Università degli Studi di Catania

Dipartimento di Matematica e Informatica

Corso di Laurea Magistrale in Matematica

Giuseppe Virgilio Minissale

On the dispersive behaviour of hyperbolic systems in periodic media

Tesi di Laurea

Relatore: prof. Giovanni Russo

Anno Accademico 2023–2024

Contents

Chapter 1

Introduction

In order to explain the reasons of this work, let us first look at an example. Let us consider the Saint-Venant model for shallow water equations given by

$$h_t + (hu)_x = 0, \tag{1.1}$$

$$(hu)_t + \left(hu^2 + \frac{1}{2}gh^2\right)_x = -ghb_x, \tag{1.2}$$

where h is the water depth, u is the depth-averaged velocity, b denotes the bottom elevation or bathymetry and g is the gravitational acceleration. What one can observe is the different behaviour of the solutions of the system with respect to the bathymetry. What one can expect, since we have a hyperbolic system of conservation laws, is that classic solutions exist until a certain time and then we have generalized solutions that present themselves as shocks. That is what happens at least for a smooth bathymetry. On the other hand, when a periodic bathymetry is considered, the numerical solution shows us a different behaviour: we observe a separation of wave packets with different velocity, similarly to what happens in the framework of dispersive equations.

In Figure 1.1, we can see what was said: while the solution of the system with a flat bottom shows the expected behaviour of N-wave formation and subsequent decay, in the periodic bottom case the initial pulse breaks up into a train of what are apparently solitary waves. This behaviour has been observed numerically but, while in the linear waves framework their propagation over periodic media have been studied profoundly (see, for example, [23]), the more general and coherent with the real world experience case of

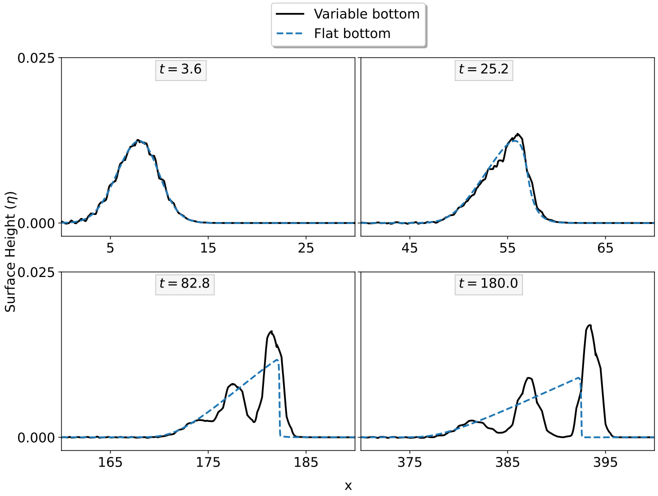


Figure 1.1: Evolution of an initial Gaussian pulse over periodic bathymetry. The surface elevation is shown, measured in meters, For comparison, the dashed blue line shows the solution for flow over a flat bottom. This figure is reproduced from [7] with the permission of the authors.

nonlinear waves in periodic structures has not received the same interest, even if it should. Hence, the purpose of this work is to explain this numerical behaviour trying to detect these dispersive properties from the equations themselves instead of just observing them from the numerical simulation.

First of all, in Chapter 2 the quasilinear hyperbolic systems of conservation laws, main object of our studies, will be introduced. We will give their definition and we will show how it is possible to construct solutions that contain shocks, which are not possible in the classic theory of differential equations. Then, we will deduce some further properties in order to have a complete understanding of this type of systems, at least for our purposes.

Chapter 3 is dedicated to the definition and study of numerical methods for systems introduced in Chapter 2, where the theoretical properties will be translated into a numerical framework. Some of these methods have been used in this work to construct the solutions of the original version of the systems in the models studied in Chapter 6.

Then, in Chapter 4, we will explain what "dispersive behaviour" means, introducing the dispersive equations. They are equations characterized by the fact that their wave solutions spread out in space as they evolve in time. The properties of these equations, such as the presence of special solutions called travelling wave and the presence of conserved quantities, will be studied afterwards, basing their description on the study of the Korteweg-de Vries equation, which is the benchmark in the study of dispersive equations and from where it is easier to explain and understand all the features of this family of equations.

In Chapter 5, we will introduce the perturbation theory, that comprises methods for the approximation of solutions of differential equations in terms of formal power series in some small parameter.

These type of techniques, and mainly the so called multiple-scale method, will be used in Chapter 6 to derive, from the original systems that describe our models, other "approximate" systems, that will be indicated as homogenized systems, which solutions are very close to the one of the original system but where the dispersive nature is clean-cut just from the definition of these systems themselves. In this chapter, everything done previously will be applied on three models, where the new versions of the original system derived with the perturbation techniques will be studied in depth. We will compare the solutions of this new system with the ones of the dispersive systems derived, observing the closeness between them at least for early times, and then we will analyse the properties of these systems that are typical of dispersive equations.

While in Chapter 2-5 everything presented is literature findable in various books, and Chapter 6 treats more advanced topics but still is the based on the works of other authors, in Chapter 7 and 8 we will present some original discoveries developed by the writer. In Chapter 7, we will construct, using an iterative method called Petviashvili method, the travelling wave for the homogenized Euler system of the fifth order presented in Chapter 6, which computation was an open problem from the original paper where this system has been introduced and analysed, while in Chapter 8 the possible presence of conserved quantities in the homogenized systems have been studied and numerically validated, with some interesting discoveries.

Chapter 2

An overview of quasilinear hyperbolic systems of conservation laws

2.1 Introduction

In this chapter, we introduce the theoretical results about the systems we are interested to, namely the *hyperbolic systems of conservation laws.* Since a conservation law is an integral relation, it may be satisfied by functions which are not differentiable, not even continuous. We will indicate them as generalized solutions, in contrast to the regular, i.e. differentiable ones. The breakdown of a regular solution may merely mean that although a generalized solution exists for all time, it ceases to be differentiable after a finite time. All available evidence indicates that this is so. It turns out however that there are many generalized solutions with the same initial data, only one of which has physical significance; the task is to give a criterion for selecting the right one. The criterions we will use are called entropy conditions, for in the gas dynamic they amount to requiring the increase of entropy of particles crossing a shock front. In this chapter, we will analyze only the mathematical side of this theory, showing results about existence and uniqueness of generalized solutions subject to an entropy condition. Finally, we will see some examples. The contents of this chapter have been mainly taken from [12].

2.2 Quasi-linear hyperbolic systems

Let us start with the basic case.

Definition 2.1. A first order system of quasi-linear equations in two independent variables has the form

$$u_t + Au_x = 0, \tag{2.1}$$

where u is a vector function of x and t and A a matrix function of x, t, u. Such system is *hyperbolic* if A is diagonalizable. When A has real and distinct eigenvalues $\lambda_j = \lambda_j(x, t, u)$ for each x, t, u we say that is *strictly hyperbolic.* We will mainly work with in this framework. Similarly, a quasi-linear system in $k+1$ independent variables $t, x_1, \ldots, x_k$

$$u_t + \sum_{i=1}^{k} A_i u_{x_i} = 0, \quad A_i = A_i(x, t, u) \tag{2.2}$$

is *strictly hyperbolic* if for each x, t, u and unit vector ω the matrix

$$\sum_{i=1}^{k} A_i \omega_i \tag{2.3}$$

has real and distinct eigenvalues $\lambda_j = \lambda_j(x, t, u, \omega)$.

The most natural question we can ask ourselves is if, prescribed $u(x, 0) = u_0(x)$ initial value at $t = 0$, this system has a solution $u(x, t)$. One thing that we can surely say is that, in the class of C^1 solutions, if the solution exists it has to be unique. Let us sketch a proof of this fact. Taken u, v different solutions of (2.1), the difference $w = u - v$ satisfies

$$w_t + A(u)w_x + (A(u) - A(v))v_x = 0. \tag{2.4}$$

To show that $w(x, t) = 0$ for all x and t, we can use an energy-based approach. Consider the L^2 norm of w, i.e. $\|w(\cdot, t)\|_{L^2}$. We calculate the time derivative of this norm:

$$\frac{d}{dt}\|w(\cdot, t)\|_{L^2}^2 = \frac{d}{dt}\int |w(x, t)|^2 \, dx. \tag{2.5}$$

Taking the time derivative and using the equation for w, we get:

$$\begin{aligned}&\frac{d}{dt}\|w\|_{L^2}^2 = 2\int w \cdot \partial_t w \, dx = \\ &-2\int w \cdot A(u)\, \partial_x w \, dx - 2\int w \cdot (A(u) - A(v))\, \partial_x v \, dx.\end{aligned} \tag{2.6}$$

For the first term, we use the fact that $A(u)$ is a hyperbolic matrix, which implies that its eigenvalues are real, so we can apply the properties of A and integration by part in order to say that this term is smaller than something of the kind $C_1\|w\|_{L^2}$. For the second term, we leverage the continuity of $A(u)$ with respect to u, i.e., $A(u) - A(v) \approx \frac{\partial A}{\partial u} \cdot w$ for $u \approx v$. This allows us to bound the difference term in terms of the L^2 norm of w. Combining these steps, we obtain an inequality of the form:

$$\frac{d}{dt}\|w\|_{L^2} \leq C\|w\|_{L^2}, \tag{2.7}$$

where C is a constant depending on u and v, but not on t. Applying Grönwall's inequality, we conclude that $\|w\|_{L^2} = 0$ for all t, which implies $w(x,t) = 0$ for all x and t. We have thus proven that $u(x,t) = v(x,t)$ for all x and t, establishing the uniqueness of the solution for the quasilinear hyperbolic system of conservation laws in the case of C^1 solutions. We need now to know something about the existence of the solution. It is possible to prove that smooth solutions exists only in a finite time interval, while they do not exist beyond them. This is a problem since the equations usually represents a physical system so a solution is required for every time t. We need to define some kind of "generalized solution", and this is possible in the framework of conservation laws.

2.3 Conservation laws

A *conservation law* is an equation of the form

$$u_t + \nabla \cdot f = 0. \tag{2.8}$$

This equation expresses the physical principle that the total amount of a given quantity in a fixed region D changes over time only due to the flux of that quantity across the boundary of the region. In a mathematical form, if u and f represent the density and flux of the quantity, respectively, the conservation principle states that

$$\frac{d}{dt}\int_D u\,dx = -\int_{\partial D} f\cdot n\,dS. \tag{2.9}$$

Applying the divergence theorem and assuming we can interchange the order of differentiation and integration, we obtain

$$\int_D (u_t + \nabla\cdot f)\,dx = 0. \tag{2.10}$$

Since this must hold for any domain D, it follows that equation (2.8) must be satisfied. We can deal with systems of conservation laws

$$u_t^j + \nabla\cdot f^j = 0, \quad j = 1,\dots,n, \tag{2.11}$$

where $f^j$is a nonlinear function of $u^1,\dots,u^n$. Carrying out the differentiation in 2.11 and using a compact notation we can rewrite it as

$$u_t + \sum_{i=1}^{k} A_i\omega_i, \tag{2.12}$$

where $A_i = \nabla_u f_i$. The matrices A_i are functions of u, and we assume that the quasi-linear system is strictly hyperbolic. Now we can generalize the definition of solution.

Definition 2.2. A function u is said to be a generalized solution of the system of conservation laws if it satisfies the integral relation

$$\int_D u^j dx\Big|_{t_1}^{t_2} + \int_{t_1}^{t_2}\int_{\partial D} f^j\cdot ndSdt = 0 \tag{2.13}$$

for every smoothly bounded domain and for every time interval (t_1, t_2).

This definition is equivalent to say that the equation holds in the sense of distribution theory. Let us analyse first the case of a single conservation law.

2.3.1 Single conservation laws

We focus on the equation of the form

$$u_t + f_x = 0, \tag{2.14}$$

where f is a nonlinear function of u. We can rewrite it as

$$u_t + a(u)u_x = 0, \quad a(u) = \frac{df}{du}, \tag{2.15}$$

which means that u is constant along trajectories $x = x(t)$ that propagates with speed $a(u) = \frac{dx}{dt}$. This special trajectories are called *characteristics.* Those characteristics are straight lines, since u is constant along these curves and so it is the velocity of propagation, and this allows us to "solve" geometrically our initial value problem with initial condition $u(x, 0) = u_0(x)$: drawing straight lines from a point y of the x-axis with speed $a(u_0(y))$, near the x-axis the solution $u(x, t)$ is uniquely determined. We can also construct the solution analytically. From the geometrical observation, we can say that

$$u - u_0(x - ta(u)) = 0 \tag{2.16}$$

holds, so, assuming u_0 differentiable and using the implicit function theorem, for x, t small enough we have

$$u_t = -\frac{u_0' a}{1 + u_0' a_u t}$$
$$u_x = \frac{u_0'}{1 + u_0' a_u t}$$

and substituting in (2.14) we see that u implicitly defined satisfies (2.14). Let us now assume now that (2.14) is genuinely non linear, i.e that $a_u \neq 0$ for all u (for example $a_u > 0$). Then if $u_0' \geq 0$ we have a solution of (2.14) for all $t > 0$, since geometrically we can see that in this case the characteristic cover the whole half-plane t>0, while if $u_0' < 0$ at some point both u_t and u_x become bigger and bigger as $1 + u_0' a_u t$ approaches zero, so in this case there will be two points y_1, y_2 such that $y_1 < y_2$ but $u_0(y_1) > u_0(y_2)$, which means that the characteristics that develop from these points will interesect at time $t = (y_2 - y_1)/(a_1 - a_2)$. This implies that at this time the solution u cannot exist since it take on both the values $u_0(y_1)$ and $u_0(y_2)$. This argument proves that if $a(u_0(x))$ is not an increasing function then there is no function $u(x,t)$ that solves the equation in the ordinary sense. We need to involve distribution solutions in order to solve this problem. We start with the simplest case, where our distribution solution satisfies the equation in the ordinary sense on each side of a smooth curve $x = y(t)$ across which u is discontinuous, denoting with u_l, u_r the values of u on the left and right sides of this curve. Choosing the $x-$interval $[a,b]$ so that the curve y intersect this interval at time t and denoting with $I(t)$ the quantity

$$I(t) = \int_a^b u(x,t)dx = \int_a^y u(x,t)dx + \int_y^b u(x,t)dx, \tag{2.17}$$

we have

$$\frac{dI(t)}{dt} = \int_a^y u_t dx + u_l s + \int_y^b u_t dx + u_r s, \tag{2.18}$$

where

$$s = \frac{dy}{dt} \tag{2.19}$$

is the speed of propagation of the discontinuity. Using the fact that $u_t = -f_x$ and carrying out the integration, we obtain

$$\frac{dI}{dt} = f(u(a)) - f(u_l) + u_l s - f(u(b)) + f(u_r) - u_r s. \tag{2.20}$$

Using the conservation law

$$\frac{dI}{dt} = f(u(a)) - f(u(b)), \tag{2.21}$$

we obtain the jump condition

$$s[u] = [f], \tag{2.22}$$

where $[u] = u_r - u_l$ and $[f] = f_r - f_l$. This relation, that is possible to generalize in the spatial multi-dimensional case, is usually called the *Rankine-Hugoniot* jump condition.

