsymbol{U})D\boldsymbol{F}(\boldsymbol{U})$ is symmetric by the definition of entropy. To see that (6.7) holds, we observe that $D\Phi(\boldsymbol{U})\boldsymbol{Q}(\boldsymbol{U})$ takes maximum values at equilibrium and that

$$\begin{aligned}\Phi_{\boldsymbol{U}}(\boldsymbol{U})\boldsymbol{Q}(\boldsymbol{U}) &= (\Phi_{\boldsymbol{U}}(\boldsymbol{U}) - v^T\mathcal{Q})\boldsymbol{Q}(\boldsymbol{U}) \\ &= (\Phi_{\boldsymbol{U}}(\boldsymbol{U}) - \Phi_{\boldsymbol{U}}(\boldsymbol{U}_e))\boldsymbol{Q}(\boldsymbol{U}).\end{aligned} \tag{6.8}$$

Hence, the Hessian matrix of (6.8) at an equilibrium value $\boldsymbol{U}_e$ is nonpositive:

$$\begin{aligned}0 &\geq [(\Phi_{\boldsymbol{U}} - \Phi_{\boldsymbol{U}}(\boldsymbol{U}_e))\boldsymbol{Q}(\boldsymbol{U})]_{\boldsymbol{UU}}(\boldsymbol{U}_e) \\ &= \Big(\Phi_{\boldsymbol{UU}}\boldsymbol{Q}_{\boldsymbol{U}} + \Phi_{\boldsymbol{U}}\boldsymbol{Q}_{\boldsymbol{UU}} + \boldsymbol{Q}^T\Phi_{\boldsymbol{UUU}} + \boldsymbol{Q}_{\boldsymbol{U}}^T\Phi_{\boldsymbol{UU}} - \Phi_{\boldsymbol{U}}(\boldsymbol{U}_e)\boldsymbol{Q}_{\boldsymbol{UU}}\Big)(\boldsymbol{U}_e) \\ &= \Phi_{\boldsymbol{UU}}\boldsymbol{Q}_{\boldsymbol{U}} + \boldsymbol{Q}_{\boldsymbol{U}}^T\Phi_{\boldsymbol{UU}}.\end{aligned} \tag{6.9}$$

□

We show that there does not need to exist a strictly convex entropy for (6.5) to be stiffly well-posed. Let us look at the 3×3 example system from Chapter 5,

$$\boldsymbol{U}_t + \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \boldsymbol{U}_x = \frac{1}{\varepsilon} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ r^2 & 0 & -1 \end{pmatrix}, \tag{6.10}$$

where $r^2 \leq 1$. It is not possible to find a symmetric matrix $\boldsymbol{A}_0$ that satisfies both conditions in Proposition 6.2 for this system. $\boldsymbol{A}_0$ has to be in the form

$$\boldsymbol{A}_0 = \begin{pmatrix} -c & 0 & c \\ 0 & c+b & d \\ c & d & b \end{pmatrix} \tag{6.11}$$

for some constants c, b and d satisfying

$$0 < -c < b \quad \text{and} \quad d^2 < (c+b)^2 \tag{6.12}$$

to be positive definite and at the same time symmetrize $\boldsymbol{A}$ in (6.10). But the matrix (6.11) cannot fulfill (6.7) with the relaxation matrix in (6.10). Thus, there does not exist a strictly convex entropy for the system (6.10). But, the system is stable and the relaxation matrix is stable and of rank one. We therefore know from Proposition 4.15 in Chapter 4 that solutions of (6.10) converge in $L^2(\mathbb{R})$ as $\varepsilon \to 0$, making the system stiffly well-posed. This shows that there exists stiffly well-posed linear constant-coefficient systems without strictly convex entropies. We can conclude that the existence of a strictly convex entropy for constant-coefficient systems is not necessary for stiffly well-posedness. The existence of a strictly convex entropy is, nevertheless, sufficient, as we will show in the following.

There is equivalence between (6.7) and the existence of a strictly convex quadratic entropy for linear systems, see Lorenz and Schroll [35]. This means that if there exists a strictly convex entropy for a constant-coefficient system, there also exists a strictly convex quadratic entropy for the system. And, if there exists a strictly convex quadratic entropy for the system, it fulfills (6.7) and symmetrizes the system. A strictly convex quadratic entropy is a function in the form

$$\Phi(\boldsymbol{U}) = \frac{1}{2} \boldsymbol{U}^T \boldsymbol{H} \boldsymbol{U}, \tag{6.13}$$

where $\boldsymbol{H}$ is a positive definite matrix that symmetrizes the system, i.e it is a symmetrizer. Summing up, the existence of a quadratic entropy is equivalent to (6.6) and (6.7). We state this as a proposition.

Proposition 6.3 *The relaxation system* (6.5) *is symmetrizable and* (6.7) *is fulfilled if and only if there exists a strictly convex quadratic entropy for the system.*

Proof. The proof is from Lorenz and Schroll [35]. Assume that there exists a symmetrizer $\boldsymbol{A}_0$ for the system (6.5) such that (6.7) is fulfilled. Then the strictly convex quadratic function (6.13) with $\boldsymbol{H} = \boldsymbol{A}_0$ is an entropy for the system if and only if $\Phi_{\boldsymbol{UU}}(\boldsymbol{U})D(\boldsymbol{AU}) = \boldsymbol{A}_0\boldsymbol{A}$ is symmetric, which holds since $\boldsymbol{A}_0$ is a symmetrizer for the system. Then we have

$$\Psi(\boldsymbol{U}) = \frac{1}{2}\boldsymbol{U}^T\boldsymbol{A}_0\boldsymbol{A}\boldsymbol{U}$$

as the corresponding entropy flux. This leads to

$$\Phi(\boldsymbol{U})_t + \Psi(\boldsymbol{U}) = \boldsymbol{U}^T\boldsymbol{A}_0\boldsymbol{U}_t + \boldsymbol{U}^T\boldsymbol{A}_0\boldsymbol{A}\boldsymbol{U}_x = \frac{1}{\varepsilon}\boldsymbol{U}^T\boldsymbol{A}_0\boldsymbol{R}\boldsymbol{U}.$$

By (6.7) we then have the entropy inequality (6.4).

Now assume that (6.13) is a strictly convex quadratic entropy for the system (6.5). Then $\boldsymbol{H}$ is clearly a symmetrizer for the system and we have

$$0 \geq \Phi(\boldsymbol{U})_t + \Psi(\boldsymbol{U}) = \boldsymbol{U}^T\boldsymbol{H}\boldsymbol{U}_t + \boldsymbol{U}^T\boldsymbol{H}\boldsymbol{A}\boldsymbol{U}_x = \frac{1}{\varepsilon}\boldsymbol{U}^T\boldsymbol{H}\boldsymbol{R}\boldsymbol{U},$$

which shows that (6.7) holds. □

We now know that the existence of a convex entropy for constant-coefficient systems implies the existence of a quadratic entropy, which implies that the system is symmetrizable in the sense that (6.6) and (6.7) holds. By the Kreiss matrix theorem [29, Theorem 2.3.2], the stiff well-posedness condition i) follows from the conditions (6.6) and (6.7). Thus, stiff well-posedness is a consequence of the existence of a convex entropy for constant-coefficient systems. And, as we have assumed that the relaxation matrix is nonoscillatory, the existence of a convex entropy implies that the solutions will converge in $L^2(\mathbb{R})$ as the relaxation parameter $\varepsilon \to 0$. We can now conclude that the existence of a convex entropy is sufficient, but not necessary, for stiffly well-posedness of constant-coefficient relaxation systems.

As the existence of a convex entropy is sufficient for well-posedness of linear constant-coefficient systems, it is reasonable to assume that the same entropy criterion also imposes some stability on nonlinear hyperbolic systems.

6.2 Nonlinear conservative systems

In this section we will look at nonlinear conservative hyperbolic systems. As the properties presented in this section, with the exception of the subcharacteristic condition, are applicable to both homogeneous systems and relaxation systems, we will, for simplicity, mainly look at homogeneous hyperbolic systems,

$$\partial_t \boldsymbol{U} + \partial_x \boldsymbol{F}(\boldsymbol{U}) = 0. \tag{6.14}$$

We still assume that $\boldsymbol{U} \in G$, where $G \subset \mathbb{R}^N$ is convex, and that $\boldsymbol{F}$ is smooth in $\boldsymbol{U}$.

We will first see that the system (6.14) can be written in a symmetric form if and only if there exists a corresponding strictly convex entropy. We will also briefly look at the vanishing viscosity approach. Finally, we state that the existence of a convex entropy for conservative relaxation systems implies the subcharacteristic condition, defined in Chapter 3.

6.2.1 Symmetric form

For systems of nonlinear conservation laws, there exists a strictly convex entropy if and only if the system can be written in a symmetric form. This is an important property as we can use the symmetric form to find an entropy and vica versa. We will prove this property, but first we define the symmetric form of a system.

Definition 6.4 (Symmetric form) *A hyperbolic conservation law* (6.14) *is on a symmetric form if it can be written in the following way:*

$$\boldsymbol{g}(\hat{\boldsymbol{U}})_t + \boldsymbol{h}(\hat{\boldsymbol{U}})_x = 0, \tag{6.15}$$

where $D\boldsymbol{g}(\hat{\boldsymbol{U}})$ and $D\boldsymbol{h}(\hat{\boldsymbol{U}})$ are symmetric and $\boldsymbol{g}(\hat{\boldsymbol{U}})$ is positive definite.

Remark 6.5 *A system endowed with a strictly convex entropy can be written in a symmetric form* [32]*. From Proposition* 6.2 *we know that if the system is endowed with a strictly convex entropy, we have*

$$D^2\Phi(\boldsymbol{U})D\boldsymbol{F}(\boldsymbol{U}) = (D^2\Phi(\boldsymbol{U})D\boldsymbol{F}(\boldsymbol{U}))^T. \tag{6.16}$$

Then $(D^2\Phi(\boldsymbol{U}))^{-1}$ will also symmetrize the system.

As the entropy is strictly convex, $\hat{\boldsymbol{U}} = D\Phi$ is strictly increasing, and we can find $\boldsymbol{U}$ as a function of $D\Phi$: $\boldsymbol{U} = \boldsymbol{g}(D\Phi) = \boldsymbol{g}(\hat{\boldsymbol{U}})$. And, we set $\boldsymbol{F}(\boldsymbol{U}) = \boldsymbol{h}(\hat{\boldsymbol{U}})$. We rewrite the conservation law (6.15)*:*

$$D_{\hat{\boldsymbol{U}}}\boldsymbol{g}(\hat{\boldsymbol{U}})\hat{\boldsymbol{U}}_t + D_{\hat{\boldsymbol{U}}}\boldsymbol{h}(\hat{\boldsymbol{U}})\hat{\boldsymbol{U}}_x = 0.$$

We see that $D_{\hat{\boldsymbol{U}}}\boldsymbol{g}(\hat{\boldsymbol{U}}) = (D^2\Phi(\boldsymbol{U}))^{-1}$ and $D_{\hat{\boldsymbol{U}}}\boldsymbol{h}(\hat{\boldsymbol{U}}) = D\boldsymbol{F}(\boldsymbol{U})(D^2\Phi(\boldsymbol{U}))^{-1}$ puts the system in a symmetric form. $D_{\hat{\boldsymbol{U}}}\boldsymbol{g}(\hat{\boldsymbol{U}}) = (D^2\Phi(\boldsymbol{U}))^{-1}$ is symmetric as it is the inverse of the Hessian of the entropy $\Phi(\boldsymbol{U})$ and $D_{\hat{\boldsymbol{U}}}\boldsymbol{h}(\hat{\boldsymbol{U}})$ is symmetric as the expression in (6.16) *is. Further, $D_{\hat{\boldsymbol{U}}}\boldsymbol{g}(\hat{\boldsymbol{U}}) = (D^2\Phi(\boldsymbol{U}))^{-1}$ is positive definite since the entropy $\Phi(\boldsymbol{U})$ is strictly convex.*

Remark 6.6 *Any symmetric matrix is diagonalizable with real eigenvalues. Hence, the symmetric system* (6.15) *is always hyperbolic. See Chen* [12]*.*

From Remark 6.5 we know that any hyperbolic system of conservation laws with a corresponding convex entropy can be written in a symmetric form. We will now prove that the opposite direction holds as well.

Proposition 6.7 (Symmetry) *A system of hyperbolic conservation laws can be written in a symmetric form if and only if there exists a strictly convex entropy for the system.*

Proof. If the system is endowed with a strictly convex entropy, it follows from Remark 6.5 that the system can be written in a symmetric form.

We prove the other direction. The proof is from LeFloch [32]. If the system is in a symmetric form, we can write it in the form in (6.15). Further, there will exist two scalar functions ψ and ϕ such that

$$\boldsymbol{g}(\hat{\boldsymbol{U}}) = D\phi(\hat{\boldsymbol{U}}) \quad \text{and} \quad \boldsymbol{h}(\hat{\boldsymbol{U}}) = D\psi(\hat{\boldsymbol{U}}), \tag{6.17}$$

since $D\boldsymbol{g}(\hat{\boldsymbol{U}})$ and $D\boldsymbol{h}(\hat{\boldsymbol{U}})$ are symmetric. Since $Dg(\hat{\boldsymbol{U}})$ is positive definite, ϕ is strictly convex, making $\boldsymbol{g}(\hat{\boldsymbol{U}})$ injective such that we can find $\hat{\boldsymbol{U}}(\boldsymbol{g})$. We can therefore use the Legendre transforms of ψ and ϕ:

$$\tilde{\Phi}(\hat{\boldsymbol{U}}) = D\phi(\hat{\boldsymbol{U}}) \cdot \hat{\boldsymbol{U}} - \phi(\hat{\boldsymbol{U}}), \quad \tilde{\Psi}(\hat{\boldsymbol{U}}) = D\psi(\hat{\boldsymbol{U}}) \cdot \hat{\boldsymbol{U}} - \psi(\hat{\boldsymbol{U}}). \tag{6.18}$$

Now, (6.18) satisfies

$$\partial_t\tilde{\Phi}(\hat{\boldsymbol{U}}) = \partial_t D\phi(\hat{\boldsymbol{U}}) \cdot \hat{\boldsymbol{U}}, \quad \partial_x\tilde{\Psi}(\hat{\boldsymbol{U}}) = \partial_x D\psi(\hat{\boldsymbol{U}}) \cdot \hat{\boldsymbol{U}}, \tag{6.19}$$

and thus, from (6.15) we can see that $\tilde{\Phi}$ and $\tilde{\Psi}$ satisfy

$$\begin{aligned}\partial_t\tilde{\Phi}(\hat{\boldsymbol{U}})+&\partial_x\tilde{\Psi}(\hat{\boldsymbol{U}})\\ &=\partial_t D\phi(\hat{\boldsymbol{U}})\cdot\hat{\boldsymbol{U}}+\partial_x D\psi(\hat{\boldsymbol{U}})\cdot\hat{\boldsymbol{U}}\\ &=(\boldsymbol{g}(\hat{\boldsymbol{U}})_t+\boldsymbol{h}(\hat{\boldsymbol{U}})_x)\cdot\hat{\boldsymbol{U}}=0.\end{aligned}$$

We have $\boldsymbol{U}=\boldsymbol{g}(\hat{\boldsymbol{U}})$ and $\boldsymbol{F}(\boldsymbol{U})=\boldsymbol{h}(\hat{\boldsymbol{U}})$, as in Remark 6.5. From (6.18), we can see that $\tilde{\Phi}(\hat{\boldsymbol{U}})$ is strictly convex in the variable $\boldsymbol{U}$ by remembering that $D\boldsymbol{g}(\hat{\boldsymbol{U}})$ is positive definite:

$$D^2_{\boldsymbol{U}}\tilde{\Phi}(\hat{\boldsymbol{U}})=D_{\boldsymbol{U}}\hat{\boldsymbol{U}}=(D\boldsymbol{g}(\hat{\boldsymbol{U}}))^{-1}>0.$$

We also have

$$D_{\boldsymbol{U}}\tilde{\Psi}(\hat{\boldsymbol{U}})=D_{\boldsymbol{U}}\tilde{\Phi}(\hat{\boldsymbol{U}})D\boldsymbol{F}(\boldsymbol{U}). \tag{6.20}$$

We can now see that $\tilde{\Phi}(\hat{\boldsymbol{U}})=\Phi(\boldsymbol{U})$, with $\tilde{\Psi}(\hat{\boldsymbol{U}})=\Psi(\boldsymbol{U})$, is an entropy-entropy flux pair for the system. □

Remark 6.8 *Proposition* 6.7 *is also applicable to relaxation systems. Adding a source term to* (6.14) *will not have an effect on the proof of Proposition* 6.7 *as it only concerns the forms of* $\boldsymbol{g}(\hat{\boldsymbol{U}})$ *and the flux term* $\boldsymbol{h}(\hat{\boldsymbol{U}})$.

6.2.2 The vanishing viscosity approach

From Chapter 3, we know that there is an equivalence between the solutions of systems of hyperbolic conservation laws, for which there exists strictly convex entropies, and the solutions, of the same systems, resulting from the vanishing viscosity approach.

Assuming that the system (6.14) is strictly hyperbolic, the limit solution of the vanishing viscosity approach, in the space of functions of bounded variation, for

$$\boldsymbol{U}^\varepsilon_t+\boldsymbol{F}(\boldsymbol{U}^\varepsilon)_x=\varepsilon\partial_x\left(\boldsymbol{D}(\boldsymbol{U})\boldsymbol{U}^\varepsilon_x\right),$$

where $\boldsymbol{D}(\boldsymbol{U})\geq 0$, does not depend on the form of the vanishing viscosity term [9]. When $\boldsymbol{D}(\boldsymbol{U})\geq 0$, the parabolic system is dissipative. Such systems, with corresponding initial conditions, are well-posed under broad assumptions [15, Ch. 6.4]. This independence of the vanishing viscosity term is not valid for nonconservative systems.

Bianchini and Bressan [5] shows that the vanishing viscosity limit when $\boldsymbol{D} = \boldsymbol{I}$, for homogeneous systems in both conservative and nonconservative form, with initial data of small total variation, is unique. Also, in the conservative case, this limit solution is the same as the entropy solution of the corresponding limiting hyperbolic system.

We will later see that the correspondence between the existence of a strictly convex entropy and the vanishing viscosity approach, valid for conservative systems, does not hold for nonconservative systems. This has to do with the path-dependence of these systems. The solution of nonconservative systems will therefore depend on the form of the vanishing viscosity term. Notice that only the path defined by $\boldsymbol{D} = \boldsymbol{I}$ is studied in [5].

6.2.3 The subcharacteristic condition

In this section, we explicitly look at relaxation systems as the subcharacterstic condition, defined in Definition 3.2, does not make sense for homogeneous systems of conservation laws.

For hyperbolic relaxation systems in conservation form, as in (6.1), the subcharacteristic condition is satisfied if the full relaxation system (6.1) is endowed with a strictly convex entropy [22]. Hence, the subcharacteristic condition, as it is an easy condition to check, is a useful criterion. If the criterion is not fulfilled, we know that there does not exist a strictly convex entropy, satisfying all conditions in Definition 6.1, for the system. Also, as the subcharacteristic condition is a weak stability criterion [8], the system may lack some useful stability properties, such as the dissipative property *ii)* in Definition 6.1.

As mentioned, the existence of a convex entropy for the full relaxation system also implies that there exists a convex entropy for the corresponding equilibrium system. We state the theorem concerning entropy and the subcharacteristic condition, which is presented by Chen et al. [22].

Theorem 6.9 *Let there exist a matrix $\mathcal{Q}$ such that each equilibrium of* (6.1), *$h(\boldsymbol{u}) = \boldsymbol{U}$, is uniquely determined by $\boldsymbol{u} = \mathcal{Q}\boldsymbol{U}$. Further, let there exist a strictly convex entropy as in Definition* 6.1 *for the system. Then there exists a strictly convex entropy pair (ϕ, ψ) for the equilibrium approximation*

$$\boldsymbol{u}_t + \boldsymbol{f}(\boldsymbol{u})_x = 0 \tag{6.21}$$

of (6.1) *with* $\boldsymbol{f}(\boldsymbol{u}) = \mathcal{Q}\boldsymbol{F}(h(\boldsymbol{u}))$. *The entropy pair is defined on the equilibrium manifold:*

$$\phi(\boldsymbol{u}) = \Phi(h(\boldsymbol{u})), \quad \psi(\boldsymbol{u}) = \Psi(h(\boldsymbol{u})) \tag{6.22}$$

Further, the equilibrium system is hyperbolic and the relaxation system (6.1) *satisfies the subcharacteristic condition.*

Proof. The proof is from Chen et al. [22].

The existence of the strictly convex entropy pair (6.22) is proven by using Legendre transforms. We restrict ourselves to proving the part involving the subcharacteristic condition. Thus, we simply assume that the entropy pair (6.22) exists for the equilibrium system.

As both the corresponding homogeneous system and the equilibrium system are endowed with strictly convex entropies, they are symmetrizable by Proposition 6.2. Then, from the min-max theorem [43, p. 75], we know that the velocities of the homogeneous system, Λ_k, and the velocities of the equilibrium system, λ_i, can be determined by the critical values of the following Rayleigh quotients:

$$\Lambda_k = \min_{\mathcal{W}\subset\mathbb{R}^N} \left\{ \max_{\boldsymbol{W}\in\mathcal{W}} \left\{ \frac{\boldsymbol{W}^T \partial_{\boldsymbol{UU}}\Phi(h(\boldsymbol{u}))\partial_{\boldsymbol{U}}\boldsymbol{F}(h(\boldsymbol{u}))\boldsymbol{W}}{\boldsymbol{W}^T \partial_{\boldsymbol{UU}}\Phi(h(\boldsymbol{u}))\boldsymbol{W}} \right\} \middle| \, \dim\mathcal{W} = k \right\} \tag{6.23}$$

and

$$\lambda_j = \min_{\mathcal{V}\subset\mathbb{R}^n} \left\{ \max_{\boldsymbol{w}\in\mathcal{V}} \left\{ \frac{\boldsymbol{w}^T \partial_{\boldsymbol{uu}}\phi(\boldsymbol{u})\partial_{\boldsymbol{u}}\boldsymbol{f}(\boldsymbol{u})\boldsymbol{w}}{\boldsymbol{w}^T \partial_{\boldsymbol{uu}}\phi(\boldsymbol{u})\boldsymbol{w}} \right\} \middle| \, \dim\mathcal{V} = i \right\}, \tag{6.24}$$

where $n < N$. In the above we have

$$\partial_{\boldsymbol{uu}}\phi(\boldsymbol{u}) = (h_{\boldsymbol{u}}(\boldsymbol{u}))^T \partial_{\boldsymbol{UU}}\Phi(h(\boldsymbol{u}))h_{\boldsymbol{u}}(\boldsymbol{u})$$

and

$$\begin{aligned}\partial_{\boldsymbol{uu}}\phi(\boldsymbol{u})\partial_{\boldsymbol{u}}\boldsymbol{f}(\boldsymbol{u}) &= \partial_{\boldsymbol{uu}}\phi(\boldsymbol{u})\mathcal{Q}\partial_{\boldsymbol{u}}\boldsymbol{F}(h(\boldsymbol{u}))h_{\boldsymbol{u}} \\ &= (h_{\boldsymbol{u}})^T \partial_{\boldsymbol{UU}}\Phi(h(\boldsymbol{u}))\partial_{\boldsymbol{u}}\boldsymbol{F}(h(\boldsymbol{u}))h_{\boldsymbol{u}},\end{aligned}$$

since $\Psi_{\boldsymbol{U}}(h(\boldsymbol{U})) = \psi_{\boldsymbol{u}}(\boldsymbol{u})\mathcal{Q}$. As the values λ_i for all $1 \le i \le n$ are determined as critical values of a restriction of (6.23), they will satisfy $\lambda_i(\boldsymbol{u}) \in [\Lambda_i(h(\boldsymbol{u})), \Lambda_{i+N-n}(h(\boldsymbol{u}))]$ and hence, the system satisfies the subcharacteristic condition. □

There does not necessarily exist a convex entropy for a system that fulfills the subcharacteristic condition, as we can easily see by using the example system (6.10). From Chapter 5, we know that the roots of the corresponding homogeneous system are

$$z_1 = -1, \quad z_2 = 0, \quad \text{and} \quad z_3 = 1,$$

and that the roots of the equilibrium system are

$$q_1 = -r \quad \text{and} \quad q_2 = r.$$

By this, we can see that the system satisfies the subcharacteristic condition for $r^2 \leq 1$. But we have already shown that there does not exist any strictly convex entropy for this system for any values of r. Hence, the subcharacteristic condition alone does not imply the existence of a strictly convex entropy for general conservative relaxation systems.

For the nonlinear 2×2 conservative relaxation system, there exists a strictly convex entropy, if and only if the strict subcharacteristic condition, together with the existence of a strictly convex entropy for the corresponding equilibrium system, holds [22]. It is uncertain if this holds for $N \times N$-systems.

6.3 Nonconservative systems

In this section we will look at nonconservative hyperbolic systems and compare some of their properties, as far as it is possible, to the corresponding properties for conservative systems. As we will see, the theory for nonconservative systems is more comprehensive than the theory for conservative ones. Some of the discussion will, therefore, only be on a hand-waving level.