Let us see a famous example, the so called *Burgers' equation*, in order to see the strength of this approach. Taken $f(u) = \frac{1}{2}u^2$ and

$$u_0(x) = \begin{cases} 1 & \text{for} \quad x \leq 0, \\ 1-x & \text{for} \quad 0 < x \leq 1 \\ 0 & \text{for} \quad x > 1 \end{cases}, \tag{2.23}$$

the "classic" solution that can be obtained with both the geometric and the analytic approach is single-valued just for $t \leq 1$, while is double-valued thereafter. Defined, for $t \geq 1$,

$$u(x,t) = \begin{cases} 1 & \text{for} \quad x < \frac{1+t}{2}, \\ 0 & \text{for} \quad \frac{1+t}{2} < x \end{cases}, \tag{2.24}$$

we have that this u verify in the generalized sense the equation. In this case, the discontinuity starts in the point $(1,1)$, that separates the state $u_l = 1$ on the left with the state u_r on the right. The speed of propagation has been chosen according to the jump condition

$$s = \frac{f(u_0(1)) - f(u_0(0))}{u_0(1) - u_0(0)} = \frac{0 - 1/2}{0 - 1} = \frac{1}{2}. \tag{2.25}$$

Introducing generalized solutions can be a double-edge sword: while it allows to solve unsolvable equations in the classical sense, at the same time there could be several generalized solutions with the same initial data. Let us now analyse in deep the Burgers' equation: consider the initial condition

$$u_0(x) = \begin{cases} 0 & \text{for} \quad x < 0, \\ 1 & \text{for} \quad x > 0 \end{cases} \tag{2.26}$$

The geometric solution is single valued for $t > 0$ but is not defined in for $0 < x < t$. We can address this point taking

$$u(x,t) = \begin{cases} 0 & \text{for} \quad x < \frac{t}{2} \\ 1 & \text{for} \quad x > \frac{t}{2} \end{cases}. \tag{2.27}$$

This u solves the equation in the distribution sense and the speed of propagation has been chosen so that the jump condition is satisfied.Conversely, the function

$$u(x,t) = \frac{x}{t} \tag{2.28}$$

solves the differential equation (2.14) and connects continuously with the rest of the solution obtained geometrically. We have two different solutions, and only one has physical meaning, so we need to decide a criterion in order to select in an appropriate way the solution of our problem. The criterion we impose is the following: The characteristics originating from both sides of the discontinuity curve, when extended in the direction of increasing t, intersect the discontinuity line. This occurs when

$$a(u_l) > s > a(u_r). \tag{2.29}$$

. This condition is called *entropy condition*, and a solution that satisfies both the jump relation (2.22) and the entropy condition (2.29) will be called a *shock*.

In our example, the solution defined by cases (2.27) violates the entropy condition since $a(u_l) = 0, s = \frac{1}{2}, a(u_r) = 1$. Our task is now to see if the initial value problem for (2.14) has exactly one generalized solution, defined for all $t \geq 0$ and with only shocks as discontinuities. Let us work with the assumption that $a_u > 0$, i.e. $f(u)$ is a convex function. This means that

$$f(u) \geq f(v) + a(v)(u - v). \tag{2.30}$$

Let u be a classic solution of (2.14) and assume $u_0(x)$ is 0 for x large enough negative; then this still holds for $u(x,t)$ for any $t > 0$ for which is defined. Introducing the integrand function $U(x,t)$ defined as

$$U(x,t) = \int_{-\infty}^{x} u(y,t)dy, \tag{2.31}$$

then $U_x = u$. Integrating (2.14) from $-\infty$ to x we obtain, supposed $f(0) = 0$,

$$U_t + f(U_x) = 0. \tag{2.32}$$

Applying the convexity inequality with $u = U_x$ and any number v to this last equality we have

$$U_t + a(v)U_x \leq a(v)v - f(v). \tag{2.33}$$

Denote by y the point such that

$$\frac{x-y}{t} = a(v), \tag{2.34}$$

integrating (2.33) along this line from 0 to t we obtain, for $t \geq 0$,

$$U(x,t) \leq U(y,0) + t(a(v)v - f(v)). \tag{2.35}$$

Calling b the inverse of the function a, we obtain

$$b\Big(\frac{x-y}{t}\Big) = v, \tag{2.36}$$

and denoting by g the function

$$g(z) = a(v)v - f(v), \quad v = b(z), \tag{2.37}$$

it is clear, since a, b are inverse functions, that

$$\frac{dg}{dz} = v\frac{da}{dv}\frac{db}{dz} = b(z). \tag{2.38}$$

Denoting $a(0)$ by c, we have $b(c) = 0$ and so $g(c) = 0$. Introducing the function g on the right side of (2.35) we obtain

$$U(x,t) \leq U(y,0) + tg(\frac{x-y}{t}). \tag{2.39}$$

This inequality holds for all choices of y, so it holds for the value of y such that v is equal to u, and for this very value the equality also holds.

We obtain the following theorem, taken from [12].

Theorem 2.3. *Given u a continuous and differentiable solution of* (2.14), *then we have*

$$u(x,t) = b(\frac{x-y}{t}), \tag{2.40}$$

where $y = y(x,t)$ is that value which minimizes

$$G(x,y,t) := U_0(y) + tg(\frac{x-y}{t}), \tag{2.41}$$

b is the inverse function of a, g is defined by

$$\frac{dg}{dz} = b(z), \quad g(c) = 0, \quad \textit{where } a(0) = c, \tag{2.42}$$

and

$$U_0(y) = \int_{-\infty}^{y} u_0(x)dx, \quad u_0(x) = u(x,0). \tag{2.43}$$

Let us try now to find something similar for generalized solutions. Let u be a solution of this kind. Since relation (2.32) is the integral form of (2.14), when f is convex the inequality (2.35) holds even for generalized solutions. If all discontinuities are shocks, every point (x,t) can be connected to a point y on the initial line by a backward characteristic. This point y is the value that satisfies the equality in (2.39), so (2.3) also applies to generalized solutions of (2.14) whose discontinuities are shocks. We can summarize this result in the following theorem, taken from [12].

Theorem 2.4. *Formulas* (2.40)-(2.41) *define a possibly discontinuous function $u(x,t)$ for arbitrary integrable initial value $u_0(x)$; the function u so defined satisfies the equation* (2.14) *in the sense of distribution, and the discontinuities are shocks.*

We omit the proof of this theorem, which is very technical. One can find the proof in [12].
Let us focus now on the uniqueness problem, introducing this theorem taken from [12].

Theorem 2.5. *Let u, v two generalized solutions of* (2.14) *with only shocks as discontinuities and assume that f is convex. Then*

$$||u(t) - v(t)|| \tag{2.44}$$

is a decreasing function of t, where the norm is the L^1 norm with respect to x.

Corollary 2.6. *If $u = v$ at $t = 0$, $u = v$ for all $t \geq 0$.*

This is the uniqueness result we were looking for. Again, the proof of this result can be found in [12]. Let us now try to omit the assumption on the convexity of f in the uniqueness theorem. We can do this replacing the entropy condition (2.29) with the following two *generalized entropy conditions*:

- if $u_r < u_l$, then

$$f(\alpha u_r + (1-\alpha)u_l) \leq \alpha f(u_r) + (1-\alpha)f(u_l), \quad 0 \leq \alpha \leq 1; \tag{2.45}$$

- if $u_r > u_l$, then

$$f(\alpha u_r + (1-\alpha)u_l) \geq \alpha f(u_r) + (1-\alpha)f(u_l), \quad 0 \leq \alpha \leq 1; \tag{2.46}$$

In the case of $f \in C^1$ arbitrary function of u, a discontinuity of u is a shock if u_r and u_l satisfy one of the entropy conditions. We can generalize the uniqueness results to this general case.

Theorem 2.7. *Let u, v two generalized solutions of* (2.14) *with only shocks as discontinuities. Then*

$$||u(t) - v(t)|| \tag{2.47}$$

is a decreasing function of t, where the norm is the L^1 norm with respect to x.

Corollary 2.8. *If u is a generalized solution of* (2.14) *of which the discontinuities fails to verify the entropy condition* (2.29)*, then there is a genuine solution v such that*

$$||u(t) - v(t)|| \tag{2.48}$$

is not a decreasing function of t.

The proof of these results are similar to the ones with the convexity assumption. For the uniqueness theorem to be interesting, we need the generalized entropy condition not to be too restrictive, so that every initial value problem has a generalized solution u. We construct this solution as a so called *viscosity solution.* Such solution can de derived as the limit for

$\epsilon \to 0$ of the solution u_ϵ of the parabolic initial value problem

$$u_t + f_x = \epsilon u_{xx}, \quad \epsilon > 0, \quad u_\epsilon(x,0) = u_0(x). \tag{2.49}$$

This problem has at most one solution from the maximum principle; moreover, the solution u_ε converges in L^1 to a limit function u as $\varepsilon \to 0$. Let us show that u is a generalized solution and that satisfies the entropy condition. For the first statement, multiplying both sides by a test function and integrating by parts we obtain

$$-\int\int [\phi_t u_\varepsilon + \phi_t f(u_\varepsilon)]dxdt = \varepsilon \int \phi_{xx} u_\varepsilon, \tag{2.50}$$

so for $\varepsilon \to 0$ we obtain

$$-\int\int [\phi_t u + \phi_t f(u)]dxdt = 0, \tag{2.51}$$

so u is a distribution solution of (2.14).

Let us now verify that the entropy condition is verified by this u. First of all, we show that for any $\varepsilon > 0$, $u_\varepsilon, v_\varepsilon$ solutions of (2.49) we have that

$$||u_\varepsilon(t) - v_\varepsilon(t)|| \tag{2.52}$$

is a decreasing function of t. First of all, we write

$$||u_\varepsilon - v_\varepsilon|| = \sum_n (-1)^n \int_{y_n}^{y_{n+1}} (u_\varepsilon - v_\varepsilon)dx, \tag{2.53}$$

where $u_\varepsilon - v_\varepsilon$ changes sign at the points y_n. In those point, we have for continuity that

$$u_\varepsilon - v_\varepsilon = 0. \tag{2.54}$$

Differentiating $||u_\varepsilon - v_\varepsilon||$ we obtain

$$\frac{d}{dt}||u_\varepsilon - v_\varepsilon|| = \sum_n (-1)^n \int_{y_n}^{y_{n+1}} \frac{d}{dt}(u_\varepsilon - v_\varepsilon)dx + \sum_n (-1)^n (u_\varepsilon - v_\varepsilon)\frac{dy}{dt}\Big|_{y_n}^{y_{n+1}}, \tag{2.55}$$

and carrying out the integration we obtain

$$\frac{d}{dt}||u_\varepsilon - v_\varepsilon|| = \sum_n (-1)^n \varepsilon \frac{\partial}{\partial x}(u_\varepsilon - v_\varepsilon)\Big|_{y_n}^{y_{n+1}} - \sum_n (-1)^n (f(u_\varepsilon) - f(v_\varepsilon))\Big|_{y_n}^{y_{n+1}}. \tag{2.56}$$

The second sum is zero from the (2.54); the first term is non positive because the term $(-1)^n\varepsilon(u_\varepsilon - v_\varepsilon)$ is nonnegative in the interval and zero in the endpoints, so its derivative is non negative at the endpoint and nonpositive at the right point, that induce that the final sum produces a nonpositive result. Now, passing to the limit for $\varepsilon \to 0$, we have $||u(t) - v(t)||$ is decreasing, so we have from the Corollary 2.8 we have that all the discontinuities of u satisfy the generalized entropy condition.

2.3.2 Hyperbolic systems of conservation laws

Let us focus now on systems like

$$\frac{\partial}{\partial t}u_i + \frac{\partial}{\partial x}f_i = 0, \quad i = 1, \dots, n, \tag{2.57}$$

where $f_i = f_i(u_1, \dots, u_n)$. Differentiating, we obtain the quasi-linear system

$$u_t + A(u)u_x = 0, \tag{2.58}$$

where $u = (u_1, \dots, u_n)^T$ and A is the matrix whose rows are the gradients of f_i with respect to u. In order to have a strictly hyperbolic system, we assume that A has real, distinct eigenvalues $\lambda_1, \dots, \lambda_n$, that depends on u as their eigenvectors. The condition for nonlinearity, that is an important side of the study that we are doing, in this case is not just $\nabla_u \lambda_k \neq 0$ anymore, but

$$\nabla_u \lambda_k \cdot r_k \neq 0, \tag{2.59}$$

i.e. the gradient of the eigenvalue λ_k has to be not orthogonal to r_k, its corresponding eigenvector. Normalizing, the previous condition becomes

$$\nabla_u \lambda_k \cdot r_k = 1. \tag{2.60}$$

Let us try to find some analogy from the (2.14) case. For example, the Rankine-Hugoniot jump condition becomes in this case

$$s[u_k] = [f_k], \quad k = 1, \dots, n, \tag{2.61}$$

i.e. each one of the conservation laws must satisfy the jump condition across every discontinuity. About the entropy condition, we require that for some index k, $1 \leq k \leq n$,

$$\lambda_k(u_l) > s > \lambda_k(u_r) \tag{2.62}$$

$$\lambda_{k-1}(u_l) < s < \lambda_{k+1}(u_r). \tag{2.63}$$

These inequalities assert that k characteristics bump on the line of discontinuity from the left and $n - k + 1$ from the right, for a total of $n + 1$. The information given by these characteristics plus the $n - 1$ relations obtained from the jump condition are sufficient to determine the $2n$ values which u takes on both side of the line of discontinuity. A discontinuity across which the jump condition and the entropy condition are satisfied is called a *k-shock*. We have a theorem that describes k-shocks, taken again from [12].

Theorem 2.9. *The set of states u_r near u_l which are connected to some given state u_l through a k-shock form a smooth one-parameter family $u_r = u(\varepsilon)$, $\varepsilon_0 < \varepsilon \leq 0, u(0) = u_l$. The shock speed s is also a smooth function of ε.*

There is a class of continuous solutions that can be useful in the construction of generalized solutions, the so called *centered rarefaction waves.* These are solution which depend only on the ratio $(x - x_0)/(t - t_0)$, so in the case $x_0 = 0, t_0 = 0$ they are solutions of the form

$$u(x,t) = h\Big(\frac{x}{t}\Big). \tag{2.64}$$

Let us see how one can construct a piecewise smooth solutions that involves a centered rarefaction wave. Let $\xi = x/t$, substituting u into (2.11) we have

$$-\frac{x}{t^2}h' + \frac{1}{t}Ah' = 0, \tag{2.65}$$

which can be seen as

$$[A(h) - \xi]h' = 0. \tag{2.66}$$

This is satisfied by

$$\xi = \lambda(h(\xi)), \quad h' = \alpha r(h), \tag{2.67}$$

where $\lambda = \lambda_k$ is one of the eigenvalues of A, and using this last relation and supposing (2.60) condition holds we can obtain

$$h' = r(h). \tag{2.68}$$

h is called *$k-$rarefaction wave.* This is a differential equation which as a unique solution h satisfying the initial condition $h'(\lambda) = u_l$ and this h is defined for all ξ close to λ. Let ε small enough such that h is defined and u_r the value $u_r = h(\lambda + \varepsilon)$, we can construct the piecewise smooth solution u defined for $t \geq 0$ as follows:

$$u(x,t) = \begin{cases} u_l & \text{for} \quad x < \lambda t, \\ h(x/t) & \text{for} \quad \lambda t \leq x \leq (\lambda + \varepsilon)t \\ u_r & \text{for} \quad (\lambda + \varepsilon)t < x \end{cases} \tag{2.69}$$

This function u satisfies the differential equation in each region and it is continuous across the separating lines. In this way, we connected two states u_l, u_r through a centred rarefaction wave. If instead condition (2.60) does not hold, then (2.11) has discontinuous solutions whose speed of propagation is

$$s = \lambda_k(u_l) = \lambda_k(u_r). \tag{2.70}$$

They are called *contact discontinuities.* All this can be useful to solve an important initial value problem in both the theoretical and application field, the so called *Riemman initial value problem.* This is the initial value problem (2.11) where the initial function is

$$u(x,t) = \begin{cases} u_0 & \text{for} \quad x < 0, \\ u_n & \text{for} \quad x > 0, \end{cases} \tag{2.71}$$

with u_0, u_n two vectors. We have the following theorem from [12].

Theorem 2.10. *If u_0, u_n are sufficiently close, the initial value problem* (2.11)-(2.71) *has a solution. This solution consists of $n+1$ constant states $u_0, u_1, \dots, u_n$ separated by centered rarefaction or shock waves.*

In [5] it is possible to find a method for solving any initial value problem where the oscillation of u_0 is small.

Let us now focus on the criteria for the selection of solutions. We saw some conditions (that we called entropy conditions), like (2.29) or (2.62), that allows us to reject certain solutions even though the conservation laws are satisfied. Let us try now to do a step further, trying to define a function, that we will call *entropy*, which is related to this criteria. Let us consider the system (2.11). What is needed for η function of $u_1, \dots, u_n$ in order to satisfy the conservation law, i.e. to have

$$\eta_t + \psi_x = 0? \tag{2.72}$$

Carrying out the differentiation we have

$$\nabla_u \eta \cdot u_t + \nabla_u \psi \cdot u_x = 0, \tag{2.73}$$

and since the system is satisfied if we find something like

$$u_t + A u_x = 0, \tag{2.74}$$

we multiply this last equality by $\nabla_u \eta$, obtaining

$$\nabla_u \eta \cdot u_t + \nabla_u \eta A u_x = 0, \tag{2.75}$$

so putting everything together we can say that η satisfies a conservation law if and only if

$$\nabla_u \eta A = \nabla_u \psi \tag{2.76}$$

holds. This is a system of n PDEs which is overdetermined and has no solutions in general for $n \geq 2$. One case where a solution exists is when A is a symmetric matrix, i.e.

$$\frac{\partial f_j}{\partial u_i} = \frac{\partial f_i}{\partial u_j}. \tag{2.77}$$

This relation is the compatibility relation for the existence of a function $g(u)$ such that $\frac{\partial g}{\partial u_i} = f_i$. It is easy to verify that

$$\eta = \sum u_j^2, \quad \psi = \sum u_j f_j - g \tag{2.78}$$

satisfy (2.76).

As said before, entropy conditions are a powerful tool for the selection of discontinuous solutions which are physically realizable from the ones that are not. Another way to do this selection, following a procedure already introduced in some way in the (2.14) case, is to see the physically realizable solutions as the limit of solution of equations with a small dissipative mechanism, for example the *viscosity solution*, which is, as already seen, the limit of the solution u_λ of the parabolic system

$$u_t + Au_x = \lambda u_{xx}, \quad \lambda > 0. \tag{2.79}$$

Multiplying (2.79) by $\nabla_u \eta$ we have, if (2.72) is satisfied,

$$\eta_t + \psi_x = \lambda \nabla_u \eta \cdot u_{xx}. \tag{2.80}$$

Since

$$\eta_{xx} = \nabla_u \eta \cdot u_x x + \eta_{ij} u_x^i u_x^j, \quad \eta_{ij} = \frac{\partial^2}{\partial u^i \partial u^j} \tag{2.81}$$

supposing η convex (so the matrix η_{ij} is positive definite) we deduce

$$\eta_{xx} \geq \nabla_u \eta \cdot u_{xx}, \tag{2.82}$$

and substituting into (2.80) we have

$$\eta_t + \psi_x \leq \lambda \eta_{xx}. \tag{2.83}$$

Letting $\lambda \to 0$ we deduce the following theorem, again taken from [12].

Theorem 2.11. *Let* (2.11) *be a system of conservation laws which implies* *an additional conservation law* (2.72) *and suppose η is strictly convex. Let u be a distribution solution of* (2.11) *which is the limit of solutions of* (2.79) *containing the artificial viscous term. Then u satisfies the inequality*

$$\eta(u)_t + \psi(u)_x \leq 0. \tag{2.84}$$

Moreover, the following statements hold:

- *$\int \eta(t)dx$ is a finite, decreasing function of t;*
- *Supposed u piecewise continuous, across a discontinuity we have*

$$s[\eta_l - \eta_r] - [\psi_l - \psi_r] \leq 0. \tag{2.85}$$

The condition (2.84)-(2.85) can be called *entropy conditions.* The following result, that we will not prove but the proof of which can be found in [12] and is based on theorem 2.9, shows the compatibility of the "new" entropy conditions with the previous versions.

Theorem 2.12. *Let* (2.11) *be a system of hyperbolic and nonlinear conservation laws which implies an additional conservation law* (2.72) *and suppose U is strictly convex; let u be a solution of* (2.11) *in the integral sense which has a discontinuity propagating with speed s. Suppose that the values on the two sides of the discontinuity are close, then the entropy condition* (2.62) *holds if and only if* (2.85) *holds.*

To conclude this section, we state a theorem that links the entropy condition of the system (2.11) with the entropy conditions of the equation (2.14).

Theorem 2.13. *The generalized entropy conditions* (2.45)-(2.46) *are satisfied if and only if they are satisfied for all η, ψ which satisfy* (2.76) *and where η is convex.*

2.4 Decay of shocks

Let us now study the asymptotic behaviour for large time of solutions of conservation laws with form (2.14) that satisfies the entropy condition (2.29), assuming $a(u)$ an increasing function of u. First of all, let us give some definitions that will be useful.

Definition 2.14. Let $f : [a,b] \to \mathbb{R}$ be a real-valued function. We define:

- the *total variation* of f on the interval $[a,b]$ as:

$$V(f;[a,b]) = \sup \sum_{i=1}^{n} |f(x_{i+1}) - f(x_i)|, \tag{2.86}$$

- the *total increasing variation* of f on $[a,b]$ as:

$$V^{+}(f;[a,b]) = \sup \sum_{i=1}^{n} \max\{f(x_{i+1}) - f(x_i), 0\}, \tag{2.87}$$

- the *total decreasing variation* of f on $[a,b]$ as:

$$V^{-}(f;[a,b]) = \sup \sum_{i=1}^{n} \max\{f(x_i) - f(x_{i+1}), 0\}, \tag{2.88}$$

where all the supremums are taken over all partitions $a = x_0 < x_1 < \cdots < x_n = b$ of the interval.

As remarked previously, any differentiable solution u is constant along characteristics $\frac{dx}{dt} = a(u) = f'(u)$. Let $x_1(t), x_2(t)$ be a pair of characteristics, for $0 \leq t \leq T$. Then there is a whole one-parameter family of characteristics connecting the points of the interval $[x_1(0), x_2(0)]$ with the points of the interval $[x_1(T), x_2(T)]$. Since u is constant along these characteristics $u(x,0)$ on the first interval and $u(x,T)$ on the second interval are equivariant, i.e. the total increasing and the total decreasing variations of a differentiable solution between any pair of characteristics are conserved.

Denoting by $D(t) = x_2(t) - x_1(t) > 0$ and differentiating it, we have

$$\frac{d}{dt}D(t) = \frac{dx_2}{dt} - \frac{dx_1}{dt} = a(u_2) - a(u_1). \tag{2.89}$$

Integrating with respect to t between 0 and T we have

$$D(T) = D(0) + [a(u_2) - a(u_1)]T. \tag{2.90}$$

Suppose there is a shock present in u between the characteristic x_1, x_2. Since, according to the entropy condition (2.29), characteristics on either side of a shock run into the shock, there exist for any given time T two characteristics y_1, y_2 that intersect the shock y at exactly time T. Assuming that there are no other shocks, we conclude that the increasing variation of u on $(x_1(t), y_1(t))$, as well as on $(x_2(t), y_2(t))$, is independent of t. According to condition (2.29), u decreases across shocks, so the increasing variation along $[x_1(T), x_2(T)]$ equals the sum of the increasing variations of u along $[x_1(0), y_1(0)]$ plus the one along $[y_2(0), x_2(0)]$, that in general is bigger than the one along $[x_1(0), x_2(0)]$. Therefore we conclude that if shocks are present, the total increasing variation of u between two characteristics decreases with time. In the following we will estimate this decrease.

Theorem 2.15. *Let u be a possible discontinuous solution of the conservation law $u_t + f_x = 0$, where f is three times differentiable and strictly convex. Suppose that all discontinuities of u satisfy the entropy condition* (2.29) *and that $u(x,0)$ has compact support. Then:*

a) the lenght of the support of $u(x,t)$ is $O(\sqrt{t})$;

b) $\max_x |u(x,t)| = O\left(\frac{1}{\sqrt{t}}\right)$.

Theorem 2.16. *Let u be a possible discontinuous solution of the conservation law $u_t + f_x = 0$, where f strictly convex and $f'' > k > 0$. Suppose that all discontinuities of u satisfy the entropy condition* (2.29) *and that u is periodic in x with period p. Then:*

a) the total variation of u at time t does not exceed $\frac{2p}{kt}$;

b) the inequality

$$|u(x,t) - \overline{u}| \leq \frac{1}{kt}, \tag{2.91}$$

where $\overline{u} = \frac{1}{p}\int_0^p u(x,t)dx$.

The proofs of these two theorems can be found in [13].

We have some surprising results, almost paradoxical:

- the inequality (2.91) holds for every solutions with period p, no matter what is the amplitude of the initial disturbance, that can only have influence on the time when (2.91) holds (namely the larger the amplitude, the sooner the convergence);

- according to the theorem we just enunciated, given u_1 the initial function which is zero outside the interval $[0,p]$, and defined $u_2(x)$ to be equal to $u_1(x)$ in $[0,p]$, and periodic, then u_2, even if represent a much larger initial disturbance than u_1, nevertheless decays faster than u_1.

2.5 Gas Dynamics and 1D Euler equations

In order to see an example or some application about the things previously introduced, let us focus now on the particular case of the *Euler Equations*, a system of partial differential equations that describes an adiabatic and inviscid flow. The system is derived from three different conservation laws of three different quantities that are required to be conserved in order to have a physically valid model. We will not describe the derivation of this equations, that can be found in several texts (as [14]). The system of Euler equations is

$$\rho_t + (\rho u)_x = 0 \tag{2.92}$$

$$(\rho u)_t + (\rho u^2 + p)_x = 0 \tag{2.93}$$

$$E_t + ((E+p)u)_x = 0, \tag{2.94}$$

where ρ is the density, p the pressure, E is the energy and u is the speed. To obtain a closed system, we still need to specify the equation of state relation the internal energy to pressure and density. In the case of a polytropic ideal gas, the equation of state that we seek is

$$E = \frac{p}{\gamma - 1} + \frac{1}{2}\rho u^2, \tag{2.95}$$

where γ is the ratio of the specific heats and is a constant. The derivation of this equation can be also found in [14]. In a physical system, we can define the entropy of the system as a physical quantity. It can be interesting if one can link the entropy with some entropy condition. The entropy per unit mass of the system in question is

$$s = c_v \log(p/\rho^\gamma) + constant. \tag{2.96}$$

We can manipulate the Euler equations to derive the relation

$$s_t + us_x = 0, \tag{2.97}$$

which allows us to obtain a new Euler equations system, even not in conservation form

$$\begin{aligned} \rho_t + (\rho u)_x &= 0 & (2.98)\\ (\rho u)_t + (\rho u^2 + p)_x &= 0 & (2.99)\\ s_t + us_x &= 0. & (2.100) \end{aligned}$$

The most important property of entropy is that in smooth flow it remains constant on each particle path, so, when a particle crosses a shock, the entropy can only jump to an higher value. This is a physical entropy condition for shocks. Note that combining the first and third equations we obtain a conservation law for $S = \rho s$ given by

$$S_t + u(S)_x = 0. \tag{2.101}$$

This equation holds for smooth solutions, but not for generalized solution, indeed the entropy is not conserved across shocks.

Let us study the nonlinearity of the system in the sense of (2.60). First of all, let us compute the matrix A: we have

$$A = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2}(\gamma-3)u^2 & (3-\gamma)u & \gamma-1 \\ \frac{1}{2}(\gamma-1)u^3 - uH & H-(\gamma-1)u^2 & \gamma u, \end{bmatrix} \tag{2.102}$$

where

$$H = \frac{E+p}{\rho} \tag{2.103}$$

is the total specific enthalpy. The eigenvalues are

$$\lambda_1 = u - c, \quad \lambda_2 = u, \quad \lambda_3 = u + c, \tag{2.104}$$

and the eigenvectors are

$$r_1 = [1, u - c, H - uc]^T, \quad r_2 = [1, u, \frac{1}{2}u^2]^T, \quad r_3 = [1, u + c, H + uc]^T. \tag{2.105}$$

Note that

$$\nabla\lambda_2 = [-\frac{u}{\rho}, \frac{1}{\rho}, 0]^T, \tag{2.106}$$

and we have $\nabla\lambda_2 \cdot r_2 = 0$, so the $2-th$ characteristic is linearly degenerate and we have contact discontinuities. In this case the discontinuities propagate with speed λ_2 on each side, while the other two characteristics are genuinely nonlinear and admit shock waves or centered rarefaction.