As mentioned in Chapter 3, the hyperbolic relaxation system

$$\partial_t \boldsymbol{U} + \boldsymbol{P}(\boldsymbol{U})\partial_x \boldsymbol{U} = \frac{1}{\varepsilon}\boldsymbol{Q}(\boldsymbol{U}), \tag{6.25}$$

is nonconservative when the term $\boldsymbol{P}(\boldsymbol{U})$ cannot be expressed as the Jacobian of some flux-term $\boldsymbol{F}(\boldsymbol{U})$. The term $\boldsymbol{P}(\boldsymbol{U})$ is assumed to be smooth in $\boldsymbol{U}$ and $\boldsymbol{U} \in G$, where G still is assumed to be a convex space.

Nonconservative hyperbolic systems are used in the modeling of, for example, multiphase flow [4, 48] and elastoplastic materials [24]. Due to the nonconservative product $\boldsymbol{P}(\boldsymbol{U})\partial_x \boldsymbol{U}$, the theory for nonconservative systems differ from the one for conservative systems. Therefore, we have to start at the beginning and define the weak solutions for these systems.

6.3.1 Defining a weak solution

Both existence of the Cauchy problem [30] and uniqueness for bounded variation solutions [32] is shown for nonconservative systems. But, due to the nonconservative product, the definition of weak solutions for nonconservative systems [37] is both different from and far more comprehensive than the definition of weak solutions for conservative ones.

As nonlinear hyperbolic systems in general develop shocks in finite time, we need to define the solutions of the nonconservative systems in a weak sense. Since the product $\boldsymbol{P}(\boldsymbol{U})\partial_x \boldsymbol{U}$ is a nonconservative product of distributions, we cannot define the weak solutions in the usual distributional sense. But, under some conditions on $\boldsymbol{U}$ and the assumption that there exists a fixed family of paths for the system, Del Maso et al. [37] defined the nonconservative product as a bounded Borel measure. This measure was then used to define weak solutions for homogeneous nonconservative systems. We will now state this definition.

We assume that $\boldsymbol{U}$ is of bounded variation on an interval (a, b) and that $\boldsymbol{P}(\boldsymbol{U})$, with values in $\mathbb{R}^{N\times N}$, is a smooth function with real and distinct eigenvalues. We further assume that a fixed family of paths is given by a Lipschitz continuous function $\phi : [0,1] \times \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^N$, which satisfies

$$\phi(0, \boldsymbol{U}_-, \boldsymbol{U}_+) = \boldsymbol{U}_-, \quad \phi(1, \boldsymbol{U}_-, \boldsymbol{U}_+) = \boldsymbol{U}_+ \quad \text{and} \quad \phi(s, \boldsymbol{U}, \boldsymbol{U}) = \boldsymbol{U},$$

and that there exists a $k > 0$ such that for all $\boldsymbol{U}_-, \boldsymbol{U}_+ \boldsymbol{V}_-, \boldsymbol{V}_+ \in \mathbb{R}^N$ and for all $s \in [0,1]$, we have

$$\begin{aligned} |\partial_s \phi(s; \boldsymbol{U}_-, \boldsymbol{U}_+)| &\leq k\,|\boldsymbol{U}_- - \boldsymbol{U}_+|\,, \\ |\partial_s \phi(s; \boldsymbol{U}_-, \boldsymbol{U}_+) - \partial_s \phi(s; \boldsymbol{V}_-, \boldsymbol{V}_+)| &\leq k\left(|\boldsymbol{U}_- - \boldsymbol{U}_+| + |\boldsymbol{V}_- - \boldsymbol{V}_+|\right). \end{aligned}$$

Then there exists a unique real-valued Borel measure μ on (a, b) satisfying

$$\mu(B) = \int_B \boldsymbol{P}(\boldsymbol{U})\partial_x \boldsymbol{U} \frac{d\boldsymbol{U}}{dx}, \tag{6.26}$$

whenever $\boldsymbol{U}$ is continuous on a Borel set $B \subset (a, b)$, and

$$\mu(\{x_0\}) = \int_0^1 \boldsymbol{P}(\phi(s; \boldsymbol{U}(x_0-), \boldsymbol{U}(x_0+)))\partial_s \phi(s; \boldsymbol{U}(x_0-), \boldsymbol{U}(x_0+))ds, \tag{6.27}$$

when $\boldsymbol{U}$ is discontinuous at a point $x_0 \in (a, b)$. The measure μ is the nonconservative product of $\boldsymbol{P}(\boldsymbol{U})$ and $\partial_x \boldsymbol{U}$ and it is usually denoted by

$$\mu = [\boldsymbol{P}(\boldsymbol{U})\partial_x \boldsymbol{U}]_\phi \,, \tag{6.28}$$

where ϕ defines the family of paths. See [37, 31] for proofs and a more thorough description.

Remark 6.10 *If $\boldsymbol{P}(\boldsymbol{U}) = D\boldsymbol{F}(\boldsymbol{U})$ for some flux-term $\boldsymbol{F}$, we have*

$$[D\boldsymbol{F}(\boldsymbol{U})\partial_x \boldsymbol{U}]_\phi = \partial_x \boldsymbol{F}(\boldsymbol{U}), \tag{6.29}$$

where the left hand side is to be understood in the sense of distributions.

We can now define weak solutions for solutions $\boldsymbol{U}$ of bounded variation [31]:

Definition 6.11 (Nonconservative weak solutions) *We say that a function $\boldsymbol{U}$ of bounded variation is a weak solution to the nonconservative system*

$$\partial_t \boldsymbol{U} + \boldsymbol{P}(\boldsymbol{U})\partial_x \boldsymbol{U} = 0 \tag{6.30}$$

if

$$\boldsymbol{U}_t + [\boldsymbol{P}(\boldsymbol{U})\partial_x \boldsymbol{U}]_\phi = 0 \tag{6.31}$$

is a bounded Borel measure on $\mathbb{R} \times (0, \infty)$.

The main difficulty when considering nonconservative systems with weak solutions as defined above, is to choose the right family of paths. Different family of paths will lead to different solutions. Hence, it can be difficult to find the right solution for nonconservative systems.

6.3.2 Modeling hyperbolic relaxation systems

As a detour, we show that nonconservative homogeneous hyperbolic systems can be used to model hyperbolic relaxation systems on conservative or nonconservative form. By reformulating the relaxation system in this way, we get rid of the source term at the cost of handling a nonconservative system with an extra equation. The approach of modeling a relaxation system as a nonconservative homogeneous system has been used when studying and developing numerical methods for relaxation systems, see Gosse [26].

We now briefly show how the transformation from relaxation system to nonconservative homogeneous system may be done. We let $q(x)$ be a function satisfying $q_x(x) = 1$ and $q_t = 0$. We can then write (6.25) as

$$\begin{aligned} \partial_t \boldsymbol{U} + \boldsymbol{P}(\boldsymbol{U})\partial_x \boldsymbol{U} - \frac{1}{\varepsilon}\boldsymbol{Q}(\boldsymbol{U})q_x &= 0, \\ \partial_t q &= 0. \end{aligned} \tag{6.32}$$

With the variable $\tilde{\boldsymbol{U}} = (\boldsymbol{U}, q)$, we can further rewrite (6.25) to the $(N+1) \times (N+1)$-system

$$\partial_t \tilde{\boldsymbol{U}} + \boldsymbol{B}(\tilde{\boldsymbol{U}})\partial_x \tilde{\boldsymbol{U}} = 0, \tag{6.33}$$

where

$$\boldsymbol{B} = \tilde{\boldsymbol{P}}(\boldsymbol{U}) - \frac{1}{\varepsilon}\tilde{\boldsymbol{Q}}(\boldsymbol{U}), \tag{6.34}$$

and

$$\tilde{\boldsymbol{P}}(\boldsymbol{U}) = \begin{pmatrix} \boldsymbol{P}(\boldsymbol{U}) & 0 \\ 0 & 0 \end{pmatrix}, \quad \tilde{\boldsymbol{Q}}(\boldsymbol{U}) = \begin{pmatrix} 0 & \boldsymbol{Q}(\boldsymbol{U}) \\ 0 & 0 \end{pmatrix}. \tag{6.35}$$

6.3.3 Symmetric form

We can define the symmetric forms for nonconservative systems in the same way as for conservative systems. Notice that the symmetric form (6.15) is conservative. Hence, nonconservative systems that can be written in a symmetric form can be rewritten in a conservative form simply by a change of variables.

Remark 6.12 *If $\boldsymbol{P}(\boldsymbol{U})$ itself is symmetric, it is the Jacobian of some flux term by the Poincaré Lemma* [40, Theorem 6.2.8]*, and hence, the system is conservative.*

Let $\Phi(\boldsymbol{U})$ be a strictly convex function. We multiply (6.30) with $D^2\Phi(\boldsymbol{U})$ from the left:

$$D^2\Phi(\boldsymbol{U})\partial_t \boldsymbol{U} + D^2\Phi(\boldsymbol{U})\boldsymbol{P}(\boldsymbol{U})\partial_x \boldsymbol{U} = 0. \tag{6.36}$$

We know that $D^2\Phi(\boldsymbol{U})$ is symmetric and positive definite. If $D^2\Phi(\boldsymbol{U})\boldsymbol{P}(\boldsymbol{U})$ is symmetric as well, we know from the Poincaré Lemma, that the system can be written in the symmetric and conservative form

$$\partial_t D\Phi(\boldsymbol{U}) + \partial_x \boldsymbol{h}(\boldsymbol{U}) = 0 \tag{6.37}$$

for some flux $\boldsymbol{h}(\boldsymbol{U})$. Hence, if there exists a symmetrizer in the form $D^2\Phi(\boldsymbol{U})$ for a nonconservative system, the system can in fact be written in a conservative form (6.37). As $\Phi(\boldsymbol{U})$ is convex, $D\Phi(\boldsymbol{U})$ is strictly increasing. Thus, there is a one to one relation between $D\Phi(\boldsymbol{U})$ and $\boldsymbol{U}$ such that we can find $\boldsymbol{U}$ as a function of $D\Phi(\boldsymbol{U})$. Let us denote $D\Phi(\boldsymbol{U}) = \hat{\boldsymbol{U}}$. Then we can rewrite (6.37) in the following conservative form:

$$\partial_t \hat{\boldsymbol{U}} + \partial_x \hat{\boldsymbol{h}} \hat{\boldsymbol{U}} = 0. \tag{6.38}$$

6.3.4 Nonconservative entropy conditions

It is possible to impose entropy conditions for nonconservative systems as well. As already mentioned, due to path-dependence, the conditions will have an impact on the form of the unique weak solutions of the system.

First, we look at the vanishing viscosity approach. For nonconservative systems, the choice of paths depend on the form of the regularization $\boldsymbol{D}$ of the system, meaning that the weak solutions of (6.30) will depend on the form of the regularization. On a formal level, LeFloch [31] showed that smooth solutions $\boldsymbol{U}^\varepsilon$ to a Cauchy problem of

$$\boldsymbol{U}_t^\varepsilon + \boldsymbol{P}(\boldsymbol{U}^\varepsilon)\boldsymbol{U}_x^\varepsilon = \varepsilon \partial_x \left(\boldsymbol{D}(\boldsymbol{U}) \boldsymbol{U}_x^\varepsilon \right) \tag{6.39}$$

converge to weak solutions of

$$\boldsymbol{U}_t + \left[\boldsymbol{P}(\boldsymbol{U}) \partial_x \boldsymbol{U} \right]_{\phi_D} = 0,$$

where the path ϕ_D depends on the form of $\boldsymbol{D}(\boldsymbol{U})$.

One can also impose convex entropy criterions for nonconservative systems, where, in the same way as for conservative systems, the solutions $\boldsymbol{U}$ have to satisfy the additional conservation law

$$\Phi(\boldsymbol{U})_t + \Psi(\boldsymbol{U}) \leq 0. \tag{6.40}$$

Notice that the entropy condition is in a conservative form.

The convex entropy defined for conservative systems in Definition 6.1 is not valid for the nonconservative systems (6.25) and (6.30). Even though the sum on the right hand side in

$$D^2\boldsymbol{\Psi}(\boldsymbol{U}) = D^2\Phi(\boldsymbol{U})\boldsymbol{P}(\boldsymbol{U}) + D\Phi(\boldsymbol{U})D\boldsymbol{P}(\boldsymbol{U}) \tag{6.41}$$

has to be symmetric for an entropy-flux to exist, the term $D\Phi(\boldsymbol{U})D\boldsymbol{P}(\boldsymbol{U})$ may not be symmetric. Hence, $D^2\Phi(\boldsymbol{U})\boldsymbol{P}(\boldsymbol{U})$ may not be symmetric either. Therefore, *i)* in Definition 6.1 cannot ensure the existence of an entropy-entropy flux pair for the nonconservative system. To ensure the existence of an entropy-entropy flux pair for the nonconservative system, the condition

$$D\Psi(\boldsymbol{U}) = D\Phi(\boldsymbol{U})\boldsymbol{P}(\boldsymbol{U}) \tag{6.42}$$

has to, explicitly, hold.

If the nonconservative system (6.30) is endowed with a strictly convex entropy, the weak solution of the entropy equation can be defined in the normal distributional sense. It then follows from the calculations in Section 3.4.4 that we can find the right path for the vanishing viscosity approach from the entropy equation by letting $\boldsymbol{D} = \boldsymbol{I}$. In general, there may not exist a strictly convex entropy for the nonconservative system. Then the search for the right path may be more extensive.

Notice that if the nonconservative system can be written in a symmetric form (6.15), Proposition 6.7 shows that it is possible to find a convex entropy for the system when in symmetric form. As already mentioned, the system is then also in a conservative form. From the above, we know that the existence of a convex entropy, with corresponding entropy flux, for nonconservative systems does in general not imply that the system can be written in a symmetric form. But, if the convex entropy also implies that $D^2\Phi(\boldsymbol{U})\boldsymbol{P}(\boldsymbol{U})$ is symmetric, we have shown that the nonconservative system can be written in a symmetric conservative form (6.38) with the new variable $\hat{\boldsymbol{U}}$ instead of $\boldsymbol{U}$.

6.3.5 The subcharacteristic condition

The proof of Theorem 6.9 relies on the symmetry of the term $D^2\Phi(\boldsymbol{U})\boldsymbol{P}(\boldsymbol{U})$. Thus, as the existence of a convex entropy alone is not enough to symmetrize a nonconservative system, the proof of Theorem 6.9 cannot be applied to nonconservative systems in general. The subcharacteristic condition may therefore not hold in general for nonconservative systems endowed with convex entropies.

If there exists a strictly convex function $\Phi(\boldsymbol{U})$ such that $D^2\Phi(\boldsymbol{U})\boldsymbol{P}(\boldsymbol{U})$ is symmetric, for a nonconservative relaxation system, we know that the system can be rewritten in a conservative symmetric form. Further, we know that there does exist a strictly convex entropy, $\hat{\Phi}(\hat{\boldsymbol{U}})$, with a corresponding entropy flux for the symmetric hyperbolic system. If the entropy for the conservative symmetric system then satisfies all conditions in Definition 6.1, this system will satisfy all conditions in Theorem 6.9 and hence, satisfy the subcharacteristic condition.

By the duality of the Legendre transform, we can see from the proof of Proposition 6.7 that the old variable $\boldsymbol{U}$ satisfies $\boldsymbol{U} = D\hat{\Phi}(\hat{\boldsymbol{U}})$. Also, $\hat{\boldsymbol{U}} = D\Phi(\boldsymbol{U})$.

From the discussion above, we can conclude with the following proposition for relaxation systems in a nonconservative form.

Proposition 6.13 *Consider the nonconservative relaxation system* (6.25)*. If there exists a strictly convex function* $\Phi(\boldsymbol{U})$ *such that*

$$D^2\Phi(\boldsymbol{U})\boldsymbol{P}(\boldsymbol{U}) \tag{6.43}$$

is symmetric, the system can be rewritten in a symmetric conservative form,

$$\partial_t \hat{\boldsymbol{U}} + \partial_x \hat{\boldsymbol{h}}(\hat{\boldsymbol{U}}) = \frac{1}{\varepsilon}\hat{\boldsymbol{Q}}(\hat{\boldsymbol{U}}), \tag{6.44}$$

There will then exist a strictly convex entropy $\hat{\Phi}(\hat{\boldsymbol{U}})$*, with corresponding entropy flux, for the symmetric conservative system. This entropy will satisfy*

$$\boldsymbol{U} = D\hat{\Phi}(\hat{\boldsymbol{U}}). \tag{6.45}$$

Further, if the entropy $\hat{\Phi}(\hat{\boldsymbol{U}})$ *satisfies all conditions in Definition* 6.1*, the conservative symmetric relaxation system will satisfy the subcharacteristic condition.*

6.4 Summary

We have looked at hyperbolic relaxation systems and homogeneous systems in connection with mathematical entropies. We have shown that the existence of a strictly convex entropy for constant-coefficient systems is sufficient, but not necessary, for stiffly well-posedness of the Cauchy problem in $L^2(\mathbb{R})$. A weaker stability assumption is sufficient for these systems. We have further shown for constant-coefficient systems that there is an equivalence between the existence of a strictly convex quadratic entropy and the dissipative mechanism (6.7) for the source term.

It was proven that a system of nonlinear homogeneous conservation laws can be written in a symmetric form if and only if there exists a strictly convex entropy for the system. We also showed that the same connection between entropy and symmetry does not hold for nonconservative systems in general. If the entropy implies that the nonconservative system is symmetrizable, the system can be rewritten in a conservative form.

A nonconservative product called for a different definition of weak solutions for nonconservative systems. By assuming that the nonconservative product can function as a Borel measure, it is possible to define weak solutions for a solution $\boldsymbol{U}$ of bounded variation. The unique solution of the system depends on the path chosen. It can therefore prove to be difficult to find the path that gives the *right* weak solution.

We also showed that the existence of a strictly convex entropy for conservative nonlinear relaxation systems implies the subcharacteristic condition. The same may not hold in general for nonconservative systems. But, if there exists a convex entropy $\Phi(\boldsymbol{U})$ s.t. $D^2\Phi(\boldsymbol{U})\boldsymbol{P}(\boldsymbol{U})$ is symmetric, we know that there exists a convex entropy for the corresponding symmetric system. If the entropy for the symmetric system is as in Definition 6.1, the symmetric conservative system will satisfy the subcharacteristic condition.

Hyperbolic relaxation systems can also be modeled as homogeneous nonconservative systems. This way of modeling relaxation systems can be useful in numerical analysis.

There does also exist conservative hyperbolic relaxation systems that are not endowed with globally defined strictly convex entropies. For these systems we cannot straightforwardly use entropy conditions when proving existence and uniqueness. We will look at such a system in the next chapter.

7 A two-phase model with well-reservoir interaction

The solution of a conservative hyperbolic relaxation system may not relax towards the solution of the corresponding equilibrium system. In this chapter we will look at a system where this seems to be the case. The hyperbolic relaxation system is a gas-liquid two-phase model with a pressure dependent well-reservoir interaction term.

We find that the velocities of the full hyperbolic relaxation system do not necessarily interlace the velocities of the equilibrium model, which eliminates the existence of a globally defined strictly convex entropy for the system. As it seems like the entropy method cannot straightforwardly be used, we add a viscous term to the system to obtain more regularity. An existence result for a class of weak solutions for this model is presented. This existence result ensures that there will exist gas and liquid at any spatial point for any finite time, which prevents the pressure form tending to ∞.

We let the relaxation time go to zero and study the corresponding formal equilibrium limit. As the estimates for the full model are relaxation parameter dependent, we have to obtain new estimates for the equilibrium model to ensure existence of solutions. With some a priori assumptions and some restrictions on the parameters in the model, we develop estimates for the equilibrium system. These estimates, together with some smallness assumptions on the initial data, are then used to obtain existence of solutions in suitable Sobolev spaces in the same manner as for the full model.

7.1 The model

We start by presenting the full model with viscosity. We will later look at the strictly hyperbolic relaxation system without viscosity. The full system is a compressible two-phase model with a pressure dependent well interaction term. This system is presented in Evje [18] and is relevant for modeling flow scenarios in oil wells. It takes the following form:

$$\begin{aligned} \partial_t n + \partial_x(nu) &= q_w[n(P_w - P(n,\rho))], \\ \partial_t \rho + \partial_x(\rho u) &= q_w[n(P_w - P(n,\rho))], \\ \partial_t(\rho u) + \partial_x(\rho u^2 + P(n,\rho)) &= \partial_x[\nu(n,\rho)\partial_x u]. \end{aligned} \tag{7.1}$$

Herein, n is the fractional gas mass, ρ the total mass, u the fluid velocity, P the well pressure and q_w a relaxation constant. The viscosity term is denoted ν and satisfies

$$\nu = \frac{(n/m+1)(n+m)^\beta}{(\rho_l - m)^{\beta+1}}, \quad \beta \in (0, 1/3), \tag{7.2}$$

where $m = \rho - n$ is the fractional liquid mass. The reference pressure P_w determines if the gas is flowing into the well, $P < P_w$, or out of the well, $P > P_w$.

In [18], the model is studied in a free boundary setting. The total mass ρ and the gas mass n are of compact support in the interval $[a, b]$ initially. We denote $a(t)$ and $b(t)$ as the particle paths initiating from $(a, 0)$ and $(b, 0)$. These are then the free boundaries determined from the equations

$$\begin{gathered} \partial_t a(t) = u(a(t), t), \quad \partial_t b(t) = u(b(t), t), \\ (-P(n,\rho) + \nu(n,\rho)u_x)(a(t)^+, t) = 0, \quad (-P(n,\rho) + \nu(n,\rho)u_x)(b(t)^-, t) = 0. \end{gathered} \tag{7.3}$$

As the model is studied in a free boundary setting, the analysis of the model in Lagrangian coordinates may be simpler than the analysis of the system on conservative form. The system in Lagrangian coordinates is equivalent to

$$\begin{aligned} \partial_t n + (n[\rho - n])\partial_x u &= q_w[n(P_w - P(n,\rho))], \\ \partial_t \rho + (\rho[\rho - n])\partial_x u &= q_w[n(P_w - P(n,\rho))], \\ g(n,\rho)\partial_t u + \partial_x P(n,\rho) &= \partial_x(E(n,\rho)\partial_x u), \quad x \in (0,1), \end{aligned} \tag{7.4}$$

with the boundary condition

$$P = Eu_x, \quad \text{at} \quad x = 0, 1. \tag{7.5}$$

Herein,

$$g(n,\rho) = \frac{\rho}{\rho - n} \tag{7.6}$$

and the pressure $P(n,\rho)$ is

$$P(n,\rho) = \left(\frac{n}{\rho_l - [\rho - n]}\right)^{\gamma}, \quad \gamma > 1, \tag{7.7}$$

where ρ_l is the liquid density, assumed to be constant. The mixture viscosity coefficient $E(n,\rho)$, in Lagrangian coordinates, is

$$E(n,\rho) = \left(\frac{\rho}{\rho_l - [\rho - n]}\right)^{\beta+1}, \quad 0 < \beta < 1/3. \tag{7.8}$$

The derivation of the model in Lagrangian coordinates is omitted, but for the interested reader it can be found in [18]. Further, an explanation of the the derivation of the boundary conditions and of the particle path equations (7.3) is made by Evje and Karlsen [19]. Also, we refer the reader to [18] and the references therein for a closer explanation of the physical process that the system (7.4) models.

We observe that the pressure (7.7) has a singular limit when the liquid mass m goes towards ρ_l. The viscosity (7.8) is therefore chosen to have a similar form to reflect the behavior of the pressure in order to ensure that certain useful estimates can be obtained [19].

7.1.1 Existence of solutions

Assuming that the initial conditions $m_0(x)$, $n_0(x)$ and $u_0(x)$ satisfy

$$\begin{aligned} &0 < n_0(x) < \infty, \quad 0 < m_0(x) < \rho_l \\ &m_0(x), n_0(x) \in W^{1,2}((0,1)), \quad u_0(x) \in L^{2q}((0,1)) \end{aligned} \tag{7.9}$$

for $q \in \mathbb{N}$, and that $P_0(0) > P_w$, the following weak global existence result is proved for (7.4) by Evje [18] for any $T > 0$:

Theorem 7.1 *Assume that the initial conditions* m_0*,* n_0 *and* u_0 *satisfy* (7.9) *and that* $P_0(0) > P_w$*, then, for any* $T > 0$*, we have*

$$n, \rho \in L^{\infty}([0,T], W^{1,2}((0,1)), \quad n_t, \rho_t \in L^2([0,T], L^2((0,1))),$$

$$u \in L^\infty([0,T], L^{2q}((0,1))) \cap L^2([0,T], H^1(0,1)).$$

We also have

$$\begin{aligned}
&0 < \inf_{x\in[0,1]} c, \sup_{x\in[0,1]} c < 1, \quad \textit{where} \quad c = n/\rho, \\
&0 < \mu \leq m(x,t) \leq \rho_l - \mu < \rho_l < \infty, \\
&0 < \mu \inf_{x\in[0,1]} c \leq n(x,t) \leq \frac{\rho_l - \mu}{1 - \sup_{x\in[0,1]} c} \sup_{x\in[0,1]} < \infty,
\end{aligned} \tag{7.10}$$

$\forall\ x \in [0,1],\ \forall\ t \in [0,T]$ *and for a small* $\mu > 0$ *depending on initial data. Further, the equations*

$$\begin{aligned}
\partial_t n + (n[\rho - n])\partial_x u = q_w[n(P_w - P(n,\rho))], \\
\partial_t \rho + (\rho[\rho - n])\partial_x u = q_w[n(P_w - P(n,\rho))],
\end{aligned}$$

with $(n,\rho)(x,0) = (n_0(x), \rho_0(x))$ *for a.e.* $x \in (0,1)$, *and*

$$\int_0^\infty \int_0^1 [ug(n,\rho)\phi_t + [P(n,\rho) - E(n,\rho)u_x]\phi_x + q_w uh(n,\rho)[P_w - P(n,\rho)]\phi]\, dxdt$$
$$+ \int_0^1 u_0(x)g(n_0(x),\rho_0(x))\phi(x,0)dx = 0,$$

where $h(n,\rho) = n/(\rho - n)$, *hold for any smooth test function* $\phi(x,t) \in C_0^\infty([0,1] \times [0,\infty])$.