Chapter 3

Numerical methods for systems of conservation laws

3.1 Introduction

In this section, we will introduce the numerical method that we need in our study and their properties and flaws. We will focus mainly on the theory that we need to develop in order to build appropriate methods, such as theorem for consistency, convergence and construction of a discrete entropy condition, defining some methods but analyzing mainly the Godunov's method while giving just some informations about other type of schemes. For more details about those schemes, see [17], while the chapter is mainly derived from [14] and [15].

3.2 Conservative finite volume methods

Consider the conservative system

$$u_t + \frac{d}{dx}f(u) = 0, \tag{3.1}$$

defined in the domain $\Omega = \{(x,t) : 0 \leq t \leq t_{final}, a \leq x \leq b\}$. Let us see how we can construct a conservative fully discrete finite volume method, where the adjective "conservative" will be explained later. Dividing the spatial domain $[a,b]$ into N subintervals $[x_{i-1/2}, x_{i+1/2}], i = 1 \ldots, N$ and the time interval in steps of size Δt, we start integrating the 3.1 in space and time over $[x_{i-1/2}, x_{i+1/2}] \times [t_n, t_{n+1}]$, obtaining

$$\int_{x_{i-1/2}}^{x_{i+1/2}} u(x, t_{n+1})dx =$$

$$= \int_{x_{i-1/2}}^{x_{i+1/2}} u(x, t_n)dx - \Big(\int_{t_n}^{t_{n+1}} f(u(x_{i+1/2}, t))dt - \int_{t_n}^{t_{n+1}} f(u(x_{i-1/2}, t))dt \Big),$$

which yields, divided everything by $\Delta x_i = x_{i+1/2} - x_{i-1/2}$ and defining the quantities *average of* u at time t^n in the i-th cell and the time average of the *Flux* between t^n and t^{n+1} at $x = x_{i+1/2}$

$$\overline{u}_i^n = \frac{1}{\Delta x_i} \int_{x_{i-1/2}}^{x_{i+1/2}} u(x, t_n)dx \tag{3.2}$$

$$F_{i+1/2} = \int_{t_n}^{t_{n+1}} f(u(x_{i-1/2}, t))dt \tag{3.3}$$

to the relation

$$\overline{u}_i^{n+1} = \overline{u}_i^n - \frac{\Delta t}{\Delta x_i}(F_{i+1/2} - F_{i-1/2}). \tag{3.4}$$

If a method can be written in this way we say that it is in *conservation form.* This is what we mean when we talk about conservative methods.

Let us work with a uniform partition of the spatial interval,i.e. $\Delta x_i = \Delta x \quad \forall i = 1, \ldots, n$. Let us try to see how one can compute, even approximately, $\overline{u}_i^{n+1}$ from $\overline{u}_i^n$. For this purpose, we need to define an approximation at time t_n, said $F^n_{i+1/2}$, of the flux $F_{i+1/2}$ in terms of known quantities. The main idea is to consider the flux $F^n_{i+1/2}$ as a function of $\overline{u}_i^n, \overline{u}_{i+1}^n$, that we can write in the form of some numerical flux function as

$$F^n_{i+1/2} = \mathcal{F}(\overline{u}_i^n, \overline{u}_{i+1}^n). \tag{3.5}$$

In this way, the value of $\overline{u}_i^{n+1}$ will depend on $\overline{u}_{i-1}^n, \overline{u}_i^n, \overline{u}_{i+1}^n$, defining an explicit method with a three-point stencil that depends on the definition of $\mathcal{F}$. For example, we can define

$$F^n_{i+1/2} = \frac{1}{2}(f(\overline{u}_i^n) + f(\overline{u}_{i+1}^n)),$$

and the equation becomes

$$\overline{u}_i^{n+1} = \overline{u}_i^n - \frac{\Delta t}{2\Delta x}(f(\overline{u}_{i+1}^n) + f(\overline{u}_{i-1}^n)).$$

However, this method is stable only under a very restrictive stability condition. Let us see another more sophisticate method.

Example 3.1. *The* Lax-Friedrich method *is*

$$\overline{u}_i^{n+1} = \frac{1}{2}(\overline{u}_{i-1}^n + \overline{u}_{i+1}^n) - \frac{\Delta t}{2\Delta x}(f(\overline{u}_{i+1}^n) + f(\overline{u}_{i-1}^n)).$$

This can be written in conservation form as

$$\overline{u}_i^{n+1} = \overline{u}_i^n - \frac{\Delta t}{\Delta x_i}(F_{i+1/2} - F_{i-1/2}). \tag{3.6}$$

where

$$F_{i+1/2} = \mathcal{F}(\overline{u}_i^n, \overline{u}_{i+1}^n),\ \mathcal{F}(W_-, W_+) = \frac{1}{2}(f(W_-)+f(W_+))-\frac{1}{2}\frac{\Delta x}{\Delta t}(W_+-W_-).$$

This method is really similar to the first one, but the diffusion term reduce the instability.

3.2.1 Consistency

Let us study the consistency of the method (3.4). This method is consistent with the original conservation law if the numerical flux function $\mathcal{F}$ reduces to the true flux in the case of constant flow, i.e.

$$\mathcal{F}(\tilde{u}, \tilde{u}) = f(\tilde{u}) \quad \forall \tilde{u} \in \mathbb{R}\,.$$

For this purpose, is sufficient to require $\mathcal{F}$ to be Lipschitz continuous function of each variable, i.e. at every point $\tilde{u}$ exists $K \geq 0$

$$|\mathcal{F}(v, w) - f(\tilde{u})| \leq K \max(|v - \tilde{u}|, |w - \tilde{u}|)$$

for all v, w with $|v - \tilde{u}|, |w - \tilde{u}|$ sufficiently small.

3.2.2 The Lax-Wendroff Theorem and discrete entropy condition

Now we ask ourselves if we can hope to correctly approximate discontinuous weak solutions to the conservation law by using a conservative method. We have the following theorem.

Theorem 3.2 (Lax-Wendroff). *Consider a sequence of grids indexed by l with mesh parameters $\Delta t_l, \Delta x_l \to 0$ as $l \to \infty$. Let $U_l(x,t)$ denote the numerical approximation computed with a consistent and conservative method on the l-th grid and suppose that U_l converges to a function u as $l \to \infty$. Then $u(x,t)$ is a weak solution of the conservation law.*

For more details about the type of convergence and for the proof see [15]. This theorem does not guarantee that weak solutions obtained in this manner satisfy some type of entropy condition as (2.84) and there are many examples of conservative numerical methods that converge to weak solutions violating the entropy condition. For some numerical methods, it is possible to show that this can never happen, and that any weak solution obtained by refining the grid must in fact satisfy the entropy condition. Of course this supposes that we have a suitable entropy condition for the system to begin with, and the most convenient form is typically the entropy inequality. Recalling the (2.84), in order to show that the weak solutoin obtained as limit of U_l satisfies this inequality, it suffices to show that a discrete entropy inequality holds, of the form

$$\eta(U_j^{n+1}) \leq \eta(U_j^n) - \frac{\Delta t}{\delta x}[\Psi(U_j^n) - \Psi(U_{j-1}^n)], \tag{3.7}$$

where Ψ is some numerical entropy flux function that must be consistent with ψ in the same manner that we required $\mathcal{F}$ to be consistent with f. If this holds for some suitable Ψ, then we can assure that the limiting weak solution u satisfies the entropy condition (2.84).

3.3 Godunov's Method

Let us start this section with an example about singular conservation laws of the kind $u_t + f(u)_x = 0$.

Example 3.3. *The* Upwind *method is*

$$\overline{u}_i^{n+1} = \overline{u}_i^n - \frac{\Delta t}{\Delta x}(f(\overline{u}_i^n) + f(\overline{u}_{i-1}^n)),$$

that can be written in conservation form as

$$\overline{u}_i^{n+1} = \overline{u}_i^n - \frac{\Delta t}{\Delta x_i}(F_{i+1/2} - F_{i-1/2}). \tag{3.8}$$

where

$$F_{i+1/2} = f(\overline{u}_i^n).$$

This one-sided method can be generalized in the case of linear systems by diagonalizing the Jacobian matrix and finding a system of equation of the type $u_{it} + \lambda_i u_{ix} = 0$. This can not be done in the nonlinear case since the eigenvalues are functions and can have mixed sign, so the method can not be one-sided. What one can do is trying to generalize the upwind method not with a diagonalization of a matrix but by the resolution of a Riemann problem. In Godunov's method, we use the numerical solution $\overline{u}^n$ to define a piecewise constant function $\tilde{u}^n(x, t^n)$ with the value $\overline{u}_i^n$ in the grid cell $[x_{i-1/2}, x_{i+1/2}]$, and we use $\tilde{u}^n(x, t^n)$ as initial data for the conservation law, which now we solve exactly in the time interval $[t_n, t_{n+1}]$. This can be done since $\tilde{u}$ is piecewise continuous defines a sequence of Riemann problems, so the exact solution at time t_{n+1} is obtained simply piecing together these Riemann solutions. After this, we can define the approximate solution $\overline{u}_{n+1}$ by averaging the exact solution at time t_{n+1}, obtaining

$$\overline{u}_i^{n+1} = \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} \tilde{u}^n(x, t_{n+1})dx.$$

Then we can iterate the process to compute $\tilde{u}^{n+1}$ and so on. Let us show that this method can be written in conservation form. Since $\tilde{u}^n$ is assumed to be an exact (weak) solution, we know that

$$\int_{x_{i-1/2}}^{x_{i+1/2}} \tilde{u}^n(x, t_{n+1})\,dx = \int_{x_{i-1/2}}^{x_{i+1/2}} \tilde{u}^n(x, t_n)\,dx -$$
$$\left(\int_{t_n}^{t_{n+1}} f(\tilde{u}^n(x_{i+1/2}, t))\,dt - \int_{t_n}^{t_{n+1}} f(\tilde{u}^n(x_{i-1/2}, t))\,dt\right).$$

Dividing by Δx and noticing that $\tilde{u}^n(x, t_n) = \overline{u}_i^n$ over the cell $[x_{i-1/2}, x_{i+1/2}]$

the equation reduces to

$$\overline{u}_i^{n+1} = \overline{u}_i^n - \frac{\Delta t}{\Delta x}(\mathcal{F}(\overline{u}_j^n, \overline{u}_{j+1}^n) - \mathcal{F}(\overline{u}_{j-1}^n, \overline{u}_j^n)),$$

where the numerical flux function is defined as

$$\mathcal{F}(\overline{u}_j^n, \overline{u}_{j+1}^n) = \frac{1}{\Delta t}\int_{t_n}^{t_{n+1}} f(\tilde{u}(x_{i+1/2}, t))dt.$$

Moreover, since $\tilde{u}^n$ is constant at point $x_{i+1/2}$ over the interval (from the structure of the Riemann problem itself), and since this constant value depend on $\overline{u}_j^n, \overline{u}_{j+1}^n$, say $u^*(\overline{u}_j^n, \overline{u}_{j+1}^n)$, the flux reduces to

$$\mathcal{F}(\overline{u}_j^n, \overline{u}_{j+1}^n) = f(u^*(\overline{u}_j^n, \overline{u}_{j+1}^n))$$

and the method becomes

$$\overline{u}_i^{n+1} = \overline{u}_i^n - \frac{\Delta t}{\Delta x}(f(u^*(\overline{u}_j^n, \overline{u}_{j+1}^n)) - f(u^*(\overline{u}_{j-1}^n, \overline{u}_j^n))).$$

For large t, of course, the solution may not remain constant at $x_{i+1/2}$ because of the effect of waves arising from neighboring Riemann problems. However, since the wave speeds are bounded by the eigenvalues of $f'(u)$ and the neighboring Riemann problems are distance Δx away, $\tilde{u}^n(x_{i+1/2}, t)$ will be constant over $[t_n, t_{n+1}]$, provided Δt is sufficiently small. We require that

$$\left|\frac{\Delta t}{\Delta x}\lambda_p(\overline{u}_i^n)\right| \leq 1$$

for all eigenvalues λ_p. This condition is a generalization of the Courant number. For more details, see [15].

3.3.1 Entropy condition

The function $\tilde{u}^n(x,t)$ for $t_n \leq t \leq t_{n+1}$, is assumed to be a weak solution of the conservation law. In situations where this weak solution is not unique, there may be several choices for $\tilde{u}^n$. Different choices may give different values of $u^*(\overline{u}_{j-1}^n, \overline{u}_j^n)$ and hence different numerical solutions. The method is conservative and consistent regardless of what choice we make, but in cases where there is a unique weak solution that satisfies some physically motivated entropy condition, it makes sense to use the entropy-satisfying weak solution for u" in each time step. We might hope that by doing this the numerical solution will satisfy a discrete version of the entropy condition. This is in fact true, as we verify below. It then follows from the section 3.2.2 about the Lax-Wendroff theorem and the discrete entropy condition that any limiting function obtained by refining the grid must then be an entropy-satisfying weak solution. If, on the other hand, we use Riemann solutions that do not satisfy the entropy condition in defining $u^*(\overline{u}^n_{j-1}, \overline{u}^n_j)$ then our numerical solution may converge to a weak solution that does not satisfy the entropy condition. Suppose that we have a convex entropy function $\eta(u)$ and entropy flux $\psi(u)$, and that every $\tilde{u}$ satisfies the entropy inequality (2.84). Then we wish to derive a discrete entropy inequality (3.7) for Godunov's method. Since $\tilde{u}^n(x,t), t_n \leq t \leq t_{n+1}$ represents the exact entropy satisfying solution, we can integrate (2.84) over the rectangle $[x_{i-1/2}, x_{i+1/2}] \times [t_n, t_{n+1}]$ to obtain

$$\begin{aligned} \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} \eta(\tilde{u}^n(x,t_{n+1}))dx \leq \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} \eta(\tilde{u}^n(x,t_n))dx- \\ \frac{1}{\Delta x}\Big[\int_{t_n}^{t_{n+1}} \psi(\tilde{u}^n(x_{i+1/2},t))dt - \int_{t_n}^{t_{n+1}} \psi(\tilde{u}^n(x_{i-1/2},t))dt\Big] \end{aligned} \tag{3.9}$$

This is almost what we need. Since $\tilde{u}^n$ is constant along three of the four sides of this rectangle, all integrals on the right hand side can be evaluated to give

$$\begin{aligned} \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} \eta(\tilde{u}^n(x,t_{n+1}))dx \leq \eta(\overline{u}^n_i)- \\ \frac{\Delta t}{\Delta x}\Big[\psi(u^*(\overline{u}^n_j, \overline{u}^n_{j+1})) - \psi(u^*(\overline{u}^n_{j-1}, \overline{u}^n_j))\Big] \end{aligned} \tag{3.10}$$

Again u^* represents the value propagating with velocity 0 in the solution of the Riemann problem. If we define the numerical entropy flux by

$$\Psi(\overline{u}^n_j, \overline{u}^n_{j+1}) = \psi(u^*(\overline{u}^n_j, \overline{u}^n_{j+1}))$$

then Ψ is consistent with ψ and the right hand side of (3.10) agrees with (2.84). Finally, since the entropy function η is convex, we can use Jensen's inequality to obtain

$$\eta(\frac{1}{\Delta x}\int_{x-1/2}^{x+1/2}\tilde{u}^n(x,t_{n+1})dx) \leq \frac{1}{\Delta x}\int_{x-1/2}^{x+1/2}\eta(\tilde{u}^n(x,t_{n+1}))dx. \qquad (3.11)$$

(13.20) Combining everything we have the desired entropy inequality

$$\eta(\overline{u}_i^{n+1}) \leq \eta(\overline{u}_i^n) - \frac{\Delta t}{\delta x}[\Psi(\overline{u}_i^n, \overline{u}_{i+1}^n) - \Psi(\overline{u}_{i-1}^n, \overline{u}_i^n)], \qquad (3.12)$$

This shows that weak solutions obtained by Godunov's method satisfy the entropy condition, provided we use entropy-satisfying Riemann solutions.

3.4 Other schemes: high-resolution Methods and ENO reconstruction

While stable, Godunov's method is only first-order accurate and suffers from excessive numerical diffusion, which leads to poor resolution of sharp gradients. For this reason, higher-order techniques that reduce numerical diffusion while maintaining stability are essential.

A significant improvement over Godunov's constant reconstruction is a piecewise linear approach, which provides second-order accuracy in smooth regions. In this case, the solution in each cell $I_j = [x_{j-1/2}, x_{j+1/2}]$ is reconstructed as

$$u_j(x) = u_j + \sigma_j(x - x_j),$$

where u_j is the cell average and σ_j the slope. To avoid oscillations near discontinuities, slope-limiters are applied to control σ_j based on local solution behavior. For example, the *Minmod* limiter is defined by

$$\text{Minmod}(a,b) = \begin{cases} a, & \text{if } |a| < |b| \text{ and } ab > 0, \\ b, & \text{if } |b| < |a| \text{ and } ab > 0, \\ 0, & \text{if } ab \leq 0, \end{cases}$$

selecting the smallest gradient when possible to prevent oscillations. The Van Leer limiter, on the other hand, is given by

$$\text{VanLeer}(a,b) = \frac{2ab}{a+b} \quad \text{for } a, b > 0,$$

resulting in smoother transitions. These limiters allow high-resolution methods to reach second-order accuracy in smooth regions, but more advanced techniques are needed for higher accuracy in complex solutions.

To address this, *Essentially Non-Oscillatory* (ENO) and *Weighted Essentially Non-Oscillatory* (WENO) methods have been developed. The ENO scheme improves accuracy by choosing in an adaptive way the smoothest stencil around each cell. For example, in a third-order ENO scheme, possible stencils around cell I_j include

$$\{u_{j-1}, u_j, u_{j+1}\}, \quad \{u_j, u_{j+1}, u_{j+2}\}, \quad \{u_{j-2}, u_{j-1}, u_j\},$$

and the stencil with the smallest divided difference is selected to minimize oscillations.

The WENO (Weighted Essentially Non-Oscillatory) method further improves efficiency by combining all stencils rather than selecting just one. In a fifth-order WENO (WENO5) scheme, the flux at an interface $x_{j+1/2}$ is approximated as a weighted sum:

$$\hat{u}_{j+1/2} = \sum_{k=0}^{2} \omega_k q_k,$$

where each q_k is calculated on a different stencil, and the weights ω_k are chosen to favor smoother stencils. These weights are defined by

$$\omega_k = \frac{\alpha_k}{\sum_{l=0}^{2} \alpha_l}, \quad \alpha_k = \frac{C_k}{(\epsilon + \beta_k)^2},$$

where ϵ prevents division by zero, C_k are constants, and β_k are smoothness indicators. These indicators ensure that stencils with large gradients are downweighted, helping to avoid oscillations.

Overall, ENO and WENO methods offer improvements in accuracy and efficiency for high-resolution solutions of hyperbolic PDEs, particularly useful in applications with strong nonlinearities and sharp gradients.

3.5 Conservative finite difference schemes

Finally, let us introduce a different kind of conservative method, a *finite difference* type. Let us consider the usual system

$$u_t + f_x = 0.$$

We write

$$\frac{\partial f}{\partial x}(u(x)) = \frac{\hat{f}(u(x+\frac{h}{2})) - \hat{f}(u(x-\frac{h}{2}))}{h}.$$

where the relation between f and $\hat{f}$ will be clarified in the following. Consider the sliding cell average operator:

$$\overline{u}(x) = \frac{1}{h}\int_{x-\frac{h}{2}}^{x+\frac{h}{2}} u(\xi)d\xi. \tag{3.13}$$

Differentiating with respect to x we get

$$\frac{\partial \overline{u}}{\partial x} = \frac{1}{h}\left(u\Big(x+\frac{h}{2}\Big) - u\Big(x-\frac{h}{2}\Big)\right).$$

Therefore the relation between f and $\hat{f}$ is the same that exist between $\overline{u}(x)$ and $u(x)$, namely the function f is the cell average of the function $\hat{f}$. This also suggests a way to compute the flux function. The technique that is used to compute pointwise values of $u(x)$ at the edge of the cell from cell averages of u can be used to compute $\hat{f}(u(x_{j+1/2}))$ from $f(u(x_j))$. This means that in finite difference method it is the flux function which is computed at x_j and then reconstructed at $x_{j+1/2}$. But the reconstruction at $x_{j+1/2}$ may be discontinuous. Which value should one use? A general answer to this question can be given if one considers flux functions that can be splitted as

$$f(u) = f^+(u) + f^-(u),$$

with the condition that

$$\frac{df^+(u)}{du} \geq 0, \quad \frac{df^-(u)}{du} \leq 0.$$

There is a close analogy between flux splitting and numerical flux functions.

In fact, if a flux f can be splitted as said with the property just said, then

$$F(a,b) = f^+(a) + f^-(b)$$

will define a monotone consistent flux. This is the case, for example, of the local Lax-Friedrichs flux.

A finite difference scheme therefore takes the following form:

$$\frac{du_j}{dt} = -\frac{1}{h}[\hat{F}_{j+1/2} - \hat{F}_{j-1/2}]$$
$$\hat{F}_{j+1/2} = \hat{f}^+(u^-_{j+1/2}) + \hat{f}^-(u^+_{j+1/2}),$$

where:

- $\hat{f}^+(u^-_{j+1/2})$ is obtained as follows:
  - compute $f^+(u_l)$ and interpret it as cell average of $\hat{f}^+$;
  - perform pointwise reconstruction of $\hat{f}^+$ in cell j and evaluate it in $x_{j+1/2}$;
- $\hat{f}^-(u^+_{j+1/2})$ is obtained as follows:
  - compute $f^-(u_l)$, interpret it as cell average of $\hat{f}^-$;
  - perform pointwise reconstruction of $\hat{f}^-$ in cell $j+1$, and evaluate it in $x_{j+1/2}$.

Chapter 4

Dispersive equations

4.1 Introduction

In this chapter, we introduce the *dispersive equations*, a type of PDE which solution are characterized by a so called dispersive behaviour, where the "waves" travel a different speeds and this leads to the spreading or distortion of a wave packet as it propagates over time. The way we can tell mathematically if a PDE is dispersive is by introducing the dispersion relation, that allows us also to describe some properties of the equations, such as the well-poseness, and the solutions, such as the asymptotic behaviour. We will focus mainly on nonlinear dispersive equation, describing the probably most important example of this kind, the Korteweg-De Vries equation. Through the study of this equation, we will introduce other theoretical aspects of the study of nonlinear dispersive equations, such as the search of special solutions, that in this work will be based on the quest for travelling waves and on the development of methods for the resolution of the ODEs that arise during this quest, or the search of conserved quantities, that are crucial on both mathematical aspect, since they can used as indicator of stability, validation of numerical schemes or for control in norms of solutions, and physical, where the conservation of quantities like the energy or the momentum can help on the description on the physical model. Most of the chapter is based on [19] and on the notes named "Nonlinear Waves" of professor Bernard Deconinck of University of Washington for his course "Coherent Structures, Pattern Formation and Solitons", findable on the website of the course.

4.2 Dispersion relation

4.2.1 Linear case

In order to understand nonlinear problems, that are the ones we are focusing on, we first have to work on the linear case, so that we will be able to describe the notions and the ideas we need. Let us start with the scalar linear first-order evolution equation with constant coefficients, i.e. equations of the form

$$u_t = F(u, u_x, u_{xx}, \ldots), \quad F(u, u_x, u_{xx}, \ldots) = a_0 u + a_1 u_x + a_2 u_{xx} + \ldots. \tag{4.1}$$

We assume vanishing boundary conditions at infinity, i.e. $u \to 0$ as $|x| \to \pm\infty$. Let us introduce the *dispersion relation.* For this purpose, we consider basic modes of the form

$$u = e^{ikx-\omega t}.$$

Substituting it in (4.1) we obtain

$$i\omega u = F(u, iku, -k^2 u, \ldots)$$

and using the linearity and homogeneity of F the quest of solutions other than $u = 0$ results in the resolution of the following equation for ω:

$$\omega = iF(1, ik, -k^2, \ldots) = \omega(k). \tag{4.2}$$

This is the *(linear) dispersion relation.* Thus, the modes take the form

$$u = e^{i(kx-\omega(k)t)},$$

and the most general solution of the equation, subject to the initial condition $u(x, 0) = u_0(x)$ with $u_0 \to 0$ as $|x| \to \pm\infty$, can be expressed as a superposition of these modes:

$$u(x,t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} a(k) e^{i(kx-\omega(k)t)}\, dk, \tag{4.3}$$

where $a(k)$ is determined by the initial condition $u_0(x)$ as its Fourier transform, i.e.,

$$a(k) = \mathcal{F}[u_0(x)](k).$$

Suppose now that u in (4.1) is a vector of dimension $N > 1$. Then, the ansatz is now given by

$$u(x,t) = A(k)e^{i(kx-\omega t)},$$

where $A(k)$ is a vector of dimension N independent from x, t. We obtain

$$-i\omega u = F(1, ik, -k^2, \ldots)u \Longrightarrow (F(1, ik, -k^2, \ldots) + i\omega I)u = 0, \tag{4.4}$$

where $F(1, ik, -k^2, \ldots)$ is a matrix of size $N \times N$.

The *dispersion relation* in the vectorial case is given by

$$Det(F(1, ik, -k^2, \ldots) + i\omega I) = 0,$$

and the roots $\omega_j(k)$ of this polynomial equation in ω are the eigenvalues of the matrix $iF(1, ik, -k^2, \ldots)$. As in the linear case, we can obtain the most general solution as the superposition

$$u(x,y) = \frac{1}{2\pi}\sum_{j=1}^{N}\int_{-\infty}^{+\infty} a_j(k)A_j(k)e^{i(kx-\omega_j(k)t)}dk, \tag{4.5}$$

where $A_j(k)$ are the eigenvectors corresponding to $\omega_j(k)$ and $a_j(k)$ are obtainable from the initial conditions through the Fourier Transform as before.

We can generalize these cases, deleting the first-order equation assumption or working with non evolutionary equations: in the first case, since equations with higher-order time derivatives can always be reduced to first-order in time system, they can be studied as in the vectorial case just studied; in the second case, we proceed as in the evolutionary case, obtaining in this case a dispersion relation that is not a polynomial expression in general.
A lot of important informations about the equation and its solutions can be deduced from the dispersion relation. We have:

- If $\text{Im}(\omega(k)) > 0$ for some $k \in \mathbb{R}$, the corresponding mode grows exponentially over time, leading to a strong sensitivity to the initial data and making the problem unstable. Moreover, if $\text{Im}(\omega(k)) \to \infty$ for any value of k, then the problem is ill-posed.

- $\text{Imm}(\omega(k)) < 0$ for all k, then the problem is dissipative and asymptotically stable;

- From Parseval's relation it is possible to say that

$$\int_{-\infty}^{\infty} |u|^2 dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\mathcal{F}[u](k)|^2 dk = \frac{1}{2\pi} \int_{-\infty}^{\infty} |a(k)|^2 e^{2\text{Imm}(\omega(k))t} dk,$$

so

$$\frac{d}{dt} \int_{-\infty}^{\infty} |u|^2 dx = 0 \Longleftrightarrow \text{Imm}(\omega(k)) = 0.$$

This means that for equations with real-valued dispersion relation the quantity $\int_{-\infty}^{\infty} |u|^2 dx$ is a *conserved quantity.* Let us assume that we are not working on the one-dimensional transport equation $u_t = cu_x$. It is possible for this solution to have an asymptotic state, as the finite amount of initial energy can spread over an infinite region (assumed that the boundary condition is not periodic). This because, since we are not dealing with transport equation, $\omega(k) \neq ck$, so the problem is not the mere translation of initial data.

Due all this consideration, we can finally define what is a dispersive equation.

Definition 4.1. We say that an equation is *dispersive* if the *phase velocity* $c_p = \omega(k)/k$ is not constant. Equivalently, the equation is dispersive if

$$\frac{\partial^2 \omega}{\partial k^2} \neq 0$$

holds.