By (7.10) we know that the gas mass n and liquid mass m are pointwise bounded, meaning that there will exist both gas and liquid at any point x for any finite time t. This prevents the pressure $P(n,\rho)$ from tending to infinity and we can say that the system regulates itself in finite time.

7.1.2 The equilibrium limit

We will now derive the formal equilibrium system from the full model (7.1) as the parameter $q_w \to \infty$. The full model (7.1) has a source term depending on q_w. By the variable transformation $q_w = 1/\varepsilon$, we see that the full system (7.1) is a relaxation system with the viscous term $\partial_x[\nu(n,\rho)\partial_x u]$. We want to study the system in the relaxation limit $\varepsilon \to 0$, i.e. in the equilibrium limit, corresponding to $q_w \to \infty$. By letting $q_w \to \infty$ in (7.1), we formally end up with the equilibrium, or reduced, system

$$\begin{aligned} \partial_t m + \partial_x(mu) &= 0, \\ \partial_t \rho u + \partial_x \rho u^2 &= \partial_x(\nu \partial_x u), \end{aligned} \tag{7.11}$$

with $P(n,\rho) = P_w$. Physically, $q_w \to \infty$ corresponds to $P(n,\rho) \to P_w$ infinitely fast. As the pressure is constant, the model (7.11) is essentially a single-phase liquid model. In Lagrangian coordinates the reduced model is

$$\begin{aligned} P(n,m) &= P_w, \\ \partial_t m + m^2 \partial_x u &= 0, \\ g(m)\partial_t u &= \partial_x(E(m)\partial_x u), \quad x \in (0,1), \end{aligned} \tag{7.12}$$

where

$$m = \rho - n. \tag{7.13}$$

As $P(n,m) = P_w$, the gas mass n can be found from the liquid mass m through the equation (7.7). From now on, we will mainly study the reduced system (7.12) and develop estimates for this model in a similar manner to the development of the estimates for the full model in [18]. But first we want to study the full model (7.1) where $\varepsilon = 0$, i.e. where the viscosity term is eliminated.

7.2 A relaxation system without viscosity

In this section we study (7.1) with $\varepsilon = 0$. We will show that there does not exist a globally defined strictly convex entropy for this system. Hence, the entropy cannot straightforwardly be used as a helpful tool in the analysis of the reduced system.

We start with the relaxation model corresponding to (7.1). We let $\varepsilon = 0$ and subtract the first equation in (7.1) from the second one. Then we get a system without any viscous term. The resulting system

$$\begin{aligned} \partial_t n + \partial_x(nu) &= q_w[n(P_w - P(n,\rho))], \\ \partial_t m + \partial_x(mu) &= 0, \\ \partial_t(\rho u) + \partial_x(\rho u^2 + P(n,\rho)) &= 0, \end{aligned} \tag{7.14}$$

is a purely conservative relaxation system. We further observe that the relaxation term is of rank one.

First, we want to check if (7.14) is hyperbolic. The system (7.14) is strictly hyperbolic if the flux-function

$$\boldsymbol{F}(\boldsymbol{U}) = \begin{pmatrix} nu \\ mu \\ \rho u^2 + P(n,\rho) \end{pmatrix}, \quad \text{where} \quad U = \begin{pmatrix} n \\ m \\ \rho u \end{pmatrix}, \tag{7.15}$$

has a Jacobian $\boldsymbol{DF}(\boldsymbol{U})$ with real and distinct eigenvalues for all $\boldsymbol{U}$ in some given solution space. The Jacobian $\boldsymbol{DF}(\boldsymbol{U})$ is

$$\boldsymbol{DF}(\boldsymbol{U}) = \begin{pmatrix} \left(1-\frac{n}{\rho}\right)u & -u\frac{n}{\rho} & \frac{n}{\rho} \\ -u\frac{m}{\rho} & \left(1-\frac{m}{\rho}\right)u & \frac{m}{\rho} \\ h(n,m)-u^2 & g(n,m)-u^2 & 2u \end{pmatrix} \tag{7.16}$$

with

$$g(m,n) = \frac{\gamma n^{\gamma}}{(\rho_l - m)^{\gamma+1}} \quad \text{and} \quad h(m,n) = \frac{\gamma n^{\gamma-1}}{(\rho_l - m)^{\gamma}}. \tag{7.17}$$

The corresponding eigenvalues are

$$\Lambda_1 = u, \quad \Lambda_2 = u + Q \quad \text{and} \quad \Lambda_3 = u - Q, \tag{7.18}$$

where

$$Q = \sqrt{\gamma \frac{P}{\rho} \cdot \frac{\rho_l}{\rho_l - m}}. \tag{7.19}$$

We observe that the full system fails to be strictly hyperbolic only when $n = 0$ or $\rho_l = m$. So, as long as there exists gas at any point $x \in [a,b]$, (7.14) will be strictly hyperbolic. We have no guarantee that this will, in fact, hold, since the regularizing viscosity term is no longer present and we cannot directly apply the results for the full model (7.4), which ensure $n > 0$ and $m < \rho_l$. When $n = 0$, we can see that the system is hyperbolic as long as $u \neq 0$ by observing that (7.16) has a complete set of eigenvectors in that case.

For the reduced system (7.11) we have the following flux function with the corresponding solution vector:

$$\boldsymbol{f}(\boldsymbol{u}) = \begin{pmatrix} mu \\ \rho u^2 \end{pmatrix}, \quad \boldsymbol{u} = \begin{pmatrix} m \\ \rho u \end{pmatrix}, \tag{7.20}$$

with $P = P_w$. The corresponding Jacobian is

$$D\boldsymbol{f}(\boldsymbol{u}) = \begin{pmatrix} u\left(1 - \frac{m}{\rho}\left[1 - P_w^{1/\gamma}\right]\right) & \frac{m}{\rho} \\ -u^2\left(1 - P_w^{1/\gamma}\right) & 2u \end{pmatrix}. \tag{7.21}$$

The eigenvalues of (7.20) are

$$\lambda_1 = u \quad \text{and} \quad \lambda_2 = u\left(2 - (1 - P_w^{1/\gamma})\frac{m}{\rho}\right). \tag{7.22}$$

We see that the full and the reduced system have one root, the velocity u, in common. If we rewrite λ_2, we get

$$\lambda_2 = u + u\frac{P_w^{1/\gamma}\rho_l}{\rho} = u + uq. \tag{7.23}$$

and we can see that the reduced system fails to be strictly hyperbolic if we have zero velocity. By checking the numbers of eigenvectors, we find that the system also fails to be merely hyperbolic for $u = 0$.

For the subcharacteristic condition to hold, the inequality

$$-Q \leq uq \leq Q \tag{7.24}$$

has to be satisfied. But, we have no guarantee that it is. As u increases, this inequality may no longer hold. We also observe that when $u = 0$ and $m < \rho_l$, we have a hyperbolic relaxation system that relaxes towards a non-hyperbolic equilibrium system. From Theorem 6.9 in Chapter 6 we can then conclude that there does not exist a strictly convex entropy for the system as the velocity u gets large enough or if $u = 0$. If it did, the velocities of the full system (7.1) would interlace the velocities of the corresponding equilibrium system, i.e. (7.14) would satisfy the subcharacteristic condition. Further, the system fails to be linearly stable when u is large enough due to the results in Chapter 4. As the system fails to be endowed with an entropy that is globally strictly convex and to be linearly stable for u large enough, we need some additional structure on (7.14) to ensure enough stability.

7.3 Analysis of the reduced system

By readding the viscous term $\partial_x[\nu(n,\rho)\partial_x u]$ to the relaxation system (7.14) we increase the regularity of the system. As we have already noted in Section 7.1.1, the viscous term ensures that we have enough stability to obtain estimates that ensure existence of a class of weak solutions for the full model (7.4). These estimates, found in [18], depend on the parameter q_w and blow up as $q_w \to \infty$. It is therefore necessary to find new and independent estimates for the reduced model (7.12). With specific assumptions on the parameters and on the initial data, we will develop estimates for the reduced model in this section.

The estimates for the full model are obtained in a way that ensures the existence of both liquid and gas at any point x for any finite time as long as there is both liquid and gas at any point in the model initially. In a similar manner, we seek estimates that ensure the existence of both liquid and gas at any point x for any finite time t,

$$0 < m < \rho_l \quad \text{and} \quad n > 0,$$

for the reduced model.

As before, the reduced system takes the form

$$P(n,m) = P_w, \tag{7.25a}$$

$$\partial_t m + m^2 \partial_x u = 0, \tag{7.25b}$$

$$g(m)\partial_t u = \partial_x (E(m)\partial_x u), \quad x \in (0,1). \tag{7.25c}$$

As the pressure is constant, the fractional gas mass n is a function of the fractional liquid mass m,

$$n(m) = P_w^{1/\gamma}(\rho_l - m), \tag{7.26}$$

and

$$g(n,m) = \frac{m + n(m)}{m} = \frac{m + P_w^{1/\gamma}(\rho_l - m)}{m} = g(m). \tag{7.27}$$

The viscosity coefficient E now satisfies

$$E(m) = \left(P_w^{1/\gamma} + \frac{m}{\rho_l - m}\right)^{\beta+1} \tag{7.28}$$

and does only depend on m.

As for the full model in [18] and in a similar fashion to [27, 19], we assume that

$$m, m_t, m_x, m_{tx}, m_{txx}, m_{xx}, u, u_x, u_t, u_{xt}, u_{xx}, u_{txx}, u_{xxx} \in C^{\alpha,\alpha/2}(D_T)$$

$$
\begin{aligned}
&\text{for some} \quad \alpha \in (0,1), \\
&0 < m(x,t) < \rho_l \quad \text{on} \quad D_T = [0,1] \times [0,T],
\end{aligned}
\tag{7.29}
$$

and that the initial data m_0 and u_0 satisfy

$$
\begin{aligned}
&\inf_{[0,1]} m_0 > 0, \quad \sup_{[0,1]} m_0 < \rho_l, \\
&m_0 \in W^{2,2}((0,1)) \quad \text{and} \quad u_0 \in W^{2,2}((0,1)).
\end{aligned}
$$

We further assume that $g_0(0)$ is large enough and that we have zero velocity at the boundaries:

$$
u(0,t) = u(1,t) = 0. \tag{7.30}
$$

The objective now is to derive a priori estimates that will allow us to obtain an existence result for any fixed time $T > 0$.

Note that the reduced model is not studied in a free boundary setting as the full model (7.4) is. Also notice that $m_0 \in W^{2,2}((0,1))$ implies $g_0 = \rho_0/m_0 \in W^{2,2}((0,1))$.

To complete some of the estimates it is necessary with some a priori assumptions. We follow the same approach as Zhang and Zhu [58] by assuming certain a priori bounds that will allow us to derive all the estimates needed. In particular, we shall assume that

$$
|u_x| \le M, \tag{7.31a}
$$

$$
|g_x| \le M, \tag{7.31b}
$$

for some $M > 0$. These a priori bounds have to be closed before we are done, i.e. we have to show that under the assumptions (7.31a)-(7.31b) and under some appropriate assumptions on M, we have

$$
|g_x|, \quad |u_x| \le 1/2M.
$$

We refer to Proposition 7.14 for a precise statement of this result. Then we can conclude that (7.31a)-(7.31b) indeed holds by standard continuity arguments.

As a consequence of the estimates, the following existence result is obtained for the reduced model:

Theorem 7.2 *Let $I = (0,1)$. Assume that the initial data m_0 and u_0 satisfies*

$$\inf_{[0,1]} m_0 > 0, \quad \sup_{[0,1]} m_0 < \rho_l,$$
$$m_0 \in W^{2,2}(I) \quad \textit{and} \quad u_0 \in W^{2,2}(I).$$

Further, let the initial data $\|u_0\|_{L^2}$, $\|u_{x,0}\|_{L^2}$, $\|g_{x,0}\|_{L^2}$, $\|u_{xx,0}\|_{L^2}$ and $\|g_{xx,0}\|_{L^2}$ be sufficiently small on $[0,1]$. Also, let $g_0(0) = g(m_0(0))$ be sufficiently large. We refer to Proposition 7.14 for a precise statement of these conditions.

Then for a given $T > 0$, the following holds:

(A) We have the estimates

$$\begin{aligned}
&n, m \in L^\infty([0,T], W^{2,2}(I)),\\
&u \in L^\infty([0,T], W^{2,2}(I)) \cap L^2([0,T], W^{3,2}(I)),\\
&n_t, m_t, g_t, u_t, (gu)_t \in L^\infty([0,T], L^2(I)).
\end{aligned}$$

(B) For all $(x,t) \in [0,1] \times [0,T]$, we further have

$$\begin{aligned}
0 < \mu \le m(x,t) \le \rho_l - \mu < \rho_l < \infty,\\
0 < P_w^{1/\gamma}\mu \le n(x,t) \le P_w^{1/\gamma}(\rho_l - \mu) < \infty,\\
|u|, \ |u_x|, \ |g_x| \le M,
\end{aligned}$$

for some constant $\mu > 0$, which depends on C_2 as described in Proposition 7.7, and where $M > 0$ is a sufficiently small constant, which is related to the initial data as described in Proposition 7.14.

(C) The following equations hold:

$$\begin{aligned}
\partial_t m + m^2 \partial_x u = 0,\\
\partial_t(gu) - P_w^{1/\gamma}\rho_l(u^2)_x = \partial_x(E u_x),
\end{aligned} \tag{7.32}$$

where g is finite, with $(m(x,0), u(x,0)) = (m_0(x), u_0(x))$, for a.e. $x \in I$ and any $t \ge 0$.

7.3.1 Variable transformations

We start with variable transformations that will be useful when deriving various estimates. We denote

$$R(m) = \frac{m}{\rho_l - m} \tag{7.33}$$

and

$$Q(m) = \frac{\rho}{\rho_l - m} = P_w^{1/\gamma} + R(m). \tag{7.34}$$

Then we have

$$E(m) = Q^{\beta+1}(m). \tag{7.35}$$

We rewrite (7.25b) with (7.33) and get

$$R_t + \rho_l R^2 u_x = 0. \tag{7.36}$$

Further observe that

$$g(m)R(m) = \frac{\rho}{m}\frac{m}{\rho_l - m} = \frac{\rho}{\rho_l - m} = Q.$$

We also have $g_t = \rho_l P_w^{1/\gamma}(1/m)_t$, as $g = \rho/m$, such that

$$g_t - \rho_l P_w^{1/\gamma} u_x = 0. \tag{7.37}$$

Now we get

$$\left(\frac{1}{R}\right)_t = \frac{1}{P_w^{1/\gamma}} g_t, \tag{7.38}$$

which we integrate w.r.t t:

$$\frac{1}{R} = \frac{1}{R_0} + \frac{1}{P_w^{1/\gamma}}\left(g - g_0\right). \tag{7.39}$$

Rearranging (7.39), we see that it is the same as

$$R = \frac{R_0 P_w^{\frac{1}{\gamma}}}{P_w^{1/\gamma} - R_0 g_0 + R_0 g}. \tag{7.40}$$

Multiplying (7.40) with g, we get

$$Q = P_w^{1/\gamma}\frac{g}{g-1}. \tag{7.41}$$

Now, since $Q_t = R_t$,

$$Q_t + \rho_l\left(Q^2 - 2P_w^{1/\gamma}Q + P_w^{2/\gamma}\right)u_x = 0, \tag{7.42}$$

$$Q_t + \rho_l Q^2 u_x = \rho_l \left(2P_w^{1/\gamma} Q - P_w^{2/\gamma} \right) u_x, \tag{7.43}$$

by inserting Q into (7.36). We multiply (7.43) by $Q^{\beta-1}\beta$:

$$\left(Q^\beta\right)_t + \beta\rho_l Q^{\beta+1} u_x = \rho_l \beta \left[2P_w^{1/\gamma} Q^\beta - P_w^{2/\gamma} Q^{\beta-1} \right] u_x, \tag{7.44}$$

$$\begin{aligned} \left(Q^\beta\right)_t + \beta\rho_l Q^{\beta+1} u_x &= \beta \left[2Q^\beta - P_w^{1/\gamma} Q^{\beta-1} \right] g_t \\ &= \beta P_w^{\beta/\gamma} \left(\frac{g}{g-1} \right)^\beta \left(\frac{g+1}{g} \right) g_t \\ &= -(Q^\beta)_t (g+1)(g-1), \end{aligned} \tag{7.45}$$

since $Q_t = -P_w^{1/\gamma} g_t/(g-1)^2$. We rewrite the expression before $(Q^\beta)_t$:

$$[1 + (g+1)(g-1)] = g^2, \tag{7.46}$$

and end up with the expression

$$\frac{g^2}{\beta\rho_l} (Q^\beta)_t = -Q^{\beta+1} u_x. \tag{7.47}$$

To derive the estimates, we will use whichever transformation is most suitable.

7.3.2 The estimates

We start with deriving the usual energy estimate.

Proposition 7.3 *For* (7.25a)-(7.25c)*, we have the following energy estimate:*

$$\int_0^1 \left(\frac{g(m)}{2} u^2 \right) dx + \int_0^t \int_0^1 E(m)(u_x)^2 dx dt \le \frac{1}{2} \sup_{[0,1]} g_0 \|u_0\|_{L^2}^2 := C_1^0. \tag{7.48}$$

Proof. We multiply (7.25c) with u and integrate over $[0,1]$:

$$\int_0^1 \left(\frac{g(m)}{2} u^2 \right)_t dx - \int_0^1 g(m)_t \frac{u^2}{2} dx = uE(m)u_x|_0^1 - \int_0^1 E(m)(u_x)^2 dx. \tag{7.49}$$

We have used partial integration of the right hand side. With

$$g_t = P_w^{1/\gamma} \rho_l \left(\frac{1}{m} \right)_t = P_w^{1/\gamma} \rho_l u_x \tag{7.50}$$

and zero velocity at the boundaries, we get

$$\begin{aligned} \int_0^1 \left(\frac{g(m)}{2} u^2 \right)_t dx + \int_0^1 E(m)(u_x)^2 dx &= uE(m)u_x|_0^1 + \int_0^1 u_x \left(P_w^{1/\gamma} \rho_l \frac{u^2}{2} \right) dx \\ &= \left(uE(m)u_x + P_w^{1/\gamma} \rho_l \frac{u^3}{6} \right) \Bigg|_0^1 = 0. \end{aligned} \tag{7.51}$$

Finally, we end up with

$$\begin{aligned} &\int_0^1 \left(\frac{g(m)}{2} u^2 \right) dx + \int_0^t \int_0^1 E(m)(u_x)^2 dx dt \\ &\qquad = \int_0^1 \left(\frac{g(m_0)}{2} u_0^2 \right) dx \le \frac{1}{2} \sup_{[0,1]} g_0 \|u_0\|_{L^2}^2. \end{aligned} \tag{7.52}$$

□

With constant pressure, a lower bound for Q can easily be obtained.

Lemma 7.4 *Q has the pointwise lower bound*

$$Q(x,t) \ge P_w^{1/\gamma}. \tag{7.53}$$

Proof.

$$Q = P_w^{1/\gamma} + \frac{m}{\rho_l - m} \ge P_w^{1/\gamma}, \quad \text{as} \quad 0 \le m \le \rho_l. \tag{7.54}$$

□

A pointwise upper bound for Q will ensure that $\sup_{x \in [0,1]} m(x,t) < \rho_l$ for any t. We will now derive an upper bound for Q which depends on the a priori assumption (7.31a) and a condition on g_0. To derive the bound, we need the two following lemmas.

Lemma 7.5 *Under the assumptions in Theorem* 7.2 *and the a priori assumption* (7.31a)*, g is pointwise bounded from below and above.*

Proof. We have

$$g = g_0 + P_w^{1/\gamma}\rho_l \int_0^t u_x dt.$$

Since u_x is assumed to be bounded and $g = m + n/m$,

$$1 \le g \le g_0 + P_w^{1/\gamma}\rho_l TM.$$

□

Lemma 7.6 *Under the assumptions in Theorem* 7.2 *and the a priori assumption* (7.31a)*, we have*

$$Q(0,t) = P_w^{1/\gamma}\frac{g}{g-1} < K_2 \tag{7.55}$$

for some constant $K_2 = K_2(T, M)$.

Proof. From the assumptions in Theorem 7.2, we have

$$g(0) \ge g_0(0) - P_w^{1/\gamma}\rho_l TM \ge K_1 - P_w^{1/\gamma}\rho_l TM > 1,$$

such that

$$Q(0,t) = P_w^{1/\gamma}\frac{g}{g-1} \le P_w^{1/\gamma}\frac{1}{K_1 - P_w^{1/\gamma}\rho_l TM - 1} < K_2$$

for some constant $K_2(T, M)$. □

Proposition 7.7 *Under the assumptions in Theorem* 7.2 *and the a priori assumption* (7.31a)*, we have*

$$g^2 Q^\beta \le C_2, \tag{7.56}$$

where $C_2 = C_2(C_1^0, \|u_0\|_{L^2}, T, \sup_{[0,1]} g_0, \|g_0\|_{L^2}, M)$.

Proof. We integrate (7.47) over $[0, t]$:

$$g^2 Q^\beta(x,t) = g_0^2 Q^\beta(x,0) + \int_0^t 2gg_t Q^\beta ds - \beta\rho_l \int_0^t Q^{\beta+1} u_x ds. \tag{7.57}$$

Then we integrate (7.25c) on $[0, x]$ and use the boundary condition (7.30) and (7.37):

$$\begin{aligned} \int_0^x (g(m)u)_t dy &= P_w^{1/\gamma} \rho_l \frac{1}{2} u^2 + Eu_x(x,t) - Eu_x(0,t) \\ &= P_w^{1/\gamma} \rho_l \frac{1}{2} u^2 + Q^{\beta+1} u_x - Q^{\beta+1} u_x(0,t). \end{aligned} \tag{7.58}$$

We insert (7.58) into (7.57) and get

$$\begin{aligned} & g^2 Q^\beta(x,t) \\ &= g_0^2 Q^\beta(x,0) + \int_0^t 2gg_t Q^\beta ds \\ &\quad - \beta\rho_l \int_0^t \left(\int_0^x (g(m)u)_t dy - P_w^{1/\gamma} \rho_l \frac{1}{2} u^2 + Q^{\beta+1} u_x(0,t) \right) ds \\ &= g_0^2 Q^\beta(x,0) + \int_0^t 2gg_t Q^\beta ds - \beta\rho_l \int_0^x \int_0^t (g(m)u)_t dy ds \\ &\quad + \beta\rho_l \int_0^t P_w^{1/\gamma} \rho_l \frac{1}{2} u^2 ds - \beta\rho_l \int_0^t Q^{\beta+1} u_x(0,t) ds \\ &\leq g_0^2 Q^\beta(x,0) + \int_0^t 2g|g_t| Q^\beta ds + \beta\rho_l \int_0^x |(g(m)u)| + |g(m_0)u_0| \, dy \\ &\quad + \beta\rho_l \int_0^t P_w^{1/\gamma} \rho_l \frac{1}{2} u^2 ds - \beta\rho_l \int_0^t Q^{\beta+1} u_x(0,t) ds. \end{aligned} \tag{7.59}$$

Now

$$\int_0^x |g(m)u| + |g(m_0)u_0|\,dy \tag{7.60}$$

$$\leq \left(\int_0^1 g\ dx\right)^{1/2} \cdot \left(\int_0^1 gu^2dx\right)^{1/2} + \sup_{[0,1]} g(m_0)\|u_0\|_{L^2} \tag{7.61}$$

$$\leq C_1^0 \left(\int_0^1 g\ dx\right)^{1/2} + \sup_{[0,1]} g(m_0)\|u_0\|_{L^2}, \tag{7.62}$$

by the energy estimate (7.48) and Hölder's inequality. Since the velocity is zero at the boundary, the integral of g over $[0,1]$ is bounded:

$$\begin{aligned}\int_0^1\int_0^t g_t\ dsdx &= P_w^{1/\gamma}\rho_l \int_0^1\int_0^t u_x\ dsdx \\ &= P_w^{1/\gamma}\rho_l \int_0^t\int_0^1 u_x\ dxds = 0,\end{aligned} \tag{7.63}$$

such that

$$\int_0^1 g\ dx = \int_0^1 g_0\ dx \leq \left(\int_0^1 g_0^2\ dx\right)^{1/2} = \|g_0\|_{L^2((0,1))}. \tag{7.64}$$

The third integral of (7.59) satisfies

$$\beta P_w^{\frac{1}{\gamma}} \rho_l^2 \int_0^t \frac{1}{2}u^2ds \leq \beta\rho_l P_w^{-\beta/\gamma}\rho_l \frac{1}{2}C_1^0 \tag{7.65}$$

by

$$u^2(x,t) = \left(\int_0^x u_x ds\right)^2 \leq \left(\int_0^x |u_x| ds\right)^2 \tag{7.66}$$

$$
\begin{aligned}
&\leq \int_0^x |u_x|^2 ds \leq \frac{1}{P_w^{(\beta+1)/\gamma}} \int_0^1 E(m)(u_x)^2 ds \\
&\leq \frac{1}{P_w^{(\beta+1)/\gamma}} C_1^0,
\end{aligned}
\tag{7.67}
$$

since

$$
\frac{E}{P_w^{(\beta+1)/\gamma}} = \left(1 + \frac{m}{P_w^{1/\gamma}(\rho_l - m)}\right)^{\beta+1} \geq 1. \tag{7.68}
$$

Now we have to obtain a bound for the first integral of (7.59). Here we have to use the a priori assumption (7.31a). We have

$$
\begin{aligned}
\int_0^t 2g|g_t|Q^\beta ds &= 2P_w^{1/\gamma}\rho_l \int_0^t gQ^\beta |u_x| dt \\
&\leq 2P_w^{1/\gamma}\rho_l M \int_0^t g^2 Q^\beta dt.
\end{aligned}
\tag{7.69}
$$

The expression (7.69) can now be bounded by Grönwall's inequality if

$$
\int_0^t Q^{\beta+1} u_x(0,t) ds \tag{7.70}
$$

is bounded. The term (7.70) can be bounded by the assumptions on g_0 in Lemma 7.6:

$$
\int_0^t Q^{\beta+1} u_x(0,t) ds \leq T K_2^{\beta+1} M.
$$

We can then conclude that

$$
\begin{aligned}
&g^2 Q^\beta(x,t) \\
&\quad \leq \left[1 + 2TP_w^{1/\gamma}\rho_l M \exp\left(2TP_w^{1/\gamma}\rho_l M\right)\right] \\
&\qquad \times \left[C_1^0 \sqrt{\|g_0\|_{L^2([0,1])}} + \sup_{[0,1]} g_0 \|u_0\|_{L^2} + \beta\rho_l^2 P_w^{-\beta/\gamma} \frac{1}{2} C_1^0 + TMK_2^{\beta+1}\right]
\end{aligned}
$$

$$:=C_2(C_1^0, \|u_0\|_{L^2}, T, \sup_{[0,1]} g_0, \|g_0\|_{L^2}, M), \tag{7.71}$$

by using Grönwall's inequality on (7.69). □

With the a priori assumption (7.31a) and the result in Proposition 7.7 we can find an upper bound for $(Q^\beta)_x$ in $L^2((0,1))$.

Proposition 7.8 *Under the assumptions in Theorem* 7.2 *and the a priori assumption* (7.31a), Q_x^β *is bounded in* $L^2((0,1))$*:*

$$\int_0^1 ((Q^\beta)_x)^2 dx \leq C_4, \tag{7.72}$$

where $C_4 = C_4(C_2, T, \|u_0\|_{W^{1,2}}, M)$*.*

Proof. We insert

$$Q^{\beta+1} u_x = -\frac{g^2}{\beta \rho_l} (Q^\beta)_t \tag{7.73}$$

into (7.25c):

$$(Q^\beta)_x = (Q^\beta)_{0,x} - \int_0^t \frac{\beta \rho_l}{g} u_t - \int_0^t \frac{2g_x}{g} (Q^\beta)_t dt. \tag{7.74}$$

Multiplying by Q_x^β and integrating over $[0,1]$ w.r.t x, we get

$$\begin{aligned}\int_0^1 ((Q^\beta)_x)^2 dx = \int_0^1 (Q^\beta)_x (Q^\beta)_{0,x} - \int_0^1 (Q^\beta)_x \int_0^t \frac{\beta \rho_l}{g} u_t dt dx \\ - \int_0^1 (Q^\beta)_x \int_0^t \frac{2g_x}{g} (Q^\beta)_t dt dx. \end{aligned} \tag{7.75}$$

By Hölder's inequality, and then Cauchy's inequality we have

$$\int_0^1 ((Q^\beta)_x)^2 dx$$

$$
\begin{aligned}
&\leq \left(\int_0^1 (Q^\beta)_x)^2 dx\right)^{1/2} \times \left[\|Q^\beta_{0,x}\|_{L^2} + \left(\int_0^1 \left(\int_0^t \frac{\beta\rho_l}{g} u_t dt\right)^2 dx\right)^{1/2}\right. \\
&\quad \left. + \left(\int_0^1 \left(\int_0^t \frac{2g_x}{g}(Q^\beta)_t dt\right)^2 dx\right)^{1/2}\right] \\
&\leq \frac{1}{2}\int_0^1 (Q^\beta)_x)^2 dx + \frac{1}{2}\left[\|Q^\beta_{0,x}\|_{L^2} + \left(\int_0^1 \left(\int_0^t \frac{\beta\rho_l}{g} u_t dt\right)^2 dx\right)^{1/2}\right. \\
&\quad \left. + \left(\int_0^1 \left(\int_0^t \frac{2g_x}{g}(Q^\beta)_t dt\right)^2 dx\right)^{1/2}\right]^2 \\
&\leq \frac{1}{2}\int_0^1 (Q^\beta)_x)^2 dx + \frac{3}{2}\|Q^\beta_{0,x}\|^2_{L^2} + \frac{3}{2}\int_0^1 \left(\int_0^t \frac{\beta\rho_l}{g} u_t dt\right)^2 dx \\
&\quad + \frac{3}{2}\int_0^1 \left(\int_0^t \frac{2g_x}{g}(Q^\beta)_t dt\right)^2 dx,
\end{aligned}
$$

where Jensen's inequality is used in the last step. We have

$$
\int_0^1 \left(\int_0^t \frac{\beta\rho_l}{g} u_t dt\right)^2 dx \leq t\beta^2\rho_l^2 \int_0^1\int_0^t (u_t)^2 dt dx, \tag{7.76}
$$

since $g \geq 1$. Now

$$
\begin{aligned}
\int_0^1\int_0^t (u_t)^2 dt dx &= \int_0^t\int_0^1 \frac{u_t}{g}\partial_x(Eu_x) dx dt \\
&\leq \int_0^t\int_0^1 u_t \partial_x(Eu_x) dx dt
\end{aligned}
$$

$$
\begin{aligned}
&= \int_0^t u_t E u_x|_0^1 \, dt - \int_0^t \int_0^1 u_{xt} E u_x dx dt \\
&= - \int_0^t \int_0^1 \frac{(u_x)^2}{2}_t E dx dt \\
&\leq C_2^{\frac{\beta+1}{\beta}} \int_0^1 \frac{1}{2} \left((u_{0,x})^2 - (u_x)^2 \right) dx + \int_0^t \int_0^1 \frac{(u_x)^2}{2} |E_t| dx dt \\
&\leq \frac{C_2^{\frac{\beta+1}{\beta}}}{2} \int_0^1 (u_{0,x})^2 + (u_x)^2 dx \\
&\quad + (\beta+1)\rho_l \int_0^t C_2^{(\beta+2)/\beta} \int_0^1 \frac{(u_x)^2}{2} |u_x| dx dt \\
&\leq \frac{C_2^{\frac{\beta+1}{\beta}}}{2} \left(\|u_0\|^2_{W^{1,2}((0,1))} + M^2 \right) + T(\beta+1)\rho_l C_2^{(\beta+2)/\beta} \frac{M^3}{2},
\end{aligned}
$$

as, from (7.57),

$$
E_t = Q_t^{\beta+1} = -\rho_l(\beta+1) Q^{\beta+2} u_x. \tag{7.77}
$$

To bound the remaining term, we do the following calculations:

$$
\begin{aligned}
\int_0^1 \left(\int_0^t 2 \frac{g_x}{g} (Q^\beta)_t dt \right)^2 dx &\leq 4t \int_0^1 \left(\int_0^t \frac{(g_x)^2}{g^2} ((Q^\beta)_t)^2 dt \right) dx \\
&\leq 4t(\beta\rho_l)^2 C_2^{(\beta+1)/\beta} M^2 \int_0^1 \int_0^t \frac{(g_x)^2}{g^6} dt dx \qquad (7.78) \\
&\leq 4T\rho_l^2 C_2^{(\beta+1)/\beta} M^2 P_w^{(1-2\beta)/\gamma} \int_0^1 \int_0^t ((Q^\beta)_x)^2 dt dx.
\end{aligned}
$$

By using Grönwall's inequality the term can be bounded. □

Remark 7.9 *Note that it is also possible to find a bound for Q_x^β in $L^2((0,1))$ by simply using the a priori assumption* (7.31b). *We have*

$$
Q_x^\beta = -\beta P_w^{1/\gamma} Q^{\beta-1} \frac{g_x}{g^2},
$$

such that

$$\int_0^1 (Q_x^\beta)^2 \leq \left(\beta P_w^{\beta/\gamma} M\right)^2,$$

as $\beta < 1/3$.

With the results above, we can now derive estimates for the original variables.

Corollary 7.10 *We have*

$$\int_0^1 (m_x)^2 dx \leq C_5, \tag{7.79}$$

where $C_5 = C_5(C_2, C_4)$.

Proof. Differentiating Q^β w.r.t. x, we get

$$\begin{aligned}(Q^\beta)_x &= (\beta) Q^{\beta-1} Q'(m) m_x \\ &= (\beta) Q^{\beta-1} \frac{\rho_l}{(\rho_l - m)^2} m_x,\end{aligned} \tag{7.80}$$

such that

$$m_x = \frac{(\rho_l - m)^2}{\rho_l \beta} Q^{1-\beta} (Q^\beta)_x. \tag{7.81}$$

and

$$\begin{aligned}\int_0^1 (m_x)^2 dx &= \int_0^1 \frac{(\rho_l - m)^4}{\rho_l^2 \beta^2} Q^{2-2\beta} ((Q^\beta)_x)^2 dx \\ &\leq \frac{\rho_l^2}{\beta^2} C_2 \int_0^1 ((Q^\beta)_x)^2 dx \\ &\leq \frac{\rho_l^2}{\beta^2} C_2 \ C_4 := C_5.\end{aligned} \tag{7.82}$$

□

7.3.3 Closing the a priori assumptions

Up til now, we have not needed the a priori assumption $|g_x| \leq M$. But we will make use of it when closing the a priori assumption $|u_x| \leq M$. We use the Sobolev inequality $\|(u_x)\|_{L^\infty} \leq \|(u_x)\|_{W^{1,1}}$ and Hölder's inequality:

$$\begin{aligned}
|u_x| &\leq \int_0^1 |u_x| dx + \int_0^1 |u_{xx}|\, dx \\
&\leq \left(\int_0^1 (u_x)^2 dx\right)^{1/2} + \left(\int_0^1 (u_{xx})^2 dx\right)^{1/2} \\
&= \|u_x\|_{L^2} + \|u_{xx}\|_{L^2}. \qquad (7.83)
\end{aligned}$$

To close the a priori assumption $|u_x| \leq M$, we have to show that we can get $|u_x| < M$ with the estimates that we have obtained so far. To do this, we find upper bounds for u_x and u_{xx} in $L^2((0,1))$ as expressed in Lemma 7.12 and 7.13. To find these upper bounds, we need $|g_x| \leq M$. It turns out that the a priori assumption $|g_x| \leq M$ can be closed with the same estimates that are used to close $|u_x| \leq M$.

Remark 7.11 *Since $u(0,t) = u(1,t) = 0$, we have $u_t(0,t) = u_t(1,t) = 0$ such that $\partial_x(Eu_x)(0,t) = \partial_x(Eu_x)(1,t) = 0$.*

Lemma 7.12 *Under the assumptions in Theorem 7.2, the a priori assumption (7.31a) and the assumptions $1/2\rho_l P_w^{-\beta/\gamma} < 1$ and*

$$M < \frac{3}{4}\frac{(2P_w^{\beta/\gamma} - \rho_l)P_w^{2/\gamma}}{(\beta+1)C_2^{(\beta+2)/\beta}}, \qquad (7.84)$$

we have

$$\int_0^1 \frac{1}{2}(u_x)^2 dx + \int_0^1 \frac{1}{2g}(g_x)^2 dx + \frac{1}{4}\int_0^t \int_0^1 \frac{E}{g}(u_{xx})^2 dx dt$$
$$\leq M_1(\|g_0\|_{W^{1,2}}, \|u_0\|_{W^{1,2}}, M, C_2, T).$$

Proof. We differentiate the expressions (7.37) and $1/g\times$(7.25c) w.r.t x:

$$g_{xt} = P_w^{1/\gamma}\rho_l u_{xx}, \qquad (7.85a)$$

$$u_{xt} = \left(\frac{1}{g}\right)_x \partial_x(Eu_x) + \frac{1}{g}\partial_{xx}(Eu_x). \tag{7.85b}$$

We multiply (7.85a) with g_x/g and get

$$\begin{aligned}\left(\frac{(g_x)^2}{2g}\right)_t &= P_w^{1/\gamma}\rho_l \frac{g_x}{g}u_{xx} + \left(\frac{1}{g}\right)_t (g_x)^2 \\ &= P_w^{1/\gamma}\rho_l \frac{g_x}{g}u_{xx} - P_w^{1/\gamma}\rho_l \frac{u_x}{g^2}(g_x)^2. \end{aligned} \tag{7.86}$$

Next we multiply (7.85b) with u_x:

$$\left(\frac{(u_x)^2}{2}\right)_t = u_x\partial_x\left(\frac{1}{g}\partial_x(Eu_x)\right) \tag{7.87}$$

We add (7.86) and (7.87) together and integrate over $[0,1]$:

$$\begin{aligned}&\int_0^1 \left(\frac{(u_x)^2}{2}\right)_t dx + \int_0^1 \left(\frac{(g_x)^2}{2g}\right)_t dx \\ &\quad= \int_0^1 u_x\partial_x\left(\frac{1}{g}\partial_x(Eu_x)\right)dx \\ &\quad\quad + P_w^{1/\gamma}\rho_l \int_0^1 \frac{g_x}{g}u_{xx}dx - P_w^{1/\gamma}\rho_l \int_0^1 \frac{u_x}{g^2}(g_x)^2 dx. \end{aligned} \tag{7.88}$$

By partial integration and using the initial conditions from Remark 7.11, we have

$$\begin{aligned}\int_0^1 u_x\partial_x\left(\frac{1}{g}\partial_x(Eu_x)\right) &= u_x\frac{1}{g}\partial_x(Eu_x)\Big|_0^1 - \int_0^1 \frac{1}{g}\partial_x(Eu_x)u_{xx}dx \\ &= -\int_0^1 \frac{E}{g}(u_{xx})^2dx - \int_0^1 \frac{E_x}{g}u_xu_{xx}dx. \end{aligned} \tag{7.89}$$

We move the first expression to the left hand side:

$$\int_0^1 \left(\frac{(u_x)^2}{2}\right)_t dx + \int_0^1 \left(\frac{(g_x)^2}{2g}\right)_t dx + \int_0^1 \frac{E}{g}(u_{xx})^2dx$$

$$
\begin{aligned}
&= -\int_0^1 \frac{E_x}{g} u_x u_{xx} dx + P_w^{1/\gamma}\rho_l \int_0^1 \frac{g_x}{g} u_{xx} dx - P_w^{1/\gamma}\rho_l \int_0^1 \frac{u_x}{2g^2}(g_x)^2 dx \\
&= -\int_0^1 \frac{1}{g} u_{xx}\left(E_x u_x + P_w^{1/\gamma}\rho_l g_x\right) dx - P_w^{1/\gamma}\rho_l \int_0^1 \frac{u_x}{2g^2}(g_x)^2 dx.
\end{aligned}
\tag{7.90}
$$

By using that $|u_x| \leq M$,

$$
|E_x| = P_w^{1/\gamma}\frac{(\beta+1)Q^\beta}{(g-1)^2}|g_x| \leq P_w^{-1/\gamma}(\beta+1)Q^{\beta+2}|g_x|
$$

and the estimate from Proposition 7.7, we get

$$
\begin{aligned}
&-\int_0^1 \frac{1}{g} u_{xx}\left(E_x u_x + P_w^{1/\gamma}\rho_l g_x\right) dx - P_w^{1/\gamma}\rho_l \int_0^1 \frac{u_x}{2g^2}(g_x)^2 dx \\
&\leq \left(M(\beta+1)C_2^{(\beta+2)\beta}P_w^{-1/\gamma} + P_w^{1/\gamma}\rho_l\right)\int_0^1 |g_x|\frac{1}{g}|u_{xx}|dx \\
&\quad + P_w^{1/\gamma}\rho_l M \int_0^1 \frac{1}{2g}(g_x)^2 dx \\
&\leq \left(M(\beta+1)C_2^{(\beta+2)\beta}P_w^{-1/\gamma} + P_w^{1/\gamma}\rho_l\right) \\
&\quad \times \left(\int_0^1 \frac{(g_x)^2}{g} dx\right)^{1/2}\left(\int_0^1 \frac{1}{g}(u_{xx})^2 dx\right)^{1/2} \\
&\quad + P_w^{1/\gamma}\rho_l M \int_0^1 \frac{1}{2g}(g_x)^2 dx \\
&\leq \left(M(\beta+1)C_2^{(\beta+2)\beta}P_w^{-1/\gamma} + P_w^{1/\gamma}\rho_l\right)\frac{1}{2}\left[\int_0^1 \frac{(g_x)^2}{g}dx + \int_0^1 \frac{1}{g}(u_{xx})^2 dx\right] \\
&\quad + P_w^{1/\gamma}\rho_l M \int_0^1 \frac{1}{2g}(g_x)^2 dx,
\end{aligned}
\tag{7.91}
$$

where we have first used Cauchy's inequality and then Young's inequality. We also have

$$\int_0^1 \frac{1}{g}(u_{xx})^2 dx \le P_w^{-(\beta+1)/\gamma} \int_0^1 \frac{E}{g}(u_{xx})^2 dx. \tag{7.92}$$

With the assumptions on M and $\rho_l P_w^{-\beta/\gamma}$, we can move the expression containing (7.92) on the left hand side of the inequality:

$$\begin{aligned}
&\int_0^1 \left(\frac{(u_x)^2}{2}\right)_t dx + \int_0^1 \left(\frac{(g_x)^2}{2g}\right)_t dx + K_4 \int_0^1 \frac{E}{g}(u_{xx})^2 dx \\
&\qquad \le K_5 \int_0^1 \frac{1}{2g}(g_x)^2 dx \le K_5 \int_0^1 \frac{1}{2g}(g_x)^2 dx + K_5 \int_0^1 \frac{1}{2}(u_x)^2 dx,
\end{aligned} \tag{7.93}$$

where

$$\begin{aligned}
K_4 &= 1 - \left(M(\beta+1)C_2^{(\beta+2)/\beta} P_w^{-1/\gamma} + P_w^{1/\gamma}\rho_l\right) \frac{1}{2} P_w^{-(\beta+1)/\gamma} \\
&= 1 - \frac{1}{2}\left(\rho_l P_w^{-\beta/\gamma} + M(\beta+1)C_2^{(\beta+2)/\beta} P_w^{-(\beta+2)/\gamma}\right) > 1/4,
\end{aligned}$$

due to (7.84), and

$$K_5 = \left(\frac{M}{2}(\beta+1)\frac{C_2^{(\beta+2)\beta}}{P_w^{1/\gamma}} + \frac{3}{2} P_w^{1/\gamma}\rho_l\right). \tag{7.94}$$

We can now integrate the expression (7.93) w.r.t. t:

$$\begin{aligned}
&\int_0^1 \frac{(u_x)^2}{2} dx + \int_0^1 \frac{(g_x)^2}{2g} dx + K_4 \int_0^t \int_0^1 \frac{E}{g}(u_{xx})^2 dx dt \\
&\qquad \le \int_0^1 \frac{(u_{0,x})^2}{2} dx + \int_0^1 \frac{(g_{0,x})^2}{2g_0} dx + K_5 \int_0^t \int_0^1 \frac{(g_x)^2}{2g} dx dt \\
&\qquad \le \int_0^1 \frac{(u_{0,x})^2}{2} dx + \int_0^1 \frac{(g_{0,x})^2}{2g_0} dx + K_5 \int_0^t \int_0^1 \frac{(g_x)^2}{2g} dx dt
\end{aligned} \tag{7.95}$$

$$+ K_5 \int_0^t \int_0^1 \frac{(u_x)^2}{2} dxdt. \tag{7.96}$$

Now, the inequality above implies that

$$\begin{aligned}
&\int_0^1 \frac{(u_x)^2}{2} dx + \int_0^1 \frac{(g_x)^2}{2g} dx + \frac{1}{4} \int_0^t \int_0^1 \frac{E}{g} (u_{xx})^2 dxdt \\
&\quad \leq \left(\int_0^1 \frac{(u_{0,x})^2}{2} dx + \int_0^1 \frac{(g_{0,x})^2}{2g_0} dx \right) (1 + TK_5 \exp(TK_5)) \qquad (7.97) \\
&\quad := M_1(\|g_0\|_{W^{1,2}}, \|u_0\|_{W^{1,2}}, M, C_2, T)
\end{aligned}$$

by the use of Grönwall's inequality. □

We find a bound for u_{xx} in $L^2((0,1))$ in the following.

Lemma 7.13 *Under the assumptions in Theorem* 7.2, *the a priori assumptions* (7.31a)-(7.31b) *and the assumptions* $1/2\rho_l P_w^{-\beta/\gamma} < 1$ *and*

$$M < \frac{3}{4} \min \left\{ \frac{(2P_w^{\beta/\gamma} - \rho_l) P_w^{2/\gamma}}{(\beta+1) C_2^{(\beta+2)/\beta}}, \frac{2(1 - \rho_l/(2P_w^{\beta/\gamma})) P_w^{1/\gamma}}{(\beta+1) C_2^{1/\beta} (2 + M + C_2^{1/\beta}) + P_w^{1/\gamma}} \right\}, \tag{7.98}$$

we have

$$\begin{aligned}
&\int_0^1 (u_{xx})^2 dx + \int_0^1 \frac{(g_{xx})^2}{2g} dx + \frac{1}{4} \int_0^t \int_0^1 \frac{E}{g} (u_{xxx})^2 dxdt \\
&\qquad \leq M_2(C_1^0, M_1, \|g_{xx,0}\|_{L^2}, \|u_{xx,0}\|_{L^2}, C_2, T). \qquad (7.99)
\end{aligned}$$

Proof. We differentiate (7.37) twice w.r.t x, multiply the resulting expression by g_{xx}/g and integrate over $[0,1]$:

$$\int_0^1 \left(\frac{(g_{xx})^2}{2g} \right)_t dx = P_w^{1/\gamma} \rho_l \int_0^1 \frac{g_{xx}}{g} u_{xxx} dx - P_w^{1/\gamma} \rho_l \int_0^1 (g_{xx})^2 \frac{u_x}{2g^2} dx.$$

By using Young's inequality and $|u_x| \le M$, we get

$$
\begin{aligned}
\int_0^1 \left(\frac{(g_{xx})^2}{2g} \right)_t dx &= P_w^{1/\gamma} \rho_l \int_0^1 g_{xx} u_{xxx} dx - P_w^{1/\gamma} \rho_l \int_0^1 (g_{xx})^2 \frac{u_x}{2g^2} dx \\
&\le P_w^{1/\gamma} \rho_l \frac{1}{2} \int_0^1 \frac{(g_{xx})^2}{g} + \frac{(u_{xxx})^2}{g} dx \\
&\le P_w^{1/\gamma} \rho_l \frac{1}{2} \int_0^1 \frac{(g_{xx})^2}{g} + \frac{1}{P_w^{(\beta+1)/\gamma}} \frac{E}{g} (u_{xxx})^2 dx. \qquad (7.100)
\end{aligned}
$$

Now we multiply (7.25c) with $1/g$, differentiate twice w.r.t. x, multiply with u_{xx} and integrate over $[0,1]$:

$$
\begin{aligned}
\int_0^1 \left(\frac{(u_{xx})^2}{2} \right)_t dx &= \int_0^1 u_{xx} \partial_{xx} \left(\frac{1}{g} \partial_x (E u_x) \right) dx \\
&= u_{xx} \partial_x \left(\frac{1}{g} \partial_x (E u_x) \right) \Big|_0^1 - \int_0^1 u_{xxx} \partial_x \left(\frac{1}{g} \partial_x (E u_x) \right) dx \\
&= u_{xx} \frac{1}{g} \partial_{xx} (E u_x) \Big|_0^1 \\
&\quad - \int_0^1 u_{xxx} \frac{1}{g} \partial_{xx} (E u_x) dx + \int_0^1 u_{xxx} \frac{g_x}{g^2} \partial_x (E u_x) dx. \qquad (7.101)
\end{aligned}
$$

We rewrite the integral expressions in (7.101):

$$
\begin{aligned}
&- \int_0^1 u_{xxx} \frac{1}{g} \partial_{xx} (E u_x) dx + \int_0^1 u_{xxx} \frac{g_x}{g^2} \partial_x (E u_x) dx \\
&= - \int_0^1 u_{xxx} \frac{1}{g} \left(E_{xx} u_x + 2 E_x u_{xx} + E u_{xxx} \right) dx \\
&\quad + \int_0^1 u_{xxx} \frac{g_x}{g^2} E_x u_x dx + \int_0^1 u_{xxx} \frac{g_x}{g^2} E u_{xx} dx. \qquad (7.102)
\end{aligned}
$$