4.2.2 Asymptotic behaviour of dispersive equations

Let us examine the asymptotic behaviour of the solution (4.3) of a dispersive equation. To do so, we use the so called method of stationary phase, developed by Kelvin. Let us consider the expression

$$u(x,t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} a(k) e^{i(\frac{kx}{t} - \omega(k))t} dk = \frac{1}{2\pi} \int_{-\infty}^{\infty} a(k) e^{i\phi(k)t} dk, \qquad (4.6)$$

with $\phi(k) = kx/t - \omega(k)$ along rays of constant x/t. Stationary points occur when $\phi'(k) = 0$, from which $x = \omega'(k)t := c_g(k)t$, where we defined the *group* *velocity* $c_g(k) = \omega'(k)$. The method is based on the assumption $\phi'(k) = 0$: this condition allows us to avoid the problems given by the fact that modes corresponding to nearby wave numbers k and $k + \delta k$ have very different exponents, leading to destructive interference. This assumption implies

$$c_g(k) = \frac{x}{t},$$

so the wave numbers (just one wave number k_0 if we assume the uniqueness of solutions) that dominate are the ones such that $c_g(k) = \frac{x}{t}$. Expanding $\phi(k)$ around k_0 we obtain

$$\phi(k) = \phi(k_0) + \frac{1}{2}(k - k_0)^2 \phi''(k_0) + O(|k - k_0|^3), \tag{4.7}$$

that substituted in (4.6) and by using the fact that the integrand is even gives us

$$\frac{1}{2\pi} \int_{k_0-\delta}^{ko+\delta} a(k) e^{i\phi(k)t} dk \approx \frac{1}{\pi} a(k_0) e^{i\phi(k_0)t} \int_{k_0}^{ko+\delta} e^{i\frac{1}{2}(k-k_0)^2 \phi''(k_0)t} dk \approx$$
$$\approx \frac{1}{\pi} a(k_0) e^{i\phi(k_0)t} \sqrt{\frac{2}{|\phi''(k_0)|t}} \frac{\sqrt{\pi}}{2} e^{i\pi sign(\phi''(k_0))/4}$$

For further details on the calculations see [19].

The final expression tells us that $|u(x,t)| \approx t^{-\frac{1}{2}}$, so the solution spreads over a larger region as $t \to \infty$. This is a proof of the dispersive character of the solution: indeed, since the energy is conserved, is the amplitude the quantity that has to decrease, spreading over the $x-$axis.

4.3 Nonlinear dispersive equations

We focus now on nonlinear equations, that are the ones of main interest for the goals of this works. We will see that the dispersion relation of an equation has consequences beyond allowing us to solve the corresponding linear equation. Indeed, when studying a nonlinear PDE, we can linearize the equation around a constant state, dropping all nonlinear terms. The dispersion relation of the resulting linear equation provides good informations about the dynamics of small solutions of the nonlinear equation. Let us introduce some important examples of nonlinear dispersive equations, where we will see what "dispersive" means in this framework.

4.3.1 The Korteweg-De Vries equation

The *Korteweg-De Vries* (KdV) equation is one of the most well-known PDEs in mathematical physics. It was introduced in 1895 by Diederik Korteweg and Gustav de Vries in their seminal work [10] on shallow water waves. The equation was derived to model the propagation of long waves in a shallow channel, such as those observed in canals or rivers. Their work was inspired by the pioneering observations of John Scott Russell, a Scottish engineer who, in 1834, first documented a phenomenon he called the "wave of translation". Russell observed that a single, localized wave could travel over a long distance without changing its shape, sparking interest among mathematicians and physicists.

The KdV equation models the interplay between two competing effects: dispersion, which causes waves to spread out, and nonlinearity, which tends to steepen wave profiles. The balance between these effects allows for the emergence of stable wave structures called *solitons*. The KdV equation, which derivation can be found in almost every textbook that treats this topic (see [19] for example), presents itself in its canonical form as

$$u_t + 6uu_x + u_{xxx} = 0, \tag{4.8}$$

where $u(x,t)$ represents the wave profile, t is time, and x is the spatial coordinate.

Let us see what does it mean for the KdV equation to be dispersive: linearizing the equation and substituting a plane wave solution of the form $u(x,t) = e^{i(kx-\omega t)}$, as done in the linear case, yields the dispersion relation:

$$\omega(k) = -k^3,$$

that allows us to say that the equation is dispersive in this new sense since $c_p(k) = \omega/k = -k^2$ is not constant. Moreover, the dependence of $c_p(k)$ implies that waves with different wavenumbers propagate at different speeds. As briefly said, one of the most remarkable properties of the KdV equation is its ability to admit exact solutions in the form of *solitons*, which are localized wave packets that maintain their shape during propagation and interaction.

A single soliton solution is given by:

$$u(x,t) = A\,\mathrm{sech}^2\left(\sqrt{\frac{A}{2}}(x-ct)\right), \tag{4.9}$$

where A is the amplitude and $c = A$ is the velocity of the soliton. Lately, it is shown how this solution has been constructed. These solutions arise from the balance between nonlinearity and dispersion.

We have other important properties of the KdV that make it an equation whose study is mandatory, for example the fact that KdV is integrable and can be solved exactly using the so called *inverse scattering transform (IST)* or the fact that KdV has an infinite number of *conserved quantities* (we will talk about them in a following section).

4.3.2 The Boussinesq Equations

The *Boussinesq equations* are a set of partial differential equations that describe the propagation of long waves in shallow water and other dispersive media. Derived in [4] by Joseph Valentin Boussinesq in 1872, they were a significant advancement in the study of fluid mechanics and nonlinear wave propagation. Boussinesq aimed to improve upon existing linear theories by accounting for both nonlinearity and dispersion, making his equations one of the first examples of a nonlinear dispersive wave model. The classical Boussinesq system is given by:

$$\eta_t + u_x + (u\eta)_x = 0, \tag{4.10}$$

$$u_t + \eta_x + uu_x - \frac{1}{3}u_{xxt} = 0. \tag{4.11}$$

Theorem 3.3 of [3] ensure the existence of solutions for this system.

Under certain conditions, the Boussinesq equations admit *soliton* solutions, which are localized, non-dispersive wave packets. For example, a single soliton solution can take the form:

$$u(x,t) = A\,\mathrm{sech}^2\left(\sqrt{\frac{A}{2}}(x-ct)\right), \tag{4.12}$$

similar to the solitons of the KdV equation. These solutions arise when the nonlinear steepening balances the dispersive spreading of waves.

4.4 Travelling waves

The equations of interest are nonlinear, and the problem we seek to solve is typically formulated as an initial or initial-boundary value problem on a given domain. Due to the complexity of nonlinear wave equations, finding explicit solutions to these problems is generally not feasible. As a result, attention is often directed toward identifying specific special solutions that can be determined.

A key observation is that the most common approach for obtaining special solutions to nonlinear wave equations relies on an educated guess. This method involves proposing an ansatz for the solution, incorporating certain free functions or parameters. By substituting this ansatz into the equation, a system of equations is obtained that constrains these functions or parameters. Unlike in the linear case, however, special solutions of nonlinear partial differential equations do not act as fundamental solutions, since superposition does not lead to more general solutions.

Despite this limitation, exact solutions remain essential. They provide a deeper understanding of the partial differential equation itself and serve as a reference to verify the accuracy of numerical methods when solving nonlinear equations computationally. Furthermore, although superposition no longer applies, special solutions can still serve as fundamental components in nonlinear theory. In many cases, the long-term behaviour of solutions to a PDE can be described asymptotically by different exact solutions, making them a powerful tool in the study of nonlinear wave phenomena.

In this section, we study a particular case of special solution: the so called *travelling wave*. Those are defined as solutions that are "fixed" with respect some frame of references, that is moving with velocity v with respect to the original frame of reference. Thus, we look for solution of the kind

$$u(x,t) = U(x - vt) = U(\xi), \tag{4.13}$$

where $\xi = x - vt$. As said, we are looking for stationary solution in a certain moving frame of reference.

Given a nonlinear partial differential equation

$$u_t = F(u, u_x, u_{xx}, \ldots), \tag{4.14}$$

substituting our ansatz into the PDEs gives

$$-vU' = F(U, U', U'', \ldots), \tag{4.15}$$

which is an ODE for U. In this section, we examine this ODE, from different points of view, introducing techniques, connected to each other, inspired by other fields of study like physics. We will introduce them through their direct construction over the KdV equation, the main case study when it comes to dispersive equations.

4.4.1 The energy integral approach

Let us consider the KdV equations

$$u_t + 6uu_x + u_{xxx} = 0. \tag{4.16}$$

Stationary solutions satisfy the ODE

$$-vU' + 6UU' + U''' = 0. \tag{4.17}$$

The equation may be integrated once to obtain

$$-vU + 3U^2 + U'' + \alpha = 0, \tag{4.18}$$

with α integration constant. Multiplying by U' and integrating the obtained equation we have

$$\begin{aligned} & -vUU' + 3U^2U' + U''U' + \alpha U' = 0 \Longrightarrow \\ & \Longrightarrow -\frac{v}{2}U^2 + U^3 + \frac{1}{2}(U')^2 + \alpha U - \beta = 0, \end{aligned} \tag{4.19}$$

where β is a second integration constant. The equation is rewritten as

$$\frac{1}{2}(U')^2 + V(U, v, \alpha) = \beta, \quad V(U, v, \alpha) = -\frac{v}{2}U^2 + U^3 + \alpha U. \tag{4.20}$$

The equation above represents the conservation of energy for the motion of a particle with unit mass subjected to a conservative force, whose potential is given by $V(U, v, \alpha)$. The first term in this equation corresponds to the kinetic energy, while the second term represents the potential energy. The right-hand side expresses the constant total energy.

This analogy allows us to draw several important conclusions. Notably, we observe that $V \to \infty$ as $U \to \infty$ and $V \to -\infty$ as $U \to -\infty$. It is evident that for every value of β, a real-valued solution exists. However, the key challenge lies in determining whether bounded solutions exist and under what conditions. If V is monotone no value of β permits the existence of bounded solutions. Conversely, if V is non-monotone and possesses a local minimum, bounded solutions can exist. In this case, they occur at the energy level β_s, corresponding to trajectories passing through the local maximum adjacent to the local minimum.

Equilibrium solutions are characterized by the condition $V'(U, v, \alpha) = 0$. Consequently, such solutions are found at the local extrema of the potential function. Those located at local minima are stable within the class of stationary solutions, whereas those at local maxima are inherently unstable.

Anticipating the phase plane analysis, we deduce that since energy remains conserved, the only possible non-degenerate equilibrium points are either centers, which are stable, or saddles, which are unstable.

4.4.2 The phase plane analysis

In the previous paragraph, we describe a way to tell if solutions exist, if they are bounded and which conditions they need to satisfy. Now, assuming everything we said before, we go further, trying to describe our solution through a phase plane analysis, which allows, for second-order autonomous systems, to visualize and study graphically the solutions without solving the equation. This approach allows us to provide motion trajectories corresponding to various initial conditions and examine their qualitative features, obtaining information regarding the stability of the equilibrium points. Let us now apply this method to our equation. Starting from (4.18), we can say that this equation can be written as

$$U'' = -\frac{\partial V}{\partial U}(U, v, \alpha), \tag{4.21}$$

that, in system form with $u_1 = U, u_2 = U'$, becomes

$$\begin{cases} u_1' = u_2 \\ u_2' = -\frac{\partial V}{\partial u_1}(u_1, v, \alpha) \end{cases} \tag{4.22}$$

From our previous results, we have that

$$E(u_1, u_2, v, \alpha) = \frac{1}{2}u_2^2 + V(u_1, u_2, v, \alpha) = \beta \tag{4.23}$$

is a conserved quantity for the first-order dynamical system. Therefore, in the (u_1, u_2)-plane, the system's trajectories are restricted to the level curves of $E(u_1, u_2, v, \alpha)$. It is important to note that all equilibrium solutions lie along the horizontal axis. Additionally, these level curves exhibit symmetry with respect to the vertical axis and, apart from the orientation of the trajectories, they are also symmetric with respect to the horizontal axis.
In the KdV equation we are studying, we have

$$\frac{\partial V}{\partial U} = 0 \Longrightarrow 3U^2 - vU + \alpha = 0$$

so the number of equilibrium solutions is determined by the discriminant of the above equation, given by

$$\Delta = v^2 - 12\alpha.$$

We have three cases to study:

- $v^2 < 12\alpha$: in this case, the equation does not admit real solutions so V is a monotone increasing function of U. This means, as said, that solutions exist for every β but they are always not bounded. In Figure 4.1 this case is shown.

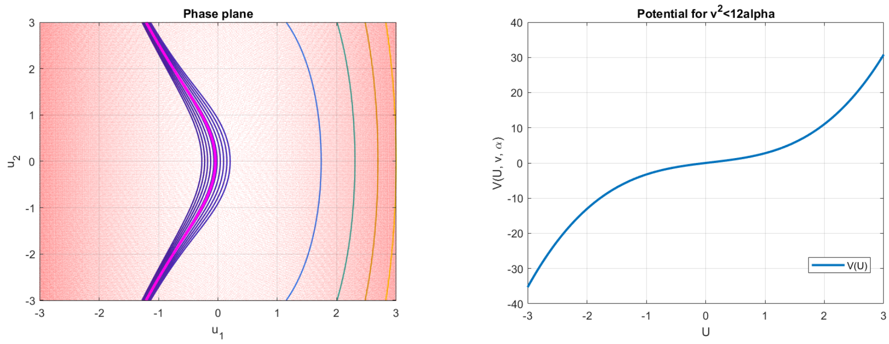


Figure 4.1: Trajectories in the phase space and potential for $v < 12\alpha$.

- $v^2 = 12\alpha$: in this case, V is still monotone increasing but there is an inflection point. So even in this case the solutions exist for every

β and they are always not bounded, and the equilibrium solution is degenerate. In Figure 4.2 this case is shown.

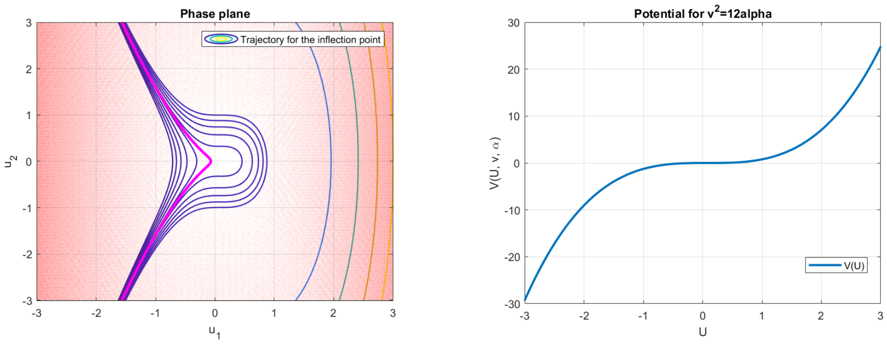


Figure 4.2: Trajectories in the phase space and potential for $v = 12\alpha$.

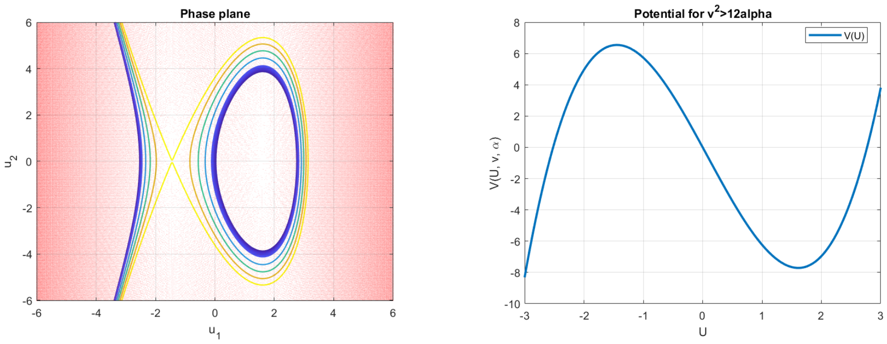


Figure 4.3: Trajectories in the phase space and potential for $v > 12\alpha$.

- $v^2 > 12\alpha$: in this case, V is no longer monotone and admits a local minimum at U_b and a local maximum at U_a. In this case, there are bounded solutions, the ones with β such that $V(U_b, v, \alpha) \leq \beta \leq V(U_a, v, \alpha)$. Moreover, there are two equilibrium solutions, one stable at U_b, that gives a saddle point in the phase plane, and one unstable at U_a, that is the center in the phase place, sorrounded by periodic solutions. Since the solution passing through U_a in the phase plane representation is the separatrix between periodic solutions and unbounded solution, it is there that we should look after a soliton solution. In Figure 4.3 this case is shown.

4.4.3 Soliton solutions by explicit integration

In order to obtain the explicit functional form of the soliton solution, we recall the energy equation (4.20), which gives

$$U' = \pm\sqrt{2(\beta - V(U,v,\alpha))}.$$

Separating variables gives an implicit solution

$$\begin{aligned} \int_{U_0}^{U} \frac{dU}{\sqrt{2(\beta - V(U,v,\alpha))}} &= \pm\int_0^z dz = \pm z. \Longrightarrow \\ \Longrightarrow \pm z &= \int_{U_0}^{U} \frac{dU}{\sqrt{2(\beta + \frac{v}{2}U^2 - U^3 - \alpha U}}. \end{aligned} \tag{4.24}$$

Since we look for solutions that vanish as $x \to \pm\infty$, we can impose $\alpha = 0, \beta = 0$, so the implicit solution simplifies as

$$z = \int_{U_0}^{U} \frac{dU}{\sqrt{2(\frac{v}{2}U^2 - U^3)}}. \tag{4.25}$$

As $z \to \pm\infty, U \to 0$, so for existence of the square root as $U \to 0$ it is required that v has to be positive, i.e. $v = a^2$, so that

$$z = \int_{U_0}^{U} \frac{dU}{U\sqrt{(a^2 - 2U)}}. \tag{4.26}$$

The integral can be calculated doing the following substitution:

$$\begin{aligned} w = \sqrt{(a^2 - 2U)} &\Longrightarrow w^2 = a^2 - 2U \Longrightarrow \\ \Longrightarrow 2wdw &= -2dU \Longrightarrow dU = -wdw, \end{aligned} \tag{4.27}$$

so the integral becomes

$$\begin{aligned} \pm z = \int -\frac{2wdw}{Uw} &= \int -\frac{2dw}{U} = \int -\frac{dw}{a^2 - w^2} = \int \frac{dw}{w^2 - a^2} = \\ &= \frac{1}{a}\int \Big(\frac{dw}{x-a} - \frac{dw}{x+a}\Big) = \frac{1}{a}\ln\Big(C\frac{w-2a}{w+2a}\Big), \end{aligned}$$

where C is a constant of integration. Solving for w, we obtain

$$w = a\frac{1 + \frac{1}{C}e^{\pm az}}{1 - \frac{1}{C}e^{\pm az}}.$$

In order to have bounded solution, we need $C < 0$. Set $C = -e^{\pm a\alpha}$, then

$$w = a\frac{1 + e^{\pm(az - a\alpha)}}{1 - e^{\pm(az - a\alpha)}} = a\frac{e^{\mp(az-a\alpha)/2} + e^{\pm(az-a\alpha)/2}}{e^{\mp(az-a\alpha)/2} - e^{\pm(az-a\alpha)/2}} = \mp a \tanh\frac{a}{2}(z - \alpha). \tag{4.28}$$

Solving for U, we obtain

$$\begin{aligned} &w^2 = a^2 \tanh^2\frac{a}{2}(z-\alpha) \Longrightarrow a^2 - 2U = a^2\tanh^2\frac{a}{2}(z-\alpha) \Longrightarrow \\ &\Longrightarrow U = \frac{1}{2}\Big(a^2 - a^2\tanh^2\frac{a}{2}(z-\alpha)\Big) = \frac{1}{2}a^2\text{sech}^2\Big(\frac{a}{2}(z-\alpha)\Big), \end{aligned} \tag{4.29}$$

that in original variables is

$$u(x,t) = \frac{1}{2}a^2\text{sech}^2\Big(\frac{a}{2}(x - vt - \alpha)\Big) = \frac{1}{2}a^2\text{sech}^2\Big(\frac{a}{2}(x - a^2t - \alpha)\Big). \tag{4.30}$$

Thsi is the soliton solution for the KdV equation, that has the same form of the one announced in (4.9) with $A = a^2/2$ and $\alpha = 0$.

4.5 Conserved Quantities

From a dispersive equation sometimes it is possible to derive some conservation laws. When it happens, we can say that some quantities are conserved, i.e. "they do not change in time". Let us be more rigorous.

Given an equation of the kind

$$u_t = N(u, u_x, \ldots),$$

a quantity $F(u, u_x, \ldots) = \int f(u, u_x, \ldots)dx$ is *conserved* if

$$\frac{d}{dt}F = 0,$$

which occurs when

$$\frac{\partial f}{\partial t}(u, u_x, \ldots) + \frac{\partial g}{\partial x}(u, u_x, \ldots) = 0$$

for some function g of u and its x-derivatives. On the other hand, if an equation of the form

$$\frac{\partial f}{\partial t} + \frac{\partial g}{\partial x} = 0$$

holds for some f, g function of u solution of the equation and its derivatives, then $F = \int f dx$ is a conserved quantity. The role of conserved quantities in the study of PDEs is crucial and can have applications in several fields, such as:

- **Physics**: Since a lot of systems come from trying to describe in a mathematical framework some physical phenomena, the conserved quantities can be related with the actual quantities involved in the physical model, and can be used as both validation of the model(for example, if in a physical phenomenon we know that a quantity is conserved, we want the model to contain this information) and mean for the discovery of other conserved quantities that cannot be found just through mere observation.
- **Numerical Analysis**: If a conserved quantity for a PDE or a system of PDEs is known, numerical methods applied on the model in study should also reveal the conservation, on a numerical level, of this quantity. This means that numerical quantities can be used for the validation on numerical schemes, which reliability is strictly connected on the representation of the conservation of the quantities, and also for the direct construction of methods, the so called *conservative numerical methods*, built starting from the conservation law.
- **Mathematical comprehension**: In the study of a PDE or a system of PDEs, the discovery of of conserved quantities can allow us to understand better the system we are studying. For example, In Hamiltonian mechanics conserved quantities like energy identify invariant surfaces in phase space, which can be used to study the system's global behaviour, or in integrable systems conserved quantities allow the equations of motion to be solved exactly.

We ask ourselves how can we find non-trivial conserved quantities. The some of most used methods are:

1. **Direct Integration:** It works by multiply the equation by a suitable test function (e.g., u, u^2) and integrate over the domain. Using integration by parts or the initial equation, one can try to simplify terms and check for cancellations at the boundaries. For example, for the

KdV equation

$$u_t + 6uu_x + u_{xxx} = 0,$$

multiplying by u and integrating, it gives the conservation of mass:

$$\frac{d}{dt}\int u^2\,dx = 0.$$

The same method can be used also for systems of dispersive equations, for example multiplying each equations by a test function, adding the equation and then trying to use integration by parts and the initial equations to simplify and find a conservation law.

2. **Lie Group Analysis:** In this method, one can analyze the equation for translational or rotational invariance, using these symmetries to construct conserved quantities. For example, for the nonlinear Schrödinger equation:

$$i\psi_t + \psi_{xx} + |\psi|^2\psi = 0,$$

the invariance under $\psi \to e^{i\theta}\psi$ implies conservation of mass:

$$M = \int |\psi|^2\,dx.$$

See [11] for further details.

3. **Hamiltonian Structure:** It works when one can write the equation in hamiltonian form. Writing in this form we have

$$u_t = J\frac{\delta H}{\delta u},$$

where J is a skew-symmetric operator and H is the Hamiltonian, and in this case the functional $H[u]$ represents a conserved quantity. For example, for the KdV equation, the Hamiltonian is:

$$H = \int \left(\frac{1}{2}u_x^2 - u^3\right) dx.$$

See Chapter 6 of the notes "Nonlinear waves" by Deconinck introduced at the beginning of the section for further details.

4. **Scaling Symmetries:** Many dispersive equations admit scaling symmetries of the form:

$$x \to \lambda^{\alpha} x, \quad t \to \lambda^{\beta} t, \quad u \to \lambda^{\gamma} u.$$

   Using these symmetries, one can deduce conserved quantities associated with scale invariance. For example, for the KdV equation:

$$u_t + 6uu_x + u_{xxx} = 0,$$

   the scaling symmetry $x \to \lambda x, t \to \lambda^3 t, u \to \lambda^{-2} u$ leads to conserved quantities involving higher-order moments, such as:

$$\int x u^2 \, dx.$$

   See Chapter 7 of the notes "Nonlinear waves" by Deconinck introduced at the beginning of the section for more details.

Unfortunately, not all the equations are the KdV equation: while this has infinite number of conserved quantities(for more details on this, see [1] or the notes cited in the introduction of the chapter), for the majority of the equations or the systems one can study is not easy to find even one conserved quantity, therefore being able to determine a conserved quantity, even if only as a stylistic exercise, still has its relevance.

Chapter 5

Perturbation theory and homogenization techniques

5.1 Introduction

The perturbations method are carried out with respect to a small parameter ε, which may be a non-dimensionalized amplitude of a typical perturbation. The coefficients in these expansions are obtained as solutions of a sequence of linear problems. The lowest-order terms are governed by problems that typically result from a linearization of the original problem and are known. The higher-order quantities are produced as solutions of linear inhomogeneous differential equations where the inhomogeneities involve only the previously determined lower-order quantities. One has a regular perturbation problem if a straightforward perturbation expansion is uniformly valid. However, the perturbation expansions are usually not uniformly valid and various techniques have been developed to render them uniformly valid. We will firstly introduced the simplest perturbation method class, the so called *regular perturbation methods*, in order to understand the philosophy of this theory and to justify the construction of other methods. Then we will briefly introduce some other class of methods, which have been studied in deep in several work such as [20],[24], whose this chapter is based on. Finally, we will focus on the technique that we will use later in our studies, the so called *multiple-scale method.*

5.2 Regular perturbation methods

Given a function $u(x, \varepsilon)$, one can always, under appropriate conditions, see it as a power series

$$u(x, \varepsilon) = \sum_{n=0}^{\infty} \varepsilon^n u_n(x).$$

In regular perturbation methods, what one does is substituting this power series into the differential equation and the boundary/initial conditions, expanding all the quantities in terms of ε, collecting the terms multiplied by the same power of ε and equating them to zero. In this way, solving sequentially this hierarchy of boundary/initial value problem, we can reconstruct our u. This method is not so reliable and the hopes of be able to construct a solution locally depends on the invertibility of the operators involved in every single equation. Let us see an example.

Example 5.1. *Consider the Van der Pol's equation*

$$\frac{du^2}{dt^2} + u = \varepsilon(1 - u^2)\frac{du}{dt}$$

for small ε. If $\varepsilon = 0$ this equation reduces to

$$\frac{d^2u}{dt^2} + u = 0,$$

which solution is

$$u = acos(t + \phi), \quad a, \phi \in \mathbb{R}\,.$$

To determine an improved approximation to the solution of our equation, we seek a perturbation expansion of the form

$$u(x, \varepsilon) = \sum_{n=0}^{\infty} \varepsilon^n u_n(x).$$

Substituting the expansion into the equation we have

$$\begin{aligned} \frac{d^2u_0}{dt^2} + u_0 + \varepsilon(\frac{d^2u_1}{dt^2} + u_1) + \varepsilon^2(\frac{d^2u_2}{dt^2} + u_2) + \ldots = \\ \varepsilon[1 - (u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \ldots)^2]\Big[\frac{du_0}{dt} + \varepsilon\frac{du_1}{dt} + \varepsilon^2\frac{du_2}{dt} + \ldots\Big], \end{aligned} \tag{5.1}$$

and expanding for small ε we obtain

$$\begin{aligned}\frac{d^2u_0}{dt^2} + u_0 + \varepsilon(\frac{d^2u_1}{dt^2} + u_1) + \varepsilon^2(\frac{d^2u_2}{dt^2} + u_2) + \ldots = \\ = \varepsilon(1-u_0)^2\frac{du_0}{dt} + \varepsilon^2\Big[(1-u_0)^2\frac{du_1}{dt} - 2u_0u_1\frac{du_0}{dt}\Big] + \ldots .\end{aligned} \tag{5.2}$$

From the independence of u_n from ε, we need to equate the coefficients of like powers of ε on both sides, obtaining

- ε^0*:* $\frac{d^2u_0}{dt^2} + u_0 = 0$ *;*
- ε^1*:* $\frac{d^2u_1}{dt^2} + u_1 = (1-u_0)^2\frac{du_0}{dt}$*;*
- ε^2*:* $\frac{d^2u_2}{dt^2} + u_2 = \Big[(1-u_0)^2\frac{du_1}{dt} - 2u_0u_1\frac{du_0}{dt}\Big]$*;*

and so on. Solving each equation from the ε^0 one and substituting the solution in the following one we can obtain in the end the solution of each equation at each step, being able to construct each coefficient of the expansion.

5.3 Alternative Perturbation Techniques

As sad before, in the majority of the problems involving a small parameter ε, alternative perturbation techniques are necessary to obtain accurate approximate solutions, especially when regular perturbation fails. The following ones are some of the most popular methods, which in this work will not be studied in deep.