We move the term containing $(u_{xxx})^2$ over to the left:

$$\int\limits_0^1 \left(\frac{(u_{xx})^2}{2} \right)_t dx + \int_0^1 \frac{E}{g} (u_{xxx})^2 dx$$
$$= \int_0^1 u_{xxx} \frac{g_x}{g^2} E_x u_x dx + \int_0^1 u_{xxx} \frac{g_x}{g^2} E u_{xx} dx$$
$$- \int_0^1 \frac{1}{g} u_x u_{xxx} E_{xx} dx - \int_0^1 2\frac{1}{g} E_x u_{xx} u_{xxx} dx + u_{xx} \frac{1}{g} \partial_{xx}(E u_x) \Big|_0^1. \quad (7.103)$$

As $\partial_x(Eu_x) = 0$ at $x = 0, 1$, we have $Eu_{xx} = -E_x u_x$ such that the last term in (7.103) can be rewritten in the following way:

$$u_{xx} \frac{1}{g} \partial_{xx}(E u_x) \Big|_0^1 = -\frac{E_x}{E} u_x u_{xt} \Big|_0^1 = -\frac{E_x}{E} \left(\frac{u_x^2}{2} \right)_t \Big|_0^1. \quad (7.104)$$

We also have

$$E_{xx} = -P_w^{1/\gamma}(\beta + 1) \left(\frac{Q^\beta}{(g-1)^2} g_{xx} - (g_x)^2 \frac{(2g + \beta)}{g(g-1)^3} Q^\beta \right) \quad (7.105)$$

by differentiating $E = Q^{\beta+1}$, where Q is as in (7.41), twice w.r.t. x. We insert the expression into (7.103):

$$\int\limits_0^1 \left(\frac{u_{xx}^2}{2} \right)_t dx + \int_0^1 \frac{E}{g} (u_{xxx})^2 dx$$
$$= - P_w^{1/\gamma}(\beta + 1) \int_0^1 \frac{Q^\beta}{(g-1)^2} (g_x)^2 u_x u_{xxx} dx$$
$$+ \int_0^1 \frac{g_x}{g^2} E u_{xxx} u_{xx} dx + P_w^{1/\gamma}(\beta + 1) \int_0^1 \frac{1}{g} u_x u_{xxx} \frac{Q^\beta}{(g-1)^2} g_{xx} dx$$
$$- P_w^{1/\gamma}(\beta + 1) \int_0^1 \frac{1}{g} u_x u_{xxx} (g_x)^2 \frac{(2g+\beta)}{g(g-1)^3} Q^\beta dx$$
$$+ 2P_w^{1/\gamma}(\beta + 1) \int_0^1 g_x u_{xx} u_{xxx} \frac{Q^\beta}{(g-1)^2} dx - \frac{E_x}{E} \left(\frac{u_x^2}{2} \right)_t \Big|_0^1. \quad (7.106)$$

We estimate the integral terms in (7.106) by using the a priori assumption (7.31b):

$$
\begin{aligned}
&-P_w^{1/\gamma}(\beta+1)\int_0^1 \frac{Q^\beta}{(g-1)^2}(g_x)^2 u_x u_{xxx} dx \\
&\quad + \int_0^1 \frac{g_x}{g^2} E u_{xxx} u_{xx} dx + P_w^{1/\gamma}(\beta+1)\int_0^1 \frac{1}{g} u_x u_{xxx} \frac{Q^\beta}{(g-1)^2} g_{xx} dx \\
&\quad - P_w^{1/\gamma}(\beta+1)\int_0^1 \frac{1}{g} u_x u_{xxx} (g_x)^2 \frac{(2g+\beta)}{g(g-1)^3} Q^\beta dx \\
&\quad + 2P_w^{1/\gamma}(\beta+1)\int_0^1 g_x u_{xx} u_{xxx} \frac{Q^\beta}{(g-1)^2} dx \\
&\leq (\beta+1)\frac{M^2}{P_w^{1/\gamma}}\frac{1}{2} C_2^{1/\beta}\int_0^1 E\left(\frac{(u_x)^2}{g} + \frac{(u_{xxx})^2}{g}\right) dx \\
&\quad + \frac{M}{2}\int_0^1 \frac{E}{g}\left((u_{xxx})^2 + (u_{xx})^2\right) dx \\
&\quad + P_w^{-1/\gamma}(\beta+1)\frac{C_2^{1/\beta} M}{2}\int_0^1 \frac{E}{g}\left((u_{xxx})^2 + (g_{xx})^2\right) dx \\
&\quad + P_w^{-1/\gamma}(\beta+1)\frac{C_2^{2/\beta} M}{2}\int_0^1 \frac{E}{g}\left((u_x)^2 + (u_{xxx})^2\right) dx \\
&\quad + (\beta+1)\frac{M C_2^{1/\beta}}{P_w^{1/\gamma}}\int_0^1 \frac{E}{g}\left((u_{xx})^2 + (u_{xxx})^2\right) dx \\
&= K_1 \int_0^1 (u_x)^2 dx + K_2 \int_0^1 \frac{E}{g}(u_{xxx})^2 dx \\
&\quad + K_3 \int_0^1 \frac{E}{g}(u_{xx})^2 dx + K_4 \int_0^1 \frac{(g_{xx})^2}{2g} dx,
\end{aligned}
$$

where we have used Young's inequality. We add (7.100) to (7.106) and, with the estimates of the integral terms, we get

$$
\begin{aligned}
&\int_0^1 \left(\frac{(u_{xx})^2}{2}\right)_t dx + \int_0^1 \left(\frac{(g_{xx})^2}{2g}\right)_t dx + \int_0^1 \frac{E}{g}(u_{xxx})^2 dx \\
&\qquad \leq K_1 \int_0^1 E(u_x)^2 dx + \tilde{K}_2 \int_0^1 \frac{E}{g}(u_{xxx})^2 dx + K_3 \int_0^1 \frac{E}{g}(u_{xx})^2 dx
\end{aligned}
$$

$$+ (K_4 + P_w^{1/\gamma}\rho_l)\int_0^1 \frac{(g_{xx})^2}{2g}dx - \frac{E_x}{E}\left(\frac{u_x^2}{2}\right)_t\Big|_0^1, \tag{7.107}$$

where

$$\tilde{K}_2 = \left(K_2 + \frac{\rho_l}{2P_w^{\beta/\gamma}}\right) \tag{7.108}$$

and

$$K_1 = (\beta+1)\frac{MC_2^{1/2}}{2P_w^{1/\gamma}}, \tag{7.109}$$

$$K_2 = (\beta+1)\frac{MC_2^{1/\beta}}{2P_w^{1/\gamma}}(M + C_2^{1/\beta} + 2) + \frac{M}{2}, \tag{7.110}$$

$$K_3 = (\beta+1)\frac{MC_2^{1/\beta}}{P_w^{1/\gamma}}, \tag{7.111}$$

$$K_4 = (\beta+1)\frac{MC_2^{1/\beta}}{2P_w^{1/\gamma}}. \tag{7.112}$$

By the assumptions on P_w and M, as expressed in (7.98), we have $\tilde{K}_2 < 3/4$, and we can move the expression containing $(u_{xxx})^2$ in (7.107) over to the left hand side:

$$\begin{aligned} \int_0^1 \left(\frac{(u_{xx})^2}{2}\right)_t dx + \int_0^1 \left(\frac{(g_{xx})^2}{2g}\right)_t dx + \frac{1}{4}\int_0^1 \frac{E}{g}(u_{xxx})^2 dx \\ \le K_1 \int_0^1 E(u_x)^2 dx + K_3 \int_0^1 \frac{E}{g}(u_{xx})^2 dx \\ + (K_4 + P_w^{1/\gamma}\rho_l)\int_0^1 \frac{(g_{xx})^2}{2g}dx - \frac{E_x}{E}\left(\frac{u_x^2}{2}\right)_t\Big|_0^1. \end{aligned} \tag{7.113}$$

We integrate over $[0,t]$:

$$\int_0^1 \left(\frac{(u_{xx})^2}{2}\right) dx + \int_0^1 \left(\frac{(g_{xx})^2}{2g}\right) dx + \frac{1}{4}\int_0^t \int_0^1 \frac{E}{g}(u_{xxx})^2 dx dt$$

$$
\begin{aligned}
&\leq \int_0^1 \left(\frac{u_{xx,0}^2}{2} \right) dx + \int_0^1 \left(\frac{(g_{xx,0})^2}{2g_0} \right) dx \\
&\quad + K_1 \int_0^t \int_0^1 E(u_x)^2 dx dt + K_3 \int_0^t \int_0^1 \frac{E}{g}(u_{xx})^2 dx dt \\
&\quad + (K_4 + P_w^{1/\gamma}\rho_l) \int_0^t \int_0^1 \frac{(g_{xx})^2}{2g} dx dt - \int_0^t \frac{E_x}{E} \left(\frac{u_x^2}{2} \right)_t \bigg|_0^1 dt.
\end{aligned}
\tag{7.114}
$$

We will now estimate the last term in (7.114):

$$
\begin{aligned}
&\int_0^t - \frac{E_x}{E} \left(\frac{u_x^2}{2} \right)_t \bigg|_0^1 dt \\
&\quad = -\frac{E_x}{E} \left(\frac{u_x^2}{2} \right) \bigg|_0^1 + \frac{E_{x,0}}{E_0} \left(\frac{u_{x,0}^2}{2} \right) \bigg|_0^1 + \int_0^t \left(\frac{E_x}{E} \right)_t \frac{u_x^2}{2} \bigg|_0^1 dt \\
&\quad \leq \frac{M^2}{P_w^{(\beta+1)/\gamma}} |E_x| + \int_0^t \left(\frac{E_x}{E} \right)_t \frac{u_x^2}{2} \bigg|_0^1 dt
\end{aligned}
\tag{7.115}
$$

Now, from (7.43) we have

$$
E_t = (\beta+1)Q^\beta Q_t = \frac{\beta+1}{\beta} Q(Q^\beta)_t = -(\beta+1)\rho_l \frac{Q^{\beta+2}}{g^2} u_x, \tag{7.116}
$$

$$
\begin{aligned}
E_{xt} &= -(\beta+1)\rho_l \left[\left(\frac{Q^{\beta+2}}{g^2} \right)_x u_x + \frac{Q^{\beta+2}}{g^2} u_{xx} \right] \\
&= -(\beta+1)\rho_l \left[Q \frac{(\beta+2)E_x}{(\beta+1)g^2} u_x - 2\frac{Q^{\beta+2}}{g^3} g_x u_x + \frac{Q^{\beta+2}}{g^2} u_{xx} \right],
\end{aligned}
\tag{7.117}
$$

and with the boundary conditions in Remark 7.11, we get

$$
\begin{aligned}
&\left(\frac{E_x}{E} \right)_t \bigg|_0^1 \\
&= \frac{E_{xt}}{E} - \frac{E_x E_t}{E^2} \bigg|_0^1 = \frac{E_{xt}}{E} + \frac{E_x}{E^2 g^2} (\beta+1)\rho_l Q^{\beta+2} u_x \bigg|_0^1 \\
&= (\beta+1)\rho_l \left[-Q \frac{(\beta+2)E_x}{(\beta+1)Eg^2} + 2\frac{Q^{\beta+2}}{Eg^3} g_x + \frac{Q^{\beta+2}}{E^2 g^2} E_x + \frac{E_x}{E^2 g^2} Q^{\beta+2} \right] u_x \bigg|_0^1
\end{aligned}
$$

$$
\begin{aligned}
&= (\beta+1)\rho_l \left[-\frac{(\beta+2)E_x}{(\beta+1)Q^\beta g^2} + 2\frac{Q}{g^3} g_x + \frac{1}{E^\beta g^2} E_x + \frac{E_x}{E^\beta g^2} \right] u_x \Big|_0^1 \\
&= (\beta+1)\rho_l \left[-\frac{(\beta+2)Q_x}{g^2} + 2\frac{Q}{g^3} g_x + (\beta+1)\frac{Q_x}{g^2} + (\beta+1)\frac{Q_x}{g^2} \right] u_x \Big|_0^1 \\
&= (\beta+1)\rho_l \left[2\frac{Q}{g^3} g_x + \beta\frac{Q_x}{g^2} \right] u_x \Big|_0^1, \qquad (7.118)
\end{aligned}
$$

such that

$$
\begin{aligned}
&-\int_0^t \frac{E_x}{E} \left(\frac{u_x^2}{2}\right)_t \Big|_0^1 dt \\
&\quad \le \frac{M^2}{P_w^{(\beta+1)/\gamma}} |E_x| + \int_0^t \left(\frac{E_x}{E}\right)_t \frac{u_x^2}{2} \Big|_0^1 dt \\
&\quad \le \frac{M^2}{P_w^{(\beta+1)/\gamma}} P_w^{1/\gamma} (\beta+1) Q^\beta \frac{-g_x}{(g-1)^2} + (\beta+1)\rho_l T M^4 \left[2C_2^{1/\beta} + \beta\frac{C_2^{2/\beta}}{P_w^{1/\gamma}} \right] \\
&\quad \le (\beta+1)\frac{M^3}{P_w^{(\beta+2)/\gamma}} C_2^{(\beta+2)/\beta} + (\beta+1)\rho_l T M^4 \left(2C_2^{1/\beta} + \beta\frac{C_2^{2/\beta}}{P_w^{1/\gamma}} \right) \\
&\quad = (\beta+1) M^3 C_2^{1/\beta} \left[\frac{C_2^{(\beta+1)/\beta}}{P_w^{(\beta+2)/\gamma}} + \rho_l T M \left(2 + \beta\frac{C_2^{1/\beta}}{P_w^{1/\gamma}} \right) \right]. \qquad (7.119)
\end{aligned}
$$

We can now estimate (7.114) in the following way:

$$
\begin{aligned}
&\int_0^1 \left(\frac{(u_{xx})^2}{2}\right) dx + \int_0^1 \left(\frac{(g_{xx})^2}{2g}\right) dx + \frac{1}{4}\int_0^t \int_0^1 \frac{E}{g}(u_{xxx})^2 dx dt \\
&\qquad \le \int_0^1 \left(\frac{(u_{xx,0})^2}{2}\right) dx + \int_0^1 \left(\frac{(g_{xx,0})^2}{2g}\right) dx \\
&\qquad\quad + K_1 C_1^0 + K_3 M_1 + (K_4 + P_w^{1/\gamma}\rho_l) \int_0^t \int_0^1 \frac{(g_{xx})^2}{2g} dx dt \\
&\qquad\quad + (\beta+1) M^3 C_2^{1/\beta} \left[\frac{C_2^{(\beta+1)/\beta}}{P_w^{(\beta+2)/\gamma}} + \rho_l T M \left(2 + \beta\frac{C_2^{1/\beta}}{P_w^{1/\gamma}} \right) \right],
\end{aligned}
$$

where we have used Proposition 7.3 and Lemma 7.12. By Grönwall's inequality we can now conclude that

$$
\begin{aligned}
&\int_0^1 \left(\frac{(u_{xx})^2}{2}\right) dx + \int_0^1 \left(\frac{(g_{xx})^2}{2g}\right) dx + \frac{1}{4}\int_0^t \int_0^1 \frac{E}{g}(u_{xxx})^2 dx dt \\
&\leq \left[1 + T(K_4 + P_w^{1/\gamma}\rho_l)\exp\left(T(K_4 + P_w^{1/\gamma}\rho_l)\right)\right] \\
&\quad \times \left[\int_0^1 \left(\frac{(u_{xx,0})^2}{2}\right) dx + \int_0^1 \left(\frac{(g_{xx,0})^2}{2g}\right) dx + K_1 C_1^0 + K_3 M_1 \right.\\
&\quad \left. +(\beta+1)M^3 C_2^{1/\beta}\left[\frac{C_2^{(\beta+1)/\beta}}{P_w^{(\beta+2)/\gamma}} + \rho_l T M\left(2 + \beta\frac{C_2^{1/\beta}}{P_w^{1/\gamma}}\right)\right]\right] \\
&:= M_2(C_1^0, M_1, \|g_{xx,0}\|_{L^2}, \|u_{xx,0}\|_{L^2}, C_2, T).
\end{aligned} \tag{7.120}
$$

□

By using Lemma 7.12 and Lemma 7.13, we show that it is possible to close the a priori assumptions with smallness assumptions on the initial data combined with a sufficiently small M.

Proposition 7.14 *Under the assumptions in Theorem* 7.2*, the a priori assumptions* (7.31a)-(7.31b) *and the assumptions* $1/2\rho_l P_w^{-\beta/\gamma} < 1$,

$$
\begin{aligned}
M < \min&\left\{\frac{3}{4}\frac{(2P_w^{\beta/\gamma} - \rho_l)P_w^{2/\gamma}}{(\beta+1)C_2^{(\beta+2)/\beta}}, \quad \frac{3}{4}\frac{2(1-\rho_l/(2P_w^{\beta/\gamma}))P_w^{1/\gamma}}{(\beta+1)C_2^{1/\beta}(2 + M + C_2^{1/\beta}) + P_w^{1/\gamma}},\right. \\
&\left. \frac{1}{32K_6}\cdot\frac{P_w^{(\beta+2)/\gamma}}{C_2^{1/\beta}(\beta+1)\left[C_2^{(\beta+1)/\beta} + 2\rho_l T M + \beta C_2^{1/\beta}P_w^{(\beta+1)/\gamma}\right]}\right\},
\end{aligned} \tag{7.121}
$$

where

$$
K_6 = \left[1 + T(K_4 + P_w^{1/\gamma}\rho_l)\exp\left(T(K_4 + P_w^{1/\gamma}\rho_l)\right)\right], \tag{7.122}
$$

$$
K_4 = (\beta+1)\frac{MC_2^{1/\beta}}{2P_w^{1/\gamma}}, \tag{7.123}
$$

and the assumption that there exists a constant $K_1(T, M)$ *such that*

$$
g_0(0) \geq K_1 > 1 + P_w^{1/\gamma}\rho_l T M, \tag{7.124}
$$

we can conclude that

$$|u_x| \leq 1/2M \quad \textit{and} \quad |g_x| \leq 1/2M \tag{7.125}$$

when the initial conditions $\|u_0\|_{L^2}$, $\|u_{x,0}\|_{L^2}$, $\|g_{x,0}\|_{L^2}$, $\|u_{xx,0}\|_{L^2}$ *and* $\|g_{xx,0}\|_{L^2}$ *are small enough on* $[0,1]$.

Proof. Remember that (7.124) has to hold for g^2Q^β to have an upper bound.

From the Sobolev inequality $\|f\|_{L^\infty} \leq \|f\|_{W^{1,1}}$ and from Lemma 7.12 and Lemma 7.13, we have

$$\begin{aligned} |u_x| &\leq \int_0^1 |u_x| dx + \int_0^1 |u_{xx}|\, dx \\ &\leq \int_0^1 (u_x)^2 dx^{1/2} + \int_0^1 (u_{xx})^2 dx^{1/2} && (7.126) \\ &\leq (2M_1)^{1/2} + (2M_2)^{1/2}, && (7.127) \end{aligned}$$

and

$$\begin{aligned} |g_x| &\leq \int_0^1 |g_x| dx + \int_0^1 |g_{xx}|\, dx \\ &\leq \int_0^1 (g_x)^2 dx^{1/2} + \int_0^1 (g_{xx})^2 dx^{1/2} \\ &\leq \frac{C_2^{1/2}}{P_w^{\beta/\gamma}} \left(\int_0^1 \frac{(g_x)^2}{g} dx^{1/2} + \int_0^1 \frac{(g_{xx})^2}{g} dx^{1/2} \right) && (7.128) \\ &\leq \frac{C_2^{1/2}}{P_w^{\beta/\gamma}} \left((2M_1)^{1/2} + (2M_2)^{1/2} \right). && (7.129) \end{aligned}$$

We do the calculations for $|u_x|$. We have

$$|u_x| \leq (2M_1)^{1/2} + (2M_2)^{1/2},$$

or, by Cauchy's inequality,

$$(u_x)^2 \leq \left((2M_1)^{1/2} + (2M_2)^{1/2}\right)^2 \leq 4(M_1 + M_2).$$

We want to show that we can make $(u_x)^2 \leq 1/4M^2$ such that $|u_x| \leq 1/2M$. Now, from the definition of M_1 in Lemma 7.12 and the definition of M_2 in Lemma 7.13,

$$\begin{aligned}
&4(M_1 + M_2)\\
&\quad = 4M_1 + 4\left[1 + T(K_4 + P_w^{1/\gamma}\rho_l)\exp\left(T(K_4 + P_w^{1/\gamma}\rho_l)\right)\right]\\
&\qquad \times \left[\int_0^1 \left(\frac{(u_{xx,0})^2}{2}\right) dx + \int_0^1 \left(\frac{(g_{xx,0})^2}{2g}\right) dx + K_1 C_1^0 + K_3 M_1\right.\\
&\qquad \left. + (\beta+1)M^3 C_2^{1/\beta}\left[\frac{C_2^{(\beta+1)/\beta}}{P_w^{(\beta+2)/\gamma}} + \rho_l T M\left(2 + \beta\frac{C_2^{1/\beta}}{P_w^{1/\gamma}}\right)\right]\right].
\end{aligned} \tag{7.130}$$

Clearly,

$$\begin{aligned}
M_1 &= \left(\int_0^1 \frac{(u_{0,x})^2}{2} dx + \int_0^1 \frac{(g_{0,x})^2}{2g_0} dx\right)(1 + TK_5\exp(TK_5)),\\
C_1^0 &= \frac{1}{2}\sup_{[0,1]} g_0 \|u_0\|_{L^2}^2,
\end{aligned}$$

from Lemma 7.12 and Proposition 7.3, can be controlled by the initial data $\|u_0\|_{L^2}$, $\|u_{x,0}\|_{L^2}$ and $\|g_{x,0}\|_{L^2}$. We can also control $\|u_{xx,0}\|_{L^2}$ and $\|g_{xx,0}\|_{L^2}$ in the expression (7.130).

K_1, K_3 are the same as in Lemma 7.13 and K_5 the same as in Lemma 7.12. We restate them here:

$$\begin{aligned}
K_1 &= (\beta+1)\frac{MC_2^{1/2}}{2P_w^{1/\gamma}},\\
K_3 &= (\beta+1)\frac{MC_2^{1/\beta}}{P_w^{1/\gamma}},\\
K_5 &= \left(M/2(\beta+1)C_2^{(\beta+2)\beta}P_w^{-1/\gamma} + 3/2P_w^{1/\gamma}\rho_l\right).
\end{aligned}$$

These constants are finite as long as M is finite, which is ensured by (7.121). Let K_6 be as in (7.122) such that we can write (7.130) in the shorter form:

$$\begin{aligned}4(M_1+M_2) =& 4M_1 + 4K_6 \int_0^1 \left(\frac{(u_{xx,0})^2}{2}\right) dx \\ &+ 4K_6 \int_0^1 \left(\frac{(g_{xx,0})^2}{2g}\right) dx + 4K_6K_1C_1^0 + 4K_6K_3M_1 \\ &+ 4K_6(\beta+1)M^3C_2^{1/\beta}\left[\frac{C_2^{(\beta+1)/\beta}}{P_w^{(\beta+2)/\gamma}} + \rho_l TM\left(2+\beta\frac{C_2^{1/\beta}}{P_w^{1/\gamma}}\right)\right].\end{aligned} \tag{7.131}$$

The five first terms in this expression can be controlled by initial data, as reflected by the above. The restriction on M in (7.121) ensures that the last expression is smaller than $1/8M^2$. Thus, by controlling the initial conditions for the three first terms we can also make them smaller than $1/8M^2$. Then we have

$$(u_x)^2 \le 4(M_1+M_2) \le \frac{M^2}{8} + \frac{M^2}{8} = \frac{M^2}{4}, \tag{7.132}$$

which we set out to achieve. The same reasoning can be used for $|g_x|$. Thus, we can make

$$\max\left\{\frac{C_2^{1/2}}{P_w^{\beta/\gamma}}\left((2M_1)^{1/2}+(2M_2)^{1/2}\right),\quad (2M_1)^{1/2}+(2M_2)^{1/2}\right\} \le \frac{1}{2}M,$$

such that $|u_x| \le 1/2M$ and $|g_x| \le 1/2M$, by controlling the initial data. □

7.4 A discussion concerning the estimates

In this section we will discuss the estimates that are obtained. We will justify the choice of a priori assumptions by identifying difficult terms when calculating the estimates and by the observations made on the relaxation model in Section 7.2. We will also look at the restrictiveness of the assumptions that were necessary when closing the a priori estimates.

First, we sum up all the results in the following corollary. The corollary is a direct consequence of the estimates that we have obtained.

Corollary 7.15 *Under the assumptions in Theorem 7.2, the assumptions that M satisfies* (7.121)*, that $P_w^{1/\gamma} > 1$ and that $1/2\rho_l P_w^{-\beta/\gamma} < 1$, we have*

$$\int_0^1 (u_t)^2 + (u_{xx})^2 + (m_{xx})^2 dx \leq N_1 \tag{7.133}$$

and

$$\int_0^t \int_0^1 (u_{xxx})^2 dxdt \leq N_2, \tag{7.134}$$

for some constants $N_1 = N_1(M, M_1, M_2, C_2)$, $N_2 = N_2(C_2, M_2)$. Further, for some constant $\mu = \mu(P_w^{1/\gamma}, C_2) > 0$, we have

$$0 < \mu \leq m \leq \rho_l - \mu < \rho_l, \tag{7.135}$$

$$0 < P_w^{1/\gamma}\mu \leq n \leq P_w^{1/\gamma}(\rho_l - \mu). \tag{7.136}$$

And g, m_t, m_x, u and u_x are pointwise bounded.

Proof. We know that u_x is bounded by assumption. As $g = \rho/m = (m + P_w^{1/\gamma}(\rho_l - m))/m$, we have $m_x = -m^2 g_x/(P_w^{1/\gamma}\rho_l)$. Then $|m_x| < M\rho_l/P_w^{1/\gamma}$, since the boundedness of Q in Proposition 7.7 implies $m < \rho_l$.

From 7.11 we have $m_t = -m^2 u_x$ such that $m_t < (\rho_l)^2 M$. Also, the boundedness of g follows from Proposition 7.7. Further,

$$u = \int_0^x u_x ds \leq M. \tag{7.137}$$

The estimate in (7.133) is a consequence of Proposition 7.7, Lemma 7.12 and Lemma 7.13. By observing that

$$u_t = \frac{1}{g}\left(E_x u_x + E u_{xx}\right), \tag{7.138}$$

we can easily see that u_t is bounded in $L^2((0,1))$. Also,

$$g_{xx} = \frac{P_w^{1/\gamma}\rho_l}{m^3}\left(2(m_x)^2 - m m_{xx}\right), \tag{7.139}$$

such that

$$m_{xx} = 2(m_x)^2 - \frac{1}{P_w^{1/\gamma}\rho_l} g_{xx} \tag{7.140}$$

and

$$\int_0^1 (m_{xx})^2 dx = \int_0^1 \left(2(m_x)^2 - \frac{1}{P_w^{1/\gamma}\rho_l} g_{xx} \right)^2 dx \tag{7.141}$$

$$\leq 8 \int_0^1 (m_x)^4 dx + 2 \int_0^1 \frac{1}{(P_w^{1/\gamma}\rho_l)^2} (g_{xx})^2 dx, \tag{7.142}$$

by Jensen's inequality. This expression is now bounded by the pointwise boundedness of m_x and Lemma 7.13.

The estimate in (7.134) follows from Proposition 7.7 and Lemma 7.13.

The inequalities in (7.135) follow directly from Proposition 7.7 and the fact that $n = P_w^{1/\gamma}(\rho_l - m)$. As $Q \leq C_2^{1/\beta}$ and

$$Q = \frac{\rho}{\rho_l - m} = \frac{m+n}{\rho_l - m}, \tag{7.143}$$

we must have $m \leq \rho_l - \mu$ for some $\mu > 0$. Also, we can see that $m > 0$ from the following. $1 \leq g = \rho/m \leq \sqrt{C_2/P_w^{1/\gamma}}$, such that m has to be bigger than 0. As n only depends on m, the inequality for n follows from the one for m. □

In other words, under the necessary assumptions, there will exist both gas and liquid at any point x in the model for any finite time t, as m and n are pointwise bounded, which we set out to obtain in the beginning. Further, the change of m, n and u in any direction x is bounded for any finite time t.

Remark 7.16 *As m, m_t, m_x, g, u and u_x are pointwise bounded on $[0, 1]$, they will also be bounded in $L^2((0, 1))$.*

7.4.1 Existence

By the same kind of compactness argument as in [18] for the full model, we can conclude from Corollary 7.15 that Theorem 7.2 holds. Due to the length of the complete argument, we include only the main steps. As the approach for n is the same as for m, we only show existence for m and u. Let $I = (0,1)$.

We denote $j_\delta(x)$ as the Friedrichs mollifier. We extend m_0 and u_0 to $[-1,2]$, i.e.

$$m_0 := \begin{cases} m_0(1), & x \in (1,2], \\ m_0(x), & x \in [0,1], \\ m_0(0), & x \in (0,-1], \end{cases} \qquad u_0 := \begin{cases} -u_0(-x), & x \in [-1,0), \\ u_0(x), & x \in [0,1], \\ -u_0(2-x), & x \in (1,2]. \end{cases} \tag{7.144}$$

The approximate initial conditions are then

$$m_0^\delta(x) = (m_0 * j_\delta)(x), \tag{7.145a}$$

$$u_0^\delta(x) = (u_0 * j_\delta)(x) + (x-1)(u_0 * j_\delta)(0) - x(u_0 * j_\delta)(1). \tag{7.145b}$$

Then $m_0^\delta \in C^{1+s}([0,1])$ and $u_0^\delta \in C^{2+s}((0,1))$ for $0 < s < 1$ such that we have enough regularity. Also, m_0^δ and u_0^δ are compatible with the boundary conditions. As $\delta \to 0$,

$$m_0^\delta \to m_0 \quad \text{uniformly in} \quad [0,1], \quad u_0^\delta \to u_0 \quad \text{in} \quad W^{2,2}(I), \tag{7.146}$$

since m_0 is bounded.

We consider the initial boundary value problem for the reduced model with (m_0^δ, u_0^δ). For this problem, standard arguments, the energy estimates and the contraction mapping theorem, can be used to obtain existence of a unique *local* solution m^δ, u^δ with m^δ, m_t^δ, m_x^δ, m_{tx}^δ, m_{txx}^δ, m_{xx}^δ, u^δ, u_x^δ, u_t^δ, u_{xt}^δ, u_{xx}^δ, u_{txx}^δ, u_{xxx}^δ $\in C^{\alpha,\alpha/2}([0,1] \times [0,T^*])$ for some $T^* > 0$. We apply the energy method to derive bounds for high-order derivatives of (m^δ, u^δ). Schauder theory for linear parabolic equations can then be applied to conclude that the $C^{\alpha,\alpha/2}([0,1] \times [0,T^*])$-norm of m^δ, m_t^δ, m_x^δ, m_{tx}^δ, m_{txx}^δ, m_{xx}^δ, u^δ, u_x^δ, u_t^δ, u_{xt}^δ, u_{xx}^δ, u_{txx}^δ, u_{xxx}^δ is a priori bounded. See [1, Ch. 3] for a similar argument. The local solution can then be continued globally in time. Now we can conclude that there exists a unique global solution m^δ, u^δ for the reduced model (7.25b)-(7.25c) with initial data m_0^δ, u_0^δ such that the regularity of (7.29) holds for any $T > 0$.

The estimates in Corollary 7.15 hold for (m^δ, u^δ) and are independent of δ. Hence, we can extract subsequences such that as $\delta \to 0$,

$$\begin{aligned} (m^\delta, u^\delta) &\rightharpoonup (m,u) && \text{weak-* in} \quad L^\infty([0,T], W^{2,2}(I)), \\ (m_t^\delta, u_t^\delta) &\rightharpoonup (m_t, u_t) && \text{weak-* in} \quad L^\infty([0,T], L^2(I)), \end{aligned} \tag{7.147}$$

$$u^\delta_{xxx} \rightharpoonup u_{xxx} \qquad\qquad \text{weakly in} \quad L^2([0,T],L^2(I)).$$

As g and g_t are pointwise bounded, one can easily see from (7.147) that g_t and $(gu)_t$ also satisfies the conditions in Theorem 7.2.

We will now show that (m,u) in (7.147) is in fact a solution of the reduced model (7.25b)-(7.25c).

As $u \in W^{2,2}(I)$, we have $u, u_x \in W^{1,2}(I)$. Also, $m \in W^{1,2}(I)$. The Sobolev embedding (Morrey's inequality, see for example [17, Sec. 5.6.2]) $W^{1,2q}(0,1) \hookrightarrow C^{1-1/2q}[0,1]$ with $q=1$ gives that for any $x_1, x_2 \in (0,1)$ and $t \in [0,T]$,

$$|m^\delta(x_1,t) - m^\delta(x_2,t)| \leq C|x_1 - x_2|^{1/2}. \tag{7.148}$$

The same does also hold for u and u_x.

We need to control continuity in time. With the embeddings $W^{1,2}(0,1) \hookrightarrow L^\infty(0,1) \hookrightarrow L^2(0,1)$, it follows from the Lions-Aubin lemma [42, Sec. 1.3.12] that for any $\nu > 0$ there exists a constant C_ν such that, for any $t_1, t_2 \in [0,T]$,

$$\begin{aligned}
&\|m^\delta(t_1) - m^\delta(t_2)\|_{L^\infty(I)} \\
&\qquad \leq \nu\|m^\delta(t_1) - m^\delta(t_2)\|_{W^{1,2}(I)} + C_\nu\|m^\delta(t_1) - m^\delta(t_2)\|_{L^2(I)} \\
&\qquad \leq 2\nu\|m^\delta(t)\|_{W^{1,2}(I)} + C_\nu|t_1 - t_2|^{1/2}\|m^\delta_t\|_{L^2([0,T],L^2(1))} \\
&\qquad \leq C\nu + C_\nu C|t_1 - t_2|^{1/2}.
\end{aligned} \tag{7.149}$$

In the above we have applied the results from Corollary 7.15 to derive the last two inequalities. By applying the triangle inequality, the above shows that $\{m^\delta\}$ is equicontinuous on $[0,1]\times[0,T]$. Hence, by the Arzelà-Ascoli theorem [16, Ch. 19] and a diagonal process for t, we can extract a subsequence such that

$$m^\delta(x,t) \to m(x,t) \quad \text{strongly} \quad \text{in } C^0([0,T]\times[0,1]). \tag{7.150}$$

In the same way, we have

$$\begin{aligned}
u^\delta(x,t) &\to u(x,t) \quad \text{strongly} \quad \text{in } C^0([0,T]\times[0,1]), \\
u^\delta_x(x,t) &\to u_x(x,t) \quad \text{strongly} \quad \text{in } C^0([0,T]\times[0,1]).
\end{aligned} \tag{7.151}$$

As $m_t \in L^\infty([0,T],L^2(I))$, $m_t \in L^2([0,T],L^2(I))$. Then from

$$\|m(t_1) - m(t_2)\|^2_{L^2(I)} = \int_0^1 |m(t_1) - m(t_2)|^2 dx$$

$$
\begin{aligned}
&= \int_0^1 \left| \int_{t_1}^{t_2} m_t ds \right|^2 dx \leq \int_0^1 \left(\int_{t_1}^{t_2} |m_t| ds \right)^2 dx \qquad (7.152) \\
&\leq |t_1 - t_2| \int_0^T \int_0^1 (m_t)^2 dx ds,
\end{aligned}
$$

where Hölder's inequality is used, we see that $m \in C^{1/2}([0,T], L^2(I))$ as well. The same can also be applied to gu (or to u).

We also know that $g, E \in C^0([0,T] \times [0,1])$ since $m \in C^0([0,T] \times [0,1])$ and $0 < m < \rho_l$. Hence, we can now conclude that the limits (m, u) satisfy the equations in (7.32) for a.e. $x \in (0,1)$ and for $t \geq 0$ in Theorem 7.2.

7.4.2 The a priori assumption

The choice of the a priori assumption $|u_x| \leq M$ is motivated by the analysis of the corresponding relaxation system in Section 7.2. As this system is unstable when the velocity u gets large enough, a bound on u_x may prevent u from getting that large. Also, the diffusive term should prevent $|u_x| \to \infty$ as long as g is bounded.

The a priori assumption $|g_x| \leq M$ seems to be necessary when estimating u_{xx} in L^2. Making this assumption did not introduce new difficulties as we were able to close it in the same way as we closed $|u_x| \leq M$ in Proposition 7.14.

Several of the estimates throughout this chapter could be simplified with the two a priori assumptions $|g_x| \leq M$ and $|u_x| \leq M$. For example, Q_x^β in Proposition 7.8 would be directly bounded by the pointwise bound of Q and g_x. Nevertheless, we have chosen to leave the calculations of the estimates as they are such that the a priori assumptions are used as little as possible. By doing this, the calculations of the estimates can more easily be altered to fit different a priori assumptions. Thus, this may make it easier to obtain different or better estimates for the model in the future.

7.4.3 Difficult terms

In this section we identify the most difficult terms in the model that made it necessary to impose further restrictions. By studying these terms one may find different or better ways to obtain estimates than the ones we have used.

The a priori assumption $|u_x| \leq M$ was first needed when finding an upper bound for Q in Proposition 7.7, where the term

$$\int_0^t g^2 Q^\beta u_x dt \tag{7.153}$$

proved to be difficult to bound as it only is an integral w.r.t t. By imposing $|u_x| \leq M$, we could bound this term by using Grönwall's ineguality. Further, we found no way of bounding

$$\int_0^t Q^{\beta+1}(0,t) u_x(0,t) dt \tag{7.154}$$

without the assumption on $g_0(0)$ in Lemma 7.6.

We also needed an additional a priori assumption in Proposition 7.8 when finding an estimate for Q_x^β in $L^2([0,1])$. In this estimate, the terms

$$\int_0^1 \int_0^t (u_t)^2 dx dt, \quad \int_0^1 \int_0^t \frac{(g_x)^2}{g^2}((Q^\beta)_t)^2 dx dt \tag{7.155}$$

were a challenge. We were able to bound both terms with the same a priori assumption $|u_x| \leq M$, which was a further motivation for using that assumption.

Even though $|u_x| \leq M$ is an a priori assumption that works for the model, it could also be useful to try other a priori assumptions to bound the difficult terms. It may, for instance, be possible to find a priori assumptions that imposes fewer restrictions on the initial data.

7.4.4 Restrictions

g_x and u_x are bounded by the constant M. Looking at the assumption 7.121 in Proposition 7.14, this constant can be small. In those cases, the change in the x-direction of the velocity u and of the ratio $g = \rho/m$ is small, meaning that the concentration of liquid mass m and gas mass $n = \rho - m$ does not change rapidly from point to point for any time t.

As $\gamma > 1$ and $0 < \beta < 1/3$, the criterion

$$1/2 P w^{-\beta/\gamma} \rho_l < 1, \tag{7.156}$$

which is necessary to close the a priori assumptions in Proposition 7.14, is a relatively strict physical criterion. Normally $\rho_l \sim 10^3$ and $P_w \sim 10^5$. This would give $P_w^{\beta/\gamma} < 10^2$, which would result in $1 < 1/2Pw^{-\beta/\gamma}\rho_l$. Thus, the ratio between the liquid density ρ_l and the pressure $P_w^{\beta/\gamma}$ has to be controlled carefully.

From the above, it is likely that $1/2Pw^{-\beta/\gamma}\rho_l$ can be close to 1, which could make M a fairly large constant.

Remark 7.17 *Notice that we do not need* $\beta \in (0, 1/3)$ *for the estimates of the reduced model to hold. The estimates would still be the same with* $\beta \in (0, 1)$*, which would make* (7.156) *less restrictive.*

Also notice in Proposition 7.7 that if we let $T \to \infty$, the pointwise bound on Q will blow up. So, we cannot let the time tend to infinity.

7.5 Summary

We shortly sum up the results and discussion from this chapter. We studied a well-reservoir model for two-phase flow, a relaxation model with a diffusive term. We restated an existence result for this model, which ensured the existence of both gas and liquid at any point for any finite time. We also studied the corresponding relaxation model without any diffusive term. We found that this model only satisfies the subcharacteristic condition when the velocity u is small. From the results in Chapter 4, we then know that the relaxation model is unstable when u gets to large. The diffusive term does therefore seem to be necessary for the stability of the model.

In particular, we have obtained estimates for the reduced model. This model is the formal limit of the full system, with the diffusive term, when the relaxation parameter $q_w \to \infty$. Theorem 7.2 shows that we have obtained estimates that, in a similar way to the estimates of the full model, ensure the existence of both gas and liquid at any spatial point in the model for any finite time t. Further, an existence result for solutions in suitable Sobolev spaces, similar to the result of the full model, is established in Theorem 7.2.

Some of the terms proved to be difficult to estimate. But, by a priori assuming that $|g_x|, |u_x| \leq M$ for a small enough $M > 0$, we obtained estimates that ensure existence of solutions when the initial data is small enough. These a priori assumptions also show that the changes of u and g in the x-direction are bounded.

To complete all the estimates, a rather strict restriction on the ratio between the liquid density ρ_l and the pressure P_w was necessary. The parameters ρ_l, P_w, β and γ must therefore be chosen carefully.

8 Conclusions

In this chapter we shortly summarize and discuss the main results in the thesis. We also suggest some topics for further work.

In this thesis we have mainly studied three subjects - constant-coefficient relaxation systems, a two-phase well-reservoir model and entropy conditions - all connected with relaxation processes. The main focus has been on the two first topics.

8.1 Linear relaxation systems of rank one

For constant-coefficient rank one relaxation systems we have obtained three general results. Assuming that the relaxation system is strictly hyperbolic and that the relaxation matrix is stable, we have proved the following:

i) The eigenvalue polynomial of the relaxation system can be written as a convex sum of the two limiting eigenvalue polynomials.

ii) There is an equivalence between stability and the subcharacteristic condition.

iii) Stability implies that the velocities of the relaxation system will never exceed the velocities of the corresponding homogeneous system.

Stability, in the sense above, implies that the solutions, with initial conditions in $L^2(\mathbb{R})$, of the relaxation system converge in $L^2(\mathbb{R})$ as the relaxation parameter $\varepsilon \to 0$. In other words, with the results obtained, we have shown that the solution of a constant-coefficient strictly hyperbolic relaxation system with a stable rank one relaxation matrix will converge in L^2 if and only if the system satisfies the subcharacteristic condition.

For some rank one relaxation systems, as the two-phase flow model in Chapter 5, a critical region for values of the relaxation time ε has been identified. The velocities of the Fourier components in this region are equal to zero. A short study of an example system with this behavior was performed. We found, as expected, that the critical region has a visible impact on the solutions. It is still uncertain how this region will affect the corresponding nonlinear model.

8.2 Entropy conditions

A literature study was performed to get a better understanding of the concept of entropy conditions. It was shown that an entropy condition is not necessary, but sufficient, for the stability and stiffly well-posedness of constant-coefficient relaxation systems with nonoscillatory relaxation matrices.

Conservative and nonconservative hyperbolic systems were compared. The comparison shows that the difference between the two systems starts *at the beginning* when defining weak solutions. The definition for nonconservative systems is far more comprehensive due to the path-dependence of the nonconservative product in the system.

It was shown that a conservative system can be written in a symmetric form if and only if the system is endowed with a strictly convex entropy. This equivalence does not hold in general for nonconservative systems. But, if the nonconservative system can be written in a symmetric form, it can be written in a conservative form. The symmetric form will then imply the existence of a convex entropy. The existence of a convex entropy will not generally imply that a nonconservative system can be written in a symmetric form. An additional symmetry assumption is needed.

The subcharacteristic condition was revisited. It was shown for conservative relaxation systems that the existence of a convex entropy implies the subcharacteristic condition. We also explained that this may not necessary hold for nonconservative relaxation systems due to lack of symmetries. We also showed that the subcharacteristic condition alone does not imply the existence of a convex entropy for the corresponding relaxation system.

8.3 The well-reservoir model

Theorem 7.2 summarizes the most important results from Chapter 7, which concern the existence of solutions for the reduced model. The reduced model is the formal limit of the full well-reservoir model, which is a rank one relaxation system with a viscosity term, as the relaxation parameter $1/q_w = \varepsilon \to 0$. The existing estimates for the full model are highly dependent on the relaxation parameter q_w and will tend to ∞ as q_w does. It was therefore necessary to find new estimates for the reduced model.

We showed that for a sufficiently small $M > 0$, such that $|u_x|, |g_x| \leq M$, there exists a solution to the reduced model which, in similarity to the solution of the full model, ensures that there will exist both liquid and gas at any point x in the model for any finite time t. The precise statement can be found in Theorem 7.2.

Setting the viscosity term of the full model to zero, we showed that the resulting relaxation system is not even linearly stable when the velocity u gets large enough. The viscosity term may therefore be necessary to obtain enough regularity for the model.

The results that we have obtained so far concerns only the reduced model with viscosity. Thus, we cannot say much about the relaxation process from the full model to the reduced model. But, the instabilities occurring when u increases for the corresponding relaxation system implies that we may have to prevent u from getting to large to obtain relaxation parameter independent estimates. The a priori assumption $|u_x| \leq M$ was chosen, amongst other reasons, to try to hinder u in the reduced model from increasing to much. It is therefore plausible that using the same a priori assumptions on the full model, with the same fixed, instead of free, boundary conditions, would make it possible to find relaxation parameter independent estimates. If this is possible, an existence result can be derived for the reduced model from the parameter independent existence result of the full model.

8.4 Suggestions for further work

In this section we suggest further work for the most important parts of the thesis. As the most original contributions in this thesis concerns the constant-coefficient relaxation systems in Chapter 4 and the well-reservoir model in Chapter 7, we will propose topics for further work based on these two chapters.

8.4.1 Linear relaxation systems

As the results in Chapter 4 are limited to rank one relaxation systems, it would be interesting to find out to what extent the results for these systems can be generalized to systems with relaxation matrices of higher ranks. For higher ranks, it will no longer be possible to write the eigenvalue polynomial as the sum of the two limiting eigenvalue polynomials. One may therefore have to rely on different techniques to find results. Stronger stability criteria may also be needed. There is a theorem in [54] supporting this assumption. The theorem shows that for higher ranks than one, a slightly stronger stability assumption will imply the subcharacteristic condition. It would be interesting to see if it is possible to prove the other direction as well.

Working with constant-coefficient relaxation systems is restrictive. An extension of the results for rank one systems to the variable coefficients case, i.e. $\boldsymbol{A} = \boldsymbol{A}(x,t)$ and $\boldsymbol{R} = \boldsymbol{R}(x,t)$ in (4.1), could be a useful task. However, how one may approach this task is uncertain. We cannot straightforwardly write the solution as the sum of its Fourier components in the same way as for the constant-coefficient case in Chapter 3. But, finding and studying the Fourier transform could be a first approach.

It is also plausible that these results can be extended to nonlinear systems when all the amplifications in the system are strictly smaller than zero, but a closer study is needed to conclude anything certain.

The critical region, appearing for some rank one relaxation systems, is an interesting phenomenon. A study of how this region affects the corresponding nonlinear model would be interesting.

8.4.2 The well-reservoir model

There is often room for improvements and, as first task, it would be interesting to see if other a priori assumptions give better estimates for the reduced well-reservoir model. Specifically, estimates with fever restrictions on the pressure and the liquid density would be preferable.

Even though we have estimates and existence results for both the full model and the reduced model, we know little about the relaxation process from the full system to the reduced system. We only know that the system is unstable for large u. It would be interesting to study the relaxation process to determine if there exists a regime, with sufficiently small u, where solutions of the full system in fact relaxes towards solutions of the reduced system. For example, could similar a priori assumptions, initial conditions and boundary conditions for the full model give relaxation parameter independent estimates that ensure convergence of solutions as the parameter q_w tends to ∞?

Evje and Karlsen [19] show uniqueness for a model that is similar to the reduced model studied in this thesis. It would therefore be interesting to try to develop uniqueness results for the reduced model under the assumptions that we have made.

Bibliography

[1] S. N. Antontsev, A. V. Kazhikhov, and V. N. Monakhov. *Boundary Value Problems in Mechanics of Nonhomogeneous Fluids.* Elsevier, 1990.

[2] P. Aursand and T. Flåtten. On the dispersive wave-dynamics of 2×2 relaxation systems. *J. Hyp. Diff. Eq.*, 9(4):641–659, 2012.

[3] A. Aw and M. Rascle. Resurrection of second order models for traffic flow. *J. Appl. Math.*, 60:916–938, 2000.

[4] C. Berthon and F. Coquel. Multiple solutions for compressible turbulent flow models. *Commun. Math. Sci.*, 4:497–511, 2006.

[5] S. Bianchini and A. Bressan. Vanishing viscosity solutions of nonlinear hyperbolic systems. *Annals of Math.*, 161:223–342, 2005.

[6] N. K. Bose and Y. Q. Shi. Network realizability theory approach to stability of complex polynomials. *IEEE Transactions on Circuits and Systems*, 34(2):216–218, 1987.

[7] F. Bouchut. A reduced stability condition for nonlinear relaxation to conservation laws. *J. Hyp. Diff. Eq*, 1:149–170, 2004.

[8] F. Bouchut. Stability of relaxation models for conservation laws. In *ECM, Stockholm, Sweden, June 27–July 2, 2004*, pages 95–101. European Mathematical Society, 2005.

[9] A. Bressan. Lecture notes on hyperbolic conservation laws, May 2009. http://php.math.unifi.it/users/cime/Courses/2009/01/200912-Notes.pdf.

[10] A. Bressan and P. G. LeFloch. Uniqueness of entropy solutions for systems of conservation laws. *Arch. Rational Mech. Anal.*, 140:301–331, 1997.

[11] G.-Q. Chen. Relaxation limits for conservation laws. *Z. Angew. Math. Mech.*, 47:381–384, 1996.

[12] G.-Q. Chen. Euler equations and related hyperbolic conservation laws. In *The Handbook of Differential Equations*, volume 2, pages 1–104. Elsevier Science, 2005.

[13] G.-Q. Chen and T.-P. Liu. Zero relaxation and dissipation limits for hyperbolic conservation laws. *Comm. Pure Appl. Math.*, 46:755–781, 1993.

[14] C. C. Christoforou. Hyperbolic systems of balance laws via vanishing viscosity. *J. Diff. Eq.*, 221:470–541, 2006.

[15] C. M. Dafermos. *Hyperbolic Conservation Laws in Continuum Physics*. Springer, 2010.

[16] E. DiBenedetto. *Real Analysis*. Birkhaüser Advanced Texts, 2002.

[17] L. C. Evans. *Partial Differential Equations*. Am. Math. Soc., 2010.

[18] S. Evje. A compressible two-phase model with pressure-dependent well-reservoir interaction. *SIAM J. Math. Anal.*, 45:518–546, 2013.

[19] S. Evje and K. H. Karlsen. Global weak solutions for a viscous liquid-gas model with singular pressure law. *Comm. Pure Appl. Anal.*, 8:1867–1894, 2009.

[20] S. Fisk. *Polynomials, Roots and Interlacing*. arXiv.org, 2008.

[21] K. O. Friedrichs and P. D. Lax. Systems of conservation equations with a convex extension. *Proc. Nat. Acad. Sci. USA*, 68:1686–1688, 1971.

[22] C. D. Levermore G.-Q. Chen and T.-P. Liu. Hyperbolic conservation laws with stiff relaxation terms and entropy. *Comm. Pure Appl. Math.*, 47:787–830, 1994.

[23] D. J. H. Garling. *Inequalities: A Journey Into Linear Analysis*. Cambridge, 2007.

[24] P. Germain and E. H. Lee. On shock waves in elastic-plastic solids. *J. Mech. Phys. Solids*, 21:359–382, 1973.

[25] S. K. Godunov. An interesting class of quasilinear systems. *Dokl. Acad. Nauk SSSR*, 139:521–523, 1961.

[26] L. Gosse. A nonconservative numerical approach for hyperbolic systems with source terms: the well-balanced schemes. In *Hyperbolic Problems: Theory, Numerics, Applications*, volume 140 of *Int. Series of Num. Math.*, pages 453–461. Birkhäuser Basel, 2001.

[27] S. Jiang, Z. Xin, and P.Zhang. Global weak solutions to 1d compressible isentropic navier-stokes equations with density-dependent viscosity. *Meth. Appl. Anal.*, 12:239–252, 2005.

[28] S. Jin and Z. Xin. The relaxation schemes of systems of conservation laws in arbitrary space dimensions. *Comm. Pure Appl. Math.*, 48:235–276, 1995.

[29] H.-O. Kreiss and J. Lorenz. *Initial-Boundary Value Problems and the Navier-Stokes Equations.* Academic Press, 1989.

[30] P. G. LeFloch. Entropy weak solutions to nonlinear hyperbolic systems in nonconservative form. *Comm. Part. Diff. Equa*, 13:669–727, 1988.

[31] P. G. LeFloch. Shock waves for nonlinear hyperbolic systems in nonconservative form. *IMA Preprint Series*, 593, 1989.

[32] P. G. LeFloch. *Hyperbolic Systems of Conservation Laws.* Birkhäuser, 2002.

[33] T.-P. Liu. Hyperbolic conservation laws with relaxation. *Comm. Math. Phys.*, 108:153–175, 1987.

[34] J. Lorenz and H. J. Schroll. Stiff well-posedness for hyperbolic systems with large relaxation terms. *Adv. in Diff. Eqns.*, 2:643–666, 1997.

[35] J. Lorenz and H.J. Schroll. Hyperbolic systems with relaxation: symmetrizers and entropies. In *Hyperbolic Problems: Theory, Numerics, Applications*, volume 130 of *Int. Series of Num. Math.*, pages 823–832. Birkhäuser Basel, 1999.

[36] H. Lund. A hierarchy of relaxation models for two-phase flows. *SIAM J. Appl. Math.*, 72:1713–1741, 2012.

[37] G. Dal Maso, P. G. LeFloch, and F. Murat. Definition and weak stability of nonconservative products. *J. Math. Pures Appl.*, 74:483–548, 1995.

[38] J. N. McDonald and N. A. Weiss. *A Course in Real Analysis.* Academic Press, 2012.

[39] A. Miroshnikov and C. Trivisa. Relative entropy in hyperbolic relaxation of balance laws. *Comm. Math. Sci.*, 12:1017–1043, 2014.

[40] M. Moskowitz and F. Paliogiannis. *Functions of Several Real Variables*. World Scientific, 2011.

[41] J. Munkres. *Topology*. Pearson Prentice Hall, 2nd edition, 1999.

[42] A. Novotny and I. Straskraba. *Introduction to the Mathematical Theory of Compressible Flows*. Oxford University Press, 2004.

[43] M. Reed and B. Simon. *Mehtods of Modern Mathematical Physics IV: Analysis of Operators*. Academic press, 1978.

[44] F. M. Reza. A note on complex reactance functions. *Computers Elect. Engang*, 18:183–188, 1992.

[45] F. M. Reza. Generation of complex reactance functions. *Journal of the Franklin Institute*, 330:423–430, 1993.

[46] S. Solem, T. Flåtten, and P. Aursand. The dispersive wave dynamics of a two-phase flow relaxation model. *Pre-print*, 2013. Available at http://www.pvv.org/~pederka/phaserelax.pdf.

[47] L. E. Spence, A. J. Insel, and S. H. Friedberg. *Elementary Linear Algebra*. Pearson Education, 1997.

[48] H. B. Stewart and B. Wendroff. Two-phase flow: models and methods. *J. Comp. Phys.*, 56:363–409, 1984.

[49] T. Flåtten and H. Lund. Relaxation two-phase flow models and the sub-characteristic condition. *Math. Models Methods Appl. Sci.*, 21:2379–2407, 2011.

[50] A. Tzavaras. Relative entropy in hyperbolic relaxation. *Comm. Math. Sci.*, 3:101–259, 2005.

[51] W. Wasow. *Asymptotic Expansions for Ordinary Differential Equations*. Wiley, New York, 1965.

[52] G. B. Whitham. *Linear and Nonlinear Waves*. Whiley, New York, 1974.

[53] W.-A. Yong. Singular pertubations of first-order hyperbolic systems with stiff source terms. *J. Diff. Eq.*, 155:89–132, 1992.

[54] W.-A. Yong. Basic aspects of hyperbolic relaxation systems. In *Advances in the Theory of Shock Waves*, volume 47 of *Progr. Nonlinear Differential Equations Appl.*, pages 259–305. Birkhäuser Boston, 2001.

[55] W.-A. Yong. Basic structures of hyperbolic relaxation systems. *P. Roy. Soc. Edinb. A*, 132:1259–1274, 2002.

[56] Z. Zahreddine. An extension of the routh array for the asymptotic stability of a system of differential equations with complex coefficients. *Appl. Anal.*, 49:61–72, 1993.

[57] Z. Zahreddine. On the interlacing property of stable complex systems of differential equations. *Indian J. pure appl. Math*, 26:643–647, 1995.

[58] M. Zhang and C. Zhu. Global existence of solutions to a hyperbolic-parabolic system. *Proc. Am. Math. Soc.*, 135:1017–1027, 2006.

[59] R. K. P. Zia, E. F. Redish, and S. R. McKay. Making sense of the legendre transform. *American J. Phys.*, 77, 2009.

A Article: Wave dynamics of linear hyperbolic relaxation systems

Journal of Hyperbolic Differential Equations

Wave dynamics of linear hyperbolic relaxation systems

SUSANNE SOLEM

Dept. of Mathematical Sciences
Norwegian University of Science and Technology (NTNU)
NO-7491 Trondheim
Norway
susanne.solem@ntnu.no

PEDER AURSAND

Dept. of Mathematical Sciences
Norwegian University of Science and Technology (NTNU)
NO-7491 Trondheim
Norway
peder.aursand@math.ntnu.no

TORE FLÅTTEN*

SINTEF Materials and Chemistry
P. O. Box 4760 Sluppen
NO-7465 Trondheim
Norway
Tore.Flatten@sintef.no

Received (Day Mth. Year)
Revised (Day Mth. Year)

Communicated by [editor]

Abstract. We consider linear hyperbolic systems with a stable rank 1 relaxation term. We establish that the characteristic polynomial for the individual Fourier components of the solution can be written as a convex combination of the eigenvalue polynomials for the formal stiff and non-stiff limits. This allows us to provide a direct and elementary proof of the equivalence between linear stability and the subcharacteristic condition. In a similar vein, a maximum principle follows: the velocity of each individual Fourier component is bounded by the minimum and maximum eigenvalues of the non-stiff limit system.

Keywords: Hyperbolic relaxation systems; sub-characteristic condition; linear stability.

*Corresponding author

1. Introduction

We are interested in hyperbolic conservation laws with relaxation source terms, acting to drive the system towards an equilibrium state. Such systems have many applications in the modeling of natural phenomena [2,7], in particular they are useful for describing the interaction between fluid-mechanical and thermodynamical processes [9,16,17].

In one space dimension, hyperbolic relaxation systems can be written in the general form

$$\partial_t \boldsymbol{U} + \partial_x \boldsymbol{F}(\boldsymbol{U}) = \frac{1}{\varepsilon}\boldsymbol{Q}(\boldsymbol{U}), \tag{1.1}$$

to be solved for the vector $\boldsymbol{U}(x,t) \in G \subseteq \mathbb{R}^N$. Herein, $\boldsymbol{F}(\boldsymbol{U})$ is the flux vector and $\boldsymbol{Q}(\boldsymbol{U})$ is the relaxation term. The parameter $\varepsilon > 0$ determines the rate of relaxation towards equilibrium. The system is hyperbolic when $\boldsymbol{F}_{\boldsymbol{U}}(\boldsymbol{U})$ has real eigenvalues and is diagonalizable and strictly hyperbolic when all the eigenvalues are real and distinct.

Two limit cases of (1.1) are of particular interest and will be central to the investigations of this paper:

- The *non-stiff limit*, characterized by $\varepsilon \to \infty$. In this limit, we may write (1.1) as

$$\partial_t \boldsymbol{U} + \partial_x \boldsymbol{F}(\boldsymbol{U}) = 0. \tag{1.2}$$

  We will denote (1.2) as the *homogeneous system*.
- The formal *equilibrium limit*, characterized by $\boldsymbol{Q}(\boldsymbol{U}) \equiv 0$. This assumption defines an equilibrium manifold [5] through

$$M = \{\boldsymbol{U} \in G : \boldsymbol{Q}(\boldsymbol{U}) = 0\}. \tag{1.3}$$

  Imposing local equilibrium, we may express (1.1) as

$$\partial_t \boldsymbol{u} + \partial_x \boldsymbol{f}(\boldsymbol{u}) = 0, \tag{1.4}$$

  for some reduced variable $\boldsymbol{u}(x,t) \in \mathbb{R}^n$, where $n \leq N$. Herein, every $\boldsymbol{u}$ uniquely defines an equilibrium state $\mathcal{E}(\boldsymbol{u}) \in M$.

  We will denote (1.4) as the *equilibrium system*.

A highly relevant question is whether solutions to (1.1) converge to solutions to (1.4) as $\varepsilon \to 0$. Chen [4] gives an overview of the existing literature where some important results are included. For the solution of a relaxation system to have a well-behaved limit, stability of the solution is a necessary criterion. Chen et al. [5] introduce an entropy condition which ensures dissipativity of the first-order Chapman-Enskog type expansion. If the full system is endowed with such an entropy, there will also exist a strictly convex entropy for the equilibrium system, implying that the equilibrium system will be hyperbolic. Further, under some suitable assumptions, it has been proved [5,6] that the solution of the 2×2 relaxation system strongly converges to that of the equilibrium system. Lattanzio and Marcati [11] prove convergence for the same system by using a compensated compactness argument. Such arguments were first developed in this context by Chen and Liu [6]. Yong [22,23] establishes a *relaxation criterion* which is necessary for the convergence of solutions as $\varepsilon \to 0$. For strictly non-linear systems, stronger stability assumptions, including the existence of a strictly convex entropy, are needed [22]. Lorenz and Schroll [14] prove equivalence between the relaxation criterion and the convergence of the solution as $\varepsilon \to 0$ in L^2 for linear systems with constant coefficients. It is also possible to show that existence of a strictly convex entropy function is not needed for linear systems to have a well-behaved limit [15,22]. Tzavaras [20] builds a framework for using the zero relaxation limit to approximate hyperbolic systems of conservation laws when the solutions of the limiting systems are assumed to be smooth.

1.1. *The subcharacteristic condition*

A key concept arising in the analysis of hyperbolic relaxation systems is the *subcharacteristic condition* [12,21], first introduced by Leray and subsequently independently found by Whitham. The modern terminology was introduced by Liu [13] for 2×2 systems.

For general $N \times N$ hyperbolic systems, the condition may be stated as follows [5].

Definition 1. *Let the N eigenvalues of the homogeneous system* (1.2) *be given by*

$$\lambda_1 \le \cdots \le \lambda_k \le \lambda_{k+1} \le \cdots \le \lambda_N, \tag{1.5}$$

i.e. λ_k are the eigenvalues of

$$\boldsymbol{A} = \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{U}}. \tag{1.6}$$

Let $\tilde{\lambda}_j$ be the n eigenvalues of the equilibrium system (1.4)*, i.e. $\tilde{\lambda}_j$ are the eigenvalues of*

$$\boldsymbol{B} = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{u}}. \tag{1.7}$$

Herein, the homogeneous system (1.2) *is applied to a local equilibrium state $\boldsymbol{U} = \mathcal{E}(\boldsymbol{u})$, such that*

$$\lambda_k = \lambda_k(\mathcal{E}(\boldsymbol{u})), \qquad \tilde{\lambda}_j = \tilde{\lambda}_j(\boldsymbol{u}). \tag{1.8}$$

The equilibrium system (1.4) *is said to satisfy the* ***subcharacteristic condition*** *with respect to* (1.2) *when the following statements hold:*

(1) all $\tilde{\lambda}_j$ are real;
(2) if the $\tilde{\lambda}_j$ are sorted in ascending order as

$$\tilde{\lambda}_1 \le \cdots \le \tilde{\lambda}_j \le \tilde{\lambda}_{j+1} \le \cdots \le \tilde{\lambda}_n, \tag{1.9}$$

then $\tilde{\lambda}_j$ are interlaced with λ_k in the following sense: Each $\tilde{\lambda}_j$ lies in the closed interval $[\lambda_j, \lambda_{j+N-m}]$.

Definition 2. *Assume that the subcharacteristic condition is satisfied in the sense of Definition 1. If in addition each* $\tilde{\lambda}_j$ *lies in the open interval* $(\lambda_j, \lambda_{j+N-m})$*, then the equilibrium system* (1.4) *is said to satisfy the* ***strict subcharacteristic condition*** *with respect to* (1.2).

Chen et al. [5] proved that for general $N \times N$ systems, their entropy condition implies the subcharacteristic condition. Yong [22] proved that for relaxation systems satisfying $n = N - 1$, the subcharacteristic condition is necessary for the linear stability of the equilibrium state; hence it is also necessary for convergence.

For 2×2 systems, it has been well established that stability of the equilibrium state is equivalent to the subcharacteristic condition [5,23]. In particular, Chen et al. [5] showed that the strict subcharacteristic condition is equivalent to their entropy condition in this case.

1.2. *Contributions of the current paper*

This paper is motivated by the observation that the modern mathematical literature focuses strongly on nonlinear analytical techniques, with the aim of obtaining asymptotic behavior and zero relaxation limits for nonlinear systems. However, there is also much physical insights to be gained through simple linear analyses on systems in the form (1.1). By linearizing the system and applying a von Neumann type analysis, one obtains dispersion relations giving the amplifications and velocities of individual Fourier components as functions of the wave number. Such analyses have been performed on two-phase flow models by for instance Städtke [19] and Solem et al. [18], and on the St. Venant equations by Barker et al. [3].

Following in the footsteps of Yong [23], a systematic study of the wave dynamics of general 2×2 systems was undertaken in [1]. Here it was established that the velocity of any isolated Fourier component will be a monotonic function of the wave number or the relaxation time. A critical phenomenon was also observed; if the eigenvalues of the homogeneous system are symmetric around the eigenvalue of the equilibrium system, a definite branching point in the wave number can be identified. At this wave number, both velocities and amplification factors are equal and non-differentiable.

Similar critical phenomena were observed for the two-phase flow model investigated in [18]. In particular, it was observed that if the ratio of the equilibrium and homogeneous sound speeds is less than 1/3, the characteristics of the sound waves cannot be continuously connected between the homogeneous and equilibrium limits as functions of the wave number or relaxation time; there exist concrete *transition points* where the system changes character in a very qualitative manner.

In this paper, we expand on the works [1,18] by considering general linear $N \times N$ systems with a stable relaxation operator of rank 1, i.e. $n = N - 1$. For this case, we prove a useful proposition:

P1: *The characteristic polynomial for any isolated Fourier component can be written* *as a convex combination of the limiting homogeneous and equilibrium eigenvalue polynomials.*

This result, following from elementary linear algebra and consistent with the observation made in [18], allows for obtaining dispersion relations for any such rank 1 hyperbolic relaxation system directly from the homogeneous and equilibrium eigenvalues; no explicit knowledge of the detailed structure of the relaxation operator is needed. Hence this proposition provides a tool for significantly simplifying the kind of von Neumann type analysis as performed in [1,18,19].

This proposition also provides an heuristic benefit in describing the fundamental relationship between stability, causality and the subcharacteristic condition. In particular, by using basic properties of polynomials established in the modern literature [8,24], we are able to provide direct and elementary proofs of the following expected results:

P2: *For strictly hyperbolic systems with a stable rank 1 relaxation term, the linear stability condition is precisely the subcharacteristic condition.*

P3: *If the subcharacteristic condition holds for such systems, a maximum principle follows: the velocity of any isolated Fourier component is bounded by the maximum and minimum eigenvalues of the homogeneous system.*

These propositions will be precisely formulated in the main part of our paper, which is organized as follows. In Section 2, we obtain the linearized system around the equilibrium state. In Section 2.2, we derive the characteristic polynomial for a Fourier component of wave number k and prove Proposition P1. In Section 3, we provide an elementary proof of Proposition P2; the equivalence between linear stability and the subcharacteristic condition. In Section 4, we provide an elementary proof of Proposition P3, which has the interpretation as a causality principle.

Finally, in Section 5, the results of our paper are summarized.

2. Linearized relaxation systems

Henceforth, we will consider linearized relaxation systems. Let $\boldsymbol{U}_{\mathrm{eq}} \in G$ be an equilibrium state, i.e. a constant N-vector characterized by $\boldsymbol{Q}(\boldsymbol{U}_{\mathrm{eq}}) = 0$. The relaxation system (1.1) linearized around $\boldsymbol{U}_{\mathrm{eq}}$ can then be written as

$$\partial_t \boldsymbol{V} + \boldsymbol{A}\,\partial_x \boldsymbol{V} = \frac{1}{\varepsilon}\boldsymbol{R}\boldsymbol{V}, \tag{2.1}$$

where $\boldsymbol{V} = \boldsymbol{U} - \boldsymbol{U}_{\mathrm{eq}}$. Herein

$$\boldsymbol{A} = \left.\frac{\partial \boldsymbol{F}(\boldsymbol{U})}{\partial \boldsymbol{U}}\right|_{\boldsymbol{U}_{\mathrm{eq}}} \quad \text{and} \quad \boldsymbol{R} = \left.\frac{\partial \boldsymbol{Q}(\boldsymbol{U})}{\partial \boldsymbol{U}}\right|_{\boldsymbol{U}_{\mathrm{eq}}} \tag{2.2}$$

are both $N \times N$ matrices with constant coefficients.

2.1. *Plane-wave solutions*

For the purpose of the present analysis, we write the solution to the linearized problem (2.1) in terms of its Fourier components. Following the approach of Yong [22,23], we assume initial data $\boldsymbol{V}(x,0) \in L^2([a,b])$, where $[a,b] \subset \mathbb{R}$ is some interval, and write the unique solution to the linear initial value problem as

$$\boldsymbol{V}(x,t) = \sum_k \boldsymbol{V}_k(x,t) = \sum_k \exp(\boldsymbol{H}(k)t)\exp(ikx)\hat{\boldsymbol{V}}(k) \tag{2.3}$$

where

$$\boldsymbol{H}(k) = \frac{1}{\varepsilon}\boldsymbol{R} - ik\boldsymbol{A}. \tag{2.4}$$

Furthermore, we can write $\boldsymbol{H} = \boldsymbol{PJP}^{-1}$, where $\boldsymbol{P}$ is the matrix of generalized eigenvectors and $\boldsymbol{J}$ is the corresponding Jordan matrix. Now let λ_j denote the eigenvalues of $\boldsymbol{H}$. The solution (2.3) can then be written as a combination of elementary waves as

$$\boldsymbol{V}(x,t) = \sum_k \sum_{j=1}^{N} \tilde{V}_j(k,t)\exp\left(ikx + \lambda_j t\right), \tag{2.5}$$

for some amplitudes $\tilde{V}_j(k,t)$, which are polynomials in t. Notice that there is a plane wave solution associated with each distinct eigenvalue. In particular, if $\boldsymbol{H}$ is diagonalizable, $\boldsymbol{J}$ reduces to the diagonal matrix consisting of the eigenvalues of $\boldsymbol{H}$, and $\tilde{V}_j(k,t) = \tilde{V}_j(k)$ for all j.

Considering (2.5), it is natural to introduce the dispersion relation

$$v_j(k) = -\frac{1}{k}\mathrm{Im}(\lambda_j) \tag{2.6}$$

and the dampening

$$f_j(k) = \mathrm{Re}(\lambda_j) \tag{2.7}$$

in order to further re-write the solution as

$$\boldsymbol{V}(x,t) = \sum_k \sum_{j=1}^{N} \tilde{V}_j(k,t)\exp(f_j(k)t)\exp\left(ik(x - v_j(k)t)\right). \tag{2.8}$$

This allows us to describe the full wave dynamics of the linear relaxation system (2.1) in terms of the eigenvalues of the matrix $\boldsymbol{H}$. Note also that since we have the symmetry

$$\boldsymbol{H}(k) = \overline{\boldsymbol{H}(-k)}, \tag{2.9}$$

we can study the system for wave numbers $k \in [0,\infty)$ without loss of generality.

2.1.1. *Stability*

We say that the relaxation system (1.1) is *linearly stable* if the solutions (2.3) to its linearization (2.1) around the equilibrium state $\boldsymbol{U}_{\mathrm{eq}}$ are bounded in L^2 for all $t \in [0, \infty)$. This is equivalent to the condition

$$|\exp(\boldsymbol{H}(k)t)| \leq C \quad \forall k \in \mathbb{R}, \tag{2.10}$$

where C is some positive constant and $|\cdot|$ denotes the L^2-norm for matrices. By making the variable transformations

$$\eta = \frac{t}{\varepsilon}, \qquad \xi = -kt \tag{2.11}$$

we may state the stability condition (2.10) in the following form:

Definition 3. *Consider the relaxation system* (1.1) *linearized as* (2.1) *around the state* $\boldsymbol{U}_{\mathrm{eq}}$*. Assume that there is a* $C > 0$ *such that*

$$|\exp(\eta R + i\xi A)| \leq C \tag{2.12}$$

for all $\eta \geq 0$ *and* $\xi \in \mathbb{R}$.

Then the equilibrium state $\boldsymbol{U}_{\mathrm{eq}}$ *is said to be* ***linearly stable****.*

This is precisely the *stability criterion* identified by Yong [22], as part of his stronger *relaxation criterion.*

We may now state the following Lemma [10].

Lemma 2.1. *Linear stability in the sense of Definition 3 is equivalent to the following statements being valid for all* k*:*

- *All eigenvalues* λ_j *of the matrix* $\boldsymbol{H}(k)$ *have a real part* $\mathrm{Re}(\lambda_j) \leq 0$.
- *If* J_r *is a Jordan block of the Jordan matrix* $\boldsymbol{J} = \boldsymbol{P}^{-1}\boldsymbol{H}\boldsymbol{P}$ *which corresponds to an eigenvalue* λ_j *with* $\mathrm{Re}(\lambda_j) = 0$*, then* J_r *has dimension* 1×1.

Proof. The proof is straightforward and can be found in [10]. □

We also define the stronger notion of *strict stability*:

Definition 4. *Assume that the equilibrium state* $\boldsymbol{U}_{\mathrm{eq}}$ *is linearly stable in the sense of Definition 3. If in addition all eigenvalues of the matrix* $\boldsymbol{H}(k)$ *have a real part* $\mathrm{Re}(\lambda_j) < 0$ *for all* k*, then the equilibrium state* $\boldsymbol{U}_{\mathrm{eq}}$ *is said to be* ***strictly linearly stable****.*

2.2. *The characteristic polynomial*

We assume that the relaxation matrix $\boldsymbol{R}$ is stable, i.e. it has no eigenvalues with positive real parts. For the general linear $N \times N$ system with rank 1 relaxation the matrix $\boldsymbol{H}(k)$ can then, up to a scaling and a similarity transform, be written as

$$\boldsymbol{H}(k) = \frac{1}{\varepsilon}\boldsymbol{R} - ik\boldsymbol{A} = \frac{1}{\varepsilon}\begin{pmatrix} 0 & \dots & 0 \\ \vdots & & \vdots \\ r_{N,1} & \dots & -1 \end{pmatrix} - ik\begin{pmatrix} a_{1,1} & \dots & a_{1,N} \\ \vdots & & \vdots \\ a_{N,1} & \dots & a_{N,N} \end{pmatrix}. \tag{2.13}$$

A crucial property of the characteristic polynomial of (2.13) is that it can be written as a convex sum of the polynomials of the homogeneous and equilibrium systems. To obtain this result, we first need to establish the following lemma:

Lemma 2.2. *Assume that the relaxation matrix is stable. In the context of* (2.13), *the characteristic polynomials for the homogeneous system and the equilibrium system are given by*

$$P_h(z) = \det(-i\boldsymbol{A} - z\boldsymbol{I}) \tag{2.14}$$

and

$$P_e(z) = -\det(-i\boldsymbol{C}\boldsymbol{D}^T - i\boldsymbol{A}_{N,N} - z\boldsymbol{I}), \tag{2.15}$$

respectively. In the above, $P_h(z)$ and $P_e(z)$ are the characteristic polynomials corresponding to the matrices

$$\begin{aligned} \boldsymbol{H}_h(k) &= -ik\boldsymbol{A} \\ \boldsymbol{H}_e(k) &= -ik\boldsymbol{B} = -ik(\boldsymbol{C}\boldsymbol{D}^T + \boldsymbol{A}_{N,N}), \end{aligned}$$

after rescaling the variable λ with $z = \lambda/k$. The vectors $\boldsymbol{D}$, $\boldsymbol{C}$ are given by

$$\boldsymbol{D} = \begin{pmatrix} r_{N,1} \\ \vdots \\ r_{N,N-1} \end{pmatrix} \tag{2.16}$$

$$\boldsymbol{C} = \begin{pmatrix} a_{1,N} \\ \vdots \\ a_{N-1,N} \end{pmatrix} \tag{2.17}$$

and

$$\boldsymbol{A}_{N,N} = \begin{pmatrix} a_{1,1} & \dots & a_{1,N-1} \\ \vdots & & \vdots \\ a_{N-1,1} & \dots & a_{N-1,N-1} \end{pmatrix}. \tag{2.18}$$

Proof. It is easily seen that (2.14) is the characteristic polynomial for the homogeneous system. To see that the characteristic polynomial for the equilibrium system satisfies (2.15), we look at solutions $\boldsymbol{V}$ satisfying $\boldsymbol{RV} = 0$. With $\boldsymbol{R}$ as in (2.13) and $\boldsymbol{V} = [V_1, V_2, \ldots, V_N]^T$, we have

$$\sum_{k=1}^{N-1} r_{Nk} V_k - V_N = 0, \tag{2.19}$$

such that the equilibrium system with $\boldsymbol{v} = [V_1, \ldots, V_{N-1}]$ is equal to

$$\partial_t \boldsymbol{v} + \begin{pmatrix} a_{11} & \cdots & a_{1,N-1} \\ \vdots & & \vdots \\ a_{N-1,1} & \cdots & a_{N-1,N-1} \end{pmatrix} + \begin{pmatrix} a_{1,N} \\ \vdots \\ a_{N-1,N} \end{pmatrix} \left(r_{N,1} \cdots r_{N,N-1} \right) \partial_x \boldsymbol{v} = 0 \tag{2.20}$$

Thus, the equilibrium system has the characteristic equation $\det(-i\boldsymbol{B} - z\boldsymbol{I}) = 0$ after rescaling λ with $z = \lambda/k$, and

$$\boldsymbol{B} = \boldsymbol{C}\boldsymbol{D}^T + \boldsymbol{A}_{N,N}. \tag{2.21}$$

□

We can now establish the following:

Proposition 2.3. *Assume that the relaxation matrix is stable. Let*

$$\chi = \frac{\varphi}{\varphi + 1} \tag{2.22}$$

with $\varphi = k\varepsilon$. *The characteristic polynomial for* (2.13) *can be written in the form*

$$\Psi(z) = \chi P_h(z) + (1 - \chi) P_e(z) = 0, \quad \chi \in [0, 1], \tag{2.23}$$

where $P_h(z)$ *and* $P_e(z)$ *are given by* (2.14) *and* (2.15), *respectively.*

Proof. We have

$$\boldsymbol{H} - \lambda \boldsymbol{I} = \frac{1}{\varepsilon}\boldsymbol{R} - ik\boldsymbol{A} - \lambda\boldsymbol{I} = \begin{pmatrix} -ika_{11} - \lambda & \cdots & -ika_{1N} \\ \vdots & & \vdots \\ \frac{r_{N1}}{\varepsilon} - ika_{N1} & \cdots & \frac{-1}{\varepsilon} - ika_{NN} - \lambda \end{pmatrix}. \tag{2.24}$$

Multiplying the characteristic equation of $\boldsymbol{H}$ with k^n, we get

$$\det\left(\frac{1}{\varphi}\boldsymbol{R} - i\boldsymbol{A} - z\boldsymbol{I}\right) = \det\begin{pmatrix} -ia_{11} - z & \cdots & -ia_{1N} \\ \vdots & & \vdots \\ \frac{r_{N1}}{\varphi} - ia_{N1} & \cdots & \frac{-1}{\varphi} - ia_{NN} - z \end{pmatrix} = 0, \tag{2.25}$$

where $\varphi = k\varepsilon$ and $z = \lambda/k$. Introducing $\boldsymbol{A}_{n,k}$ as the sub-matrix of $-i\boldsymbol{A} - z\boldsymbol{I}$ where the nth row and the kth column are removed, we have the characteristic equation in the following form,

$$\tilde{\Psi}(z) = \sum_{k=1}^{N-1} (-1)^{k-1} r_{Nk} \cdot \det(\boldsymbol{A}_{N,k}) - \det(\boldsymbol{A}_{N,N}) + \varphi \cdot \det(-i\boldsymbol{A} - z\boldsymbol{I}) = 0 \tag{2.26}$$

when expanding along the bottom row of (2.25). By (2.14), we may write (2.26) as

$$\tilde{\Psi}(z) = \tilde{P}_e(z) + \varphi P_h(z), \tag{2.27}$$

where

$$\tilde{P}_e(z) = \sum_{k=1}^{N-1} (-1)^{k-1} r_{Nk} \cdot \det(\boldsymbol{A}_{N,k}) - \det(\boldsymbol{A}_{N,N}). \tag{2.28}$$

We can now observe that (2.28) is equal to the characteristic polynomial for the equilibrium system (2.15) by the following calculation,

$$\begin{aligned}
\tilde{P}_e(z) =& \det \begin{pmatrix} -ia_{11} - z \cdots & -ia_{1,N-1} & -ia_{1N} \\ \vdots \quad \ddots & & \vdots \\ -ia_{N-1,1} \cdots & -ia_{N-1,N-1} - z & -ia_{N-1,N} \\ r_{N,1} \quad \cdots & r_{N,N-1} & -1 \end{pmatrix} \\
=& \det \begin{pmatrix} -ia_{11} - ia_{1N} r_{N,1} - z & \cdots & -ia_{1,N-1} - ia_{1N} r_{N,N-1} & 0 \\ \vdots & \ddots & & \vdots \\ -ia_{N-1,1} - ia_{N-1,n} r_{N,1} & \cdots & -ia_{N-1,N-1} - ia_{N-1,N} r_{N,N-1} - z & 0 \\ 0 & \cdots & 0 & -1 \end{pmatrix} \\
=& - \det(-i\boldsymbol{B} - z\boldsymbol{I}) = P_e(z),
\end{aligned} \tag{2.29}$$

where we have first added $-ia_{jN}$ multiplied with the last row to the jth row of (2.29) and then added $r_{N,i}$ multiplied with the last column to the jth column for $j = 1, \ldots, N-1$. Substituting (2.22) into (2.27) we obtain

$$\Psi(z) = (1-\chi)\tilde{\Psi}(z) = \chi P_h(z) + (1-\chi) P_e(z). \tag{2.30}$$

□

3. Linear stability

In this section we prove that strictly hyperbolic relaxation systems with stable rank 1 relaxation matrices are linearly stable if and only if the roots of the two limiting polynomials interlace on the imaginary axis, i.e. if and only if the relaxation system satisfies the subcharacteristic condition of Definition 1.

Let

$$\Psi(z) = \chi P_h(z) + (1-\chi) P_e(z) \tag{3.1}$$

be the eigenvalue polynomial for the $N \times N$ linear hyperbolic relaxation system (2.1) where $\boldsymbol{R}$ is of rank one and stable. Further, let $P_h(z)$ and $P_e(z)$ be as in Lemma 2.2. Since $\boldsymbol{A}$ is a $N \times N$ real matrix and $\boldsymbol{B}$ a $(N-1) \times (N-1)$ real matrix, the coefficients of $P_h(z)$ and $P_e(z)$ alternate between being purely real and purely imaginary in the following way:

$$P_h(z) = z^N + ib_{N-1}z^{N-1} + b_{N-2}z^{N-2} + \dots \tag{3.2}$$

$$P_e(z) = z^{N-1} + ic_{N-2}z^{N-2} + c_{N-3}z^{N-3} + \dots, \tag{3.3}$$

such that the full polynomial (3.1) satisfies

$$\begin{aligned}\Psi(z) =& \chi P_h(z) + (1-\chi)P_e(z) \\ =& \chi(z^N + ib_{N-1}z^{N-1} + b_{N-2}z^{N-2} + \dots) \\ & + (1-\chi)(z^{N-1} + ic_{N-2}z^{N-2} + c_{N-3}z^{N-3} + \dots).\end{aligned} \tag{3.4}$$

By rewriting the polynomial in this form, we are able to prove the following proposition.

Proposition 3.1. *Assume that* $\boldsymbol{R}$ *is a stable rank 1 relaxation matrix for the linearized relaxation system* (2.1)*. Let* $\Psi(z)$ *in* (3.1) *be its characteristic polynomial. Further, assume that the system is strictly hyperbolic. Then the system* (1.1) *is linearly stable for all* $\chi \in [0,1]$ *if and only if the roots of* $P_e(z)$ *are purely imaginary and interlace the roots of* $P_h(z)$ *on the imaginary axis, i.e. the subcharacteristic condition is satisfied.*

Further, the subcharacteristic condition is strictly satisfied if and only if the system is strictly linearly stable for all $\chi \in (0,1)$*.*

Before we prove Proposition 3.1 for relaxation systems, let us take a look at a general complex polynomial

$$P(z) = \sum_{k=0}^{N}(a_k + ib_k)z^k, \quad a_N + ib_N \neq 0. \tag{3.5}$$

This polynomial can be rewritten as $P(z) = m(z) + p(z)$, where $m(z)$ and $p(z)$ are the two axially complementary polynomials

$$m(z) = \frac{1}{2}\left[P(z) + \overline{P(-\overline{z})}\right] \tag{3.6}$$

$$p(z) = \frac{1}{2}\left[P(z) - \overline{P(-\overline{z})}\right]. \tag{3.7}$$

Let us assume that $m(z)$ and $p(z)$ have no roots in common. The order of $m(z)$ is one more than the order of $p(z)$ if N is even and the opposite if N is odd. Further, observe that the inequality

$$a_N a_{N-1} + b_N b_{N-1} > 0 \tag{3.8}$$

is necessary for (3.5) to be stable, as the coefficient for z^{n-1} is equal to minus the sum of all the roots and that the real part of the sum of the roots have to be less than zero. The following stability lemma exists for general polynomials [24].

Lemma 3.2. *The general complex polynomial* (3.5) *is strictly stable, i. e.* $\mathrm{Re}(\lambda) < 0$ *for all roots* λ*, if and only if* (3.8) *holds,* $m(z)$ *and* $p(z)$ *have no roots in common,* *their roots are distinct and purely imaginary and their roots interlace on the imaginary axis.*

Proof. The proof is presented in Zhareddine [24]. □

Now we are ready to prove Proposition 3.1.

Proof of Proposition 3.1. We assume that all roots $P_h(z)$ and $P_e(z)$ have in common have been factored out such that we are left with the reduced polynomial $\Psi_r(z)$. All the roots that $P_{e,r}$ and $P_{h,r}$ have in common satisfy $\mathrm{Re}(\lambda) = 0$. Since the system is assumed to be strictly hyperbolic, all the roots that $P_e(z)$ and $P_h(z)$ have in common are distinct. Thus, the Jordan blocks corresponding to these eigenvalues will have dimension 1×1 and, according to Lemma 2.1, they will not cause any linear instability in the sense of Definition 3.

We now split the remaining polynomial $\Psi_r(z)$ into two axially complementary polynomials as in (3.6),

$$m(z) = \frac{1}{2}\left[\Psi_r(z) + \overline{\Psi_r(-\overline{z})}\right] \tag{3.9}$$

$$p(z) = \frac{1}{2}\left[\Psi_r(z) - \overline{\Psi_r(-\overline{z})}\right]. \tag{3.10}$$

We observe that $m(z) = \chi P_{h,r}(z)$ and $p(z) = (1-\chi)P_{e,r}(z)$ if N is even and $p(z) = \chi P_{h,r}(z)$ and $m(z) = (1-\chi)P_{e,r}(z)$ if N is odd, making $P_h(z)$ and $P_e(z)$ axially complementary. We easily see that for $\chi = 1$ we have the homogeneous eigenvalue polynomial and for $\chi = 0$ we have the equilibrium eigenvalue polynomial.

From now on, we look at $\chi \in (0,1)$. Corresponding to the coefficients for the general polynomial (3.5), $\Psi_r(z)$ has

$$a_N = \chi, \quad b_N = 0, \quad a_{N-1} = (1-\chi), \tag{3.11}$$

such that (3.8) always is fulfilled. It now follows from Lemma 3.2 that the roots $\{z_j\}$ of $\Psi_r(z)$ satisfy $\mathrm{Re}(z_j) < 0$ if and only if the roots of $P_{h,r}(z)$ interlace the roots of $P_{e,r}$ and their roots are distinct and purely imaginary.

If $P_e(z)$ and $P_h(z)$ have no roots in common, we have $\mathrm{Re}(z_j) < 0$ for all roots of $\Psi(z)$ when $\chi \in (0,1)$, making the system strictly linearly stable.

If all roots of $P_e(z)$ are roots of $P_h(z)$, the remaining eigenvalue polynomial will have one root, $\Psi_r(z) = \chi(z - z_k) + (\chi - 1)$, where z_k is a root of $P_h(z)$. This root is always stable as $0 \leq \chi \leq 1$. □

With Proposition 3.1, we have now shown that there is an equivalence between linear stability and the subcharacteristic condition for hyperbolic relaxation systems with stable rank one relaxation matrices. We can further observe that linear stability implies that the linear equilibrium system must be strictly hyperbolic.

4. A maximum principle

We show that the velocities of the linearized hyperbolic relaxation system 2.1 can never exceed the velocities of the corresponding homogeneous system when the system is linearly stable. We prove this with the help of some properties for polynomials found in Fisk [8].

Let (3.1) be the eigenvalue polynomial for the strictly hyperbolic $N \times N$ linearized relaxation system with a relaxation matrix of rank 1. Assume that the system is linearly stable. Let $\Psi_r(z)$ be the reduced polynomial where all the roots that $P_e(z)$ and $P_h(z)$ have in common are factored out. Then, by Proposition 3.1, the roots of $P_{h,r}(z)$ strictly interlace the roots of $P_{e,r}(z)$ on the imaginary axis. We make a translation of the roots from the left half plane to the lower half plane,

$$\begin{aligned}\hat{\Psi}_r(z) &= i^N \Psi_r(-iz) \\ &= i^N \chi P_{h,r}(-iz) + i^N (1-\chi) P_{e,r}(-iz) \\ &= h(z) + ig(z)\end{aligned} \tag{4.1}$$

The roots of $h(z)$ and $g(z)$ in (4.1) interlace on the real axis. Further, the real roots of $h(z)$ and $g(z)$ correspond to the roots of $P_{h,r}(z)$ and $P_{e,r}(z)$ on the imaginary axis. Since the roots $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ of $h(z)$ strictly interlace the roots $\{\alpha_1, \alpha_2, \dots, \alpha_{n-1}\}$ of $g(z)$, $g(z)$ changes sign on the roots of $h(z)$.

$h(z)$ is of order N and of one order more than $g(z)$. The homogeneous system is assumed to be strictly hyperbolic, making the roots of $h(z)$ distinct such that

$$\frac{h(z)}{z-\lambda_1}, \quad \frac{h(z)}{z-\lambda_2}, \dots, \frac{h(z)}{z-\lambda_N} \tag{4.2}$$

is a basis for all real polynomials with real roots of order $N-1$. We can therefore express $g(z)$ with basis

$$g(z) = \sum_{k=1}^{N} c_k \frac{h(z)}{z-\lambda_k}. \tag{4.3}$$

The c_ks have the same sign if the eigenvalue polynomial (3.1) is strictly stable. For a root λ_k of $h(z)$, we have

$$g(\lambda_k) = c_k(\lambda_k - \lambda_1)\dots(\lambda_k - \lambda_{k-1})(\lambda_k - \lambda_{k+1})\dots(\lambda_k - \lambda_N), \tag{4.4}$$

such that

$$\operatorname{sgn}(g(\lambda_k)) = \operatorname{sgn}(c_k)(-1)^{k+N} \tag{4.5}$$

and we can see that $g(z)$ changes sign on the roots of $h(z)$ if all the c_ks have the same sign. The c_ks also have to be strictly greater than zero. If not, λ_k would also be a root of $g(z)$, which contradicts the fact that $h(z)$ and $g(z)$ have no roots in common. $P_e(z)$ in (3.1) has a positive leading coefficient which makes all the c_ks positive.

We can now prove the following proposition.

Proposition 4.1. *Let* (2.1) *be a strictly hyperbolic relaxation system with a stable rank 1 relaxation matrix. Let the system be linearly stable. Then the imaginary parts of the roots,* $\mathrm{Im}(z_k)$*, for* $k = 1, \ldots, N$*, of* (3.1) *satisfies*

$$\min_i\{\lambda_i)\} \leq \mathrm{Im}(z_k) \leq \max_i\{\lambda_i\} \tag{4.6}$$

where $i\lambda_k$ *are the roots of* $P_h(z)$*, for all* $\chi \in [0, 1]$*.*

Proof. When the system is stable, the reduced polynomial of (3.1) is strictly stable for all $\chi \in (0, 1)$, the roots of $\chi P_{h,r}(z)$ strictly interlace the roots of $(1 - \chi)P_{e,r}(z)$ on the imaginary axis. We look at the translated polynomial in (4.1). We can write (4.1) as

$$\hat{\Psi}_r(z) = h(z) + i \sum_{k=1}^{N} c_k \frac{h(z)}{z - \lambda_k}. \tag{4.7}$$

For a root z_i of (4.1), we will have

$$0 = 1 + i \sum_{k=1}^{N} c_k \frac{1}{z_i - \lambda_k}. \tag{4.8}$$

Assume that z_i is a root of $\hat{\Psi}_r(z)$ with $\mathrm{Re}(z_i) > \lambda_k$ for all $k = 1, \ldots, N$. All the c_ks are greater than zero, making the real part of the sum in (4.8) greater than zero, such that the right hand side cannot be equal to zero. Therefore, there are no roots z_i of (3.1) with real part greater than all the roots of $h(z)$. The proof for $\mathrm{Re}(z_i) < \lambda_k$ is similar.

We conclude that (4.1) has no roots with real part greater than or smaller than the real roots of $h(z)$. Translating (4.1) back to (3.1), we observe that the real part of the roots in (4.1) correspond to the imaginary parts of the roots in (3.1).

The roots that $P_e(z)$ and $P_h(z)$ have in common are constant for any $\chi \in [0, 1]$ and will never be able to exceed any maximum or minimum value. □

Remark 4.2. The converse direction of Proposition 4.1 does not hold. We can easily generate two polynomials $P_1(z)$ and $P_2(z)$ satisfying the maximum principle that do not interlace, making the convex combination unstable,

$$P_1(z) = (z + i5)(z + i)(z - i2) \tag{4.9}$$

$$P_2(z) = (z + i4)(z + i2). \tag{4.10}$$

5. Summary

We have provided some fundamental and elementary results pertaining to the von Neumann type analysis of linearized hyperbolic relaxation systems where the relaxation operator is assumed to be stable and of rank 1. Our results may be briefly summarized as follows:

P1: The characteristic polynomial for any Fourier component of wave number k may be directly obtained as a convex combination of the eigenvalue polynomials for the homogeneous and equilibrium limits.

P2: A strictly hyperbolic relaxation system with a stable rank 1 relaxation operator is linearly stable if and only if the subcharacteristic condition is satisfied.

P3: If the subcharacteristic condition is satisfied, the velocity of any isolated Fourier component is bounded by the maximum and minimum eigenvalue of the homogeneous system.

Herein, it should be noted that the proof of P1 is obtained from elementary linear algebra and the statements of P2 and P3 are unsurprising given the already established strong relationship between stability and the subcharacteristic condition [5,23]. In our opinion, the main interest of our paper lies in the connection provided between theory describing general properties of roots of polynomials and fundamental causality and stability properties of hyperbolic relaxation systems. These connections seem so far to have been given little emphasis in the literature.

References

[1] P. Aursand and T. Flåtten, On the dispersive wave dynamics of 2×2 relaxation systems, *J. Hyperbol. Differ. Eq.* **9**, 641–659, (2012).

[2] A. Aw and M. Rascle, Resurrection of second order models for traffic flow, *J. Appl. Math.* **60**, 916–938, (2000).

[3] B. Barker, M. A. Johnson, L. M. Rodrigues and K. Zumbrun, Metastability of solitary roll wave solutions of the St. Venant equations with viscosity, *Phys D* **240**, 1289–1310, (2011).

[4] G.-Q. Chen, Relaxation limit for conservation laws, *Z. Angew. Math. Mech.* **76**, 381–384, (1996).

[5] G.-Q. Chen, C. D. Levermore and T.-P. Liu, Hyperbolic conservation laws with stiff relaxation terms and entropy, *Comm. Pure Appl. Math.* **47**, 787–830, (1994).

[6] G. Q. Chen and T. P. Liu, Zero Relaxation and Dissipation Limits for Hyperbolic Conservation Laws, *Comm. Pure Appl. Math.* **46**, 755–781, (1993).

[7] S. Evje, A compressible two-phase model with pressure-dependent well-reservoir interaction, *SIAM J. Math. Anal.* **45**, 518–546, (2013).

[8] S. Fisk, *Polynomials, Roots and Interlacing*, (2008). `http://arxiv.org/abs/math/0612833v2`

[9] T. Flåtten and H. Lund, Relaxation two-phase flow models and the subcharacteristic condition, *Math. Models Methods Appl. Sci.* **21**, 2379–2407, (2011).

[10] H.-O. Kreiss and J. Lorenz, *Initial-boundary value problems and the Navier-Stokes equations*, Academic Press, (1989).

[11] C. Lattanzio and P. Marcati, The Zero Relaxation Limit for 2×2 Hyperbolic Systems, *Non-linear anal.* **38**, 375–389, (1999).

[12] J. Leray, *Hyperbolic Differential Equations*, The Institute for Advanced Study, (1953).

[13] T. P. Liu, Hyperbolic Conservation Laws with Relaxation, *Comm. Math. Phys.* **108**, 153–175, (1987).

[14] J. Lorenz and H. J. Schroll, Stiff well-posedness for hyperbolic systems with large relaxation terms (linear constant-coefficient problems), *Adv. Differential Equations* **2**, 643–666, (1997).

[15] J. Lorenz and H. J. Schroll, Hyperbolic Systems With Relaxation: Characterization of Stiff Well-Posedness and Asymptotic Expansions, *J. Math. Anal. Appl.* **235**, 497–532, (1999).

[16] H. Lund, A hierarchy of relaxation models for two-phase flow, *SIAM J. Appl. Math.* **72**, 1713–1741, (2012).

[17] M. Pelanti and K.-M. Shyue, A mixture-energy-consistent six-equation two-phase numerical model for fluids with interfaces, cavitation and evaporation waves, *J. Comput. Phys.* **259**, 331–357, (2014).

[18] S. Solem, P. Aursand and T. Flåtten, The dispersive wave dynamics of a two-phase flow relaxation model, *submitted*, (2013).
Available online at `http://www.pvv.org/~pederka/phaserelax.pdf`

[19] H. Städtke, *Gasdynamic Aspects of Two-Phase Flow* Wiley-VCH, Weinheim, Germany, (2006)

[20] A. Tzavaras, Relative Entropy in Hyperbolic Relaxation, *Comm. Math. Sci.* **3**, 101–259, (2005).

[21] G. B. Whitham, *Linear and Nonlinear Waves* John Wiley & Sons, New York, (1974).

[22] W.-A. Yong, Basic aspects of hyperbolic relaxation systems, in Advances in the Theory of Shock Waves, *Progr. Nonlinear Differential Equations Appl.* **47**, 259–305, Birkhäuser Boston, Boston, (2001).

[23] W.-A. Yong, Basic structures of hyperbolic relaxation systems, *P. Roy. Soc. Edinb. A*, **132**, 1259–1274, (2002).

[24] Z. Zahreddine, On the interlacing property of stable complex systems of differential equations, *Indian J. pure appl. Math*, **26**, 643–647, (1995)