- **Method of Strained Coordinates/Parameters**: This method is particularly useful when the presence of ε causes phase shifts or parameter adjustments in the solution, leading to the emergence of secular terms that grow unboundedly. To avoid these terms, a *strained* or modified variable is introduced, which compensates for the cumulative effects of the small parameter. For example, if t is the time variable, a strained time variable τ is defined as:

  $$\tau = t + \varepsilon\tau_1 + \varepsilon^2\tau_2 + \ldots,$$

  where the terms $\tau_1, \tau_2, \ldots$ are chosen to prevent the growth of unwanted terms. This method is particularly useful in oscillatory systems, allowing for the correction of cumulative phase deviations and resulting in a more accurate solution over long periods.

- **Method of Averaging**: The Method of Averaging is effective in systems with rapid oscillations combined with slower dynamics. For problems where a function exhibits periodic oscillatory behavior in time, averaging out these oscillations provides a simplified system that captures the longer-term behavior. For a dynamical system such as

  $$\frac{dx}{dt} = \varepsilon f(x,t),$$

  where $f(x,t)$ is periodic in t, one can compute an averaged version $\bar{f}(x)$ of $f(x,t)$:

  $$\bar{f}(x) = \frac{1}{T}\int_0^T f(x,t)\,dt,$$

  leading to the approximate equation

  $$\frac{dx}{dt} = \varepsilon \bar{f}(x).$$

  This method is widely applied in fields like celestial mechanics and in oscillatory systems, since it simplifies the analysis of slow dynamics while still providing an accurate approximation of the system's overall behavior.

- **Method of Matched Asymptotic Expansions**: When a problem exhibits regions where the solution changes rapidly, such as near boundaries or transitions, the Method of Matched Asymptotic Expansions is used to construct a globally valid solution. This technique divides the solution into two parts: an *outer solution*, valid away from the boundary layer, and an *inner solution* that captures the behaviour in the boundary layer. For a function $u(x)$, the outer expansion might take the form:

  $$u_{\text{outer}}(x) = u_0(x) + \varepsilon u_1(x) + \dots,$$

  while for the inner solution, a rescaled variable $\xi = x/\varepsilon$ is used, yielding an expansion:

  $$u_{\text{inner}}(\xi) = v_0(\xi) + \varepsilon v_1(\xi) + \dots.$$

> The two solutions are then *matched* in an overlapping region where both are valid, ensuring a smooth transition. This method is especially useful in fluid dynamics and thermal conduction problems, where boundary layers are common.

These perturbation methods provide effective alternatives when regular perturbation does not yield satisfactory solutions. Each technique is suited to specific problem configurations and enables uniformly valid approximations, even in cases of complex behaviour such as rapid oscillations or steep gradients.

5.4 Multiple-scales method

Some natural processes have more than one characteristic length or time scales associated with them, for example the turbulent flow consists of various length scales of the turbulent eddies along with the length scale of the objects over which the fluid flows. The failure to recognize a dependence on more than one space/time scale is a common source of nonuniformity in perturbation expansions. The method of multiple scales (also called the *multiple-scale analysis*) comprises techniques used to construct uniformly valid approximations to the solutions of perturbation problems in which the solutions depend simultaneously on widely different scales. This is done by introducing fast-scale and slow-scale variables for an independent variable, and subsequently treating these variables, fast and slow, as if they are independent.

Essentially, given a function $u(x)$, in this method what one does is to modify the dependence of the function u from x, introducing new variables of the kind $x_k = \varepsilon^k x$, where ε is a small parameter, and then to develop the regular perturbation $u = u_0 + \varepsilon u_1 + \ldots$.

Let us see the classical example for what concern the illustration of this method, the so called "Linear Damped Oscillator".

Example 5.2. *Consider the differential equation for the linear damped mass-spring system with no external forces. The equation for $y(\tau)$ is*

$$my'' + cy' + ky = 0, \quad y(0) = y_i, \quad y'(0) = 0$$

Assuming $c << m, k$ and adimensionalizing the equation by introducing the dimensionless variables

$$x = \frac{y}{y_i}, \quad t = \frac{\tau}{\sqrt{m/k}},$$

we obtain the equation

$$x'' + 2\varepsilon x' + x = 0, \quad x(0) = 1, \quad x'(0) = 0, \tag{5.3}$$

where $\varepsilon = \frac{c}{2\sqrt{mk}} << 1$. The exact solution is given by

$$x(t) = e^{-\varepsilon t}\Big(cos(\sqrt{1-\varepsilon^2}t + \frac{\varepsilon}{1-\varepsilon^2}sin(\sqrt{1-\varepsilon^2}t\Big),$$

so in the undamped case ($\varepsilon = 0$) we have $x(t) = cos(t)$. However, presence of damping shows that both amplitude and phase change with time. In fact, the amplitude drifts on the time scale ε^{-1}, while the phase drifts on the longer time scale ε^{-2}. This is the good enviroment for the application of the method. We note from analytical solution (3) that the functional dependence of x on t and ε is not disjoint because x depends on the combination of εt as well as on the individual t and ε. Thus in place of $x = x(t; \varepsilon)$, we write $x = \tilde{x}(t, \varepsilon t; \varepsilon)$. Applying the method as in [6], where the case with only two time scales involved is considered, we seek a solution in the form

$$\tilde{x}(t_0, t_1, \varepsilon) = x_0(t_0, t_1) + \varepsilon x_1(t_0, t_1) + \dots,$$
$$t_0 = t, \quad t_1 = \varepsilon t.$$

Therefore, substituting

$$x' = \frac{\partial x}{\partial t_0} + \varepsilon\frac{\partial x}{\partial t_1} \tag{5.4}$$

$$x'' = \frac{\partial^2 x}{\partial t_0^2} + 2\varepsilon\frac{\partial^2 x}{\partial t_1 \partial t_2} + \varepsilon^2\frac{\partial^2 x}{\partial t_1^2}, \tag{5.5}$$

in (5.3) *and imposing that every coefficient of every power of ε has to be zero, we obtain the following initial value problem, which will be needed to be* *solved sequentially:*

$$\frac{\partial^2 x_0}{\partial^2 t_0} + x_0 = 0, \quad x_0(0,0) = 1, \frac{\partial x_0}{\partial t_0}(0,0) = 0, \tag{5.6}$$

$$\frac{\partial^2 x_1}{\partial^2 t_0} + x_1 = -2(\frac{\partial^2 x_0}{\partial t_0 \partial t_1} + \frac{\partial x_0}{\partial t_0}), \quad x_1(0,0) = 0, \frac{\partial x_1}{\partial t_0}(0,0) + \frac{\partial x_0}{\partial t_1}(0,0) = 0 \tag{5.7}$$

$$\frac{\partial^2 x_2}{\partial^2 t_0} + x_2 = -2(\frac{\partial^2 x_1}{\partial t_0 \partial t_1} + \frac{\partial x_1}{\partial t_0}) - \frac{\partial^2 x_0}{\partial t_1^2} - 2\frac{\partial x_0}{\partial t_1} - 2\frac{\partial^2 x_0}{\partial t_0 \partial t_1} \tag{5.8}$$

$$x_2(0,0) = 0, \frac{\partial x_2}{\partial t_0}(0,0) + \frac{\partial x_1}{\partial t_1}(0,0) = 0 \tag{5.9}$$

solving sequentially this equations as done in [6] we can construct the solution.

In a certain sense, we can see this method as combination of the method of strained coordinates and the regular perturbation method. In the following chapter, we will use both the multiple scale analysis and the averaging method, in order to derive a set of *homogenized* equations, that are equations for the averaged quantities involved in the original system, where the multiple scale method has been applied previously.

Chapter 6

Application on systems: 1D Shallow Water, 2D Shallow Water and Euler Equations

6.1 Introduction

In this chapter, we finally apply what was said in the previous chapters to some systems of quasi-linear hyperbolic conservation laws. The systems we are going to analyse are the 1D and 2D *Saint-Venant Equations*, usually known as *Shallow Water Equations*, and the 1D Euler system. For each system, we will study the dispersive behaviour of the solutions when the background is periodic, trying to detect it using the multiple-scale method combined with the averaging of the unknowns that varies in a much faster scale then the other (see [16] for more), and constructing in this way systems for the averages, or *homogenized systems*, derived from the original ones, which dispersiveness of the solutions can be observed from an analytic point of view and not just from a numerical observation anymore.

Then, for each system, we will focus on the computation of travelling waves of these homogenized systems and the comparison between this travelling waves and the solitary wave of the original system, in order to observe the closeness between them. We will also give, for the 1D shallow water system and the Euler system, the comparison also between the numerical solution of the original system and the one of the homogenized system. This chapter is entirely based on some very recent works, already published such as [7], under review as [8], or recently submitted like [9] of professors Giovanni Russo, David Ketcheson et al. on this topic, while all the original results will be shown in the next chapters.

6.2 1D Shallow water equations

The system we study, already introduced in Chapter 1 in this section is

$$h_t + (hu)_x = 0, \tag{6.1}$$

$$(hu)_t + \left(hu^2 + \frac{1}{2}gh^2\right)_x = -ghb_x, \tag{6.2}$$

where h is the water depth, u is the depth-averaged velocity, b denotes the bottom elevation or bathymetry and g is the gravitational acceleration. We are interested in the behavior of waves propagating over periodic bathymetry with period δ, i.e. $b(x) = b(x+\delta)$. We consider waves with a wavelength significantly larger than δ, and situations where the variation in $b(x)$ is comparable to the total depth, yet not large enough to produce dry states. The notation and scales involved are depicted in Figure 6.1.

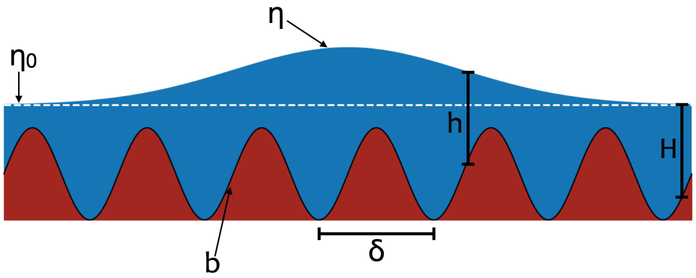


Figure 6.1: Physical description of the quantities involved. This figure is reproduced from [7] with the permission of the authors.

We already saw in Chapter 1 that the behaviour of the solution of this system is different whether the bathymetry is periodic or not, so our purpose is to find another version of this system where this behaviour can be observe even theoretically instead of just numerically as shown in figure 1.1. Let us rewrite the system (6.1)-(6.2) in terms of the surface elevation $\eta = h + b$ and the discharge $q = hu$. We obtain

$$\eta_t + q_x = 0, \tag{6.3}$$

$$q_t + \Big(\frac{q^2}{\eta - b}\Big)_x + g(\eta - b)\eta_x = 0. \tag{6.4}$$

This system is not in conservative form, but since the solution we are interested more does not contain shock, we do not actually need to work with a conservative version of the equations.

Let us now apply to (6.3)-(6.4) the multiple scale technique, following the example of [16]. We introduce a small parameter δ and fast spatial scale $y = x/\delta$ and treat the two spatial scales formally as independent variables, so that

$$\frac{\partial}{\partial x} \to \frac{\partial}{\partial x} + \frac{1}{\delta}\frac{\partial}{\partial y}.$$

We assume that both η, q can be written as power series expansion in δ of the form:

$$\eta = \eta^0(x,t) + \delta\eta^1(x,y,t) + \delta^2\eta^2(x,y,t) + \dots \tag{6.5}$$

$$q = q^0(x,t) + \delta q^1(x,y,t) + \delta^2 q^2(x,y,t) + \dots \tag{6.6}$$

where we suppose that every term of the expansion is periodic with respect to y in its period, and where superscripts on η and q are indices, while superscripts on other quantities are exponents. Substituting these expansions in the equations (6.3)-(6.4) we see that there is only one term proportional to δ^{-1}, given by

$$\frac{q_0^2 H'(y)}{(\eta - b)^2},$$

where we define

$$H(y) = \eta^0 - b(y).$$

This makes sense because later we will see that η^0 is a constant. We will assume $H' > 0$. This implies $q^0 = 0$. Up to $O(\delta^3)$, we obtain the following:

$$\eta_y^0 + q_y^1 + \delta(q_y^2 + q_y^1 + \eta_t^1) + \delta^2(q_y^3 + q_x^2 + \eta_t^2) = O(\delta^3) \tag{6.7}$$

$$
\begin{gathered}
g(\eta_x^0+\eta_y^1)H+\\
+\delta\Big(q_t^1+2q^1q_y^1H^{-1}+g((\eta_x^0+\eta_y^1)\eta^1+(\eta_x^1+\eta_y^2)H)-(q^1)^2H'H^{-2}\Big)+\\
+\delta^2\Big(2q^1q_x^1H^{-2}+q_t^2+2((q^2-q^1\eta^1H^{-1})q_y^1+q^1q_y^2)H^{-1}-(q^1)^2\eta_y^1H^{-2}+\\
g((\eta_x^0+\eta_y^1)\eta^2+(\eta_x^1+\eta_y^2)\eta^1+(\eta_x^2+\eta_y^3)H)-2(q^2-q^1\eta^1H^{-1})q^1H'H^{-2}\Big)=\\
=O(\delta^3).
\end{gathered}
\tag{6.8}
$$

Now we will proceed in the following way.

- We impose that every coefficient of every power of δ in the expansions has to be zero, solving the equations we obtain for the terms with highest index;
- we average the resulting equations with respect to y and solve the obtained system for the averaged quantities;
- Integrate in order to Determine expressions for the variables with the highest indices by integrating, expressing them in terms of y-averages of lower-index variables and the function $H(y)$.

The description of the averaging operators used is shown in [7], as the step-by-step resolution of the increasingly cumbersome and complicated systems obtained for each δ that will be omitted here but that the interesting reader can find in this paper. We just need to say for our purpose that we define the average of f over one period, and denote it as $\langle f\rangle$, the quantity

$$\langle f\rangle=\int_0^1 f(y)dy.$$

We obtain, defined the averaged variables

$$\overline{\eta}=\langle\eta^1\rangle+\delta\langle\eta^2\rangle+\dots, \tag{6.9}$$

$$\overline{q}=\langle q^1\rangle+\delta\langle q^2\rangle+\dots, \tag{6.10}$$

and summing terms up to $O(\delta^5)$, the following equations:

$$\delta(\overline{\eta}+\overline{q})=O(\delta^6) \tag{6.11}$$

$$
\begin{gathered}
\delta(\overline{q}_t + c^2\overline{\eta}_x) + \delta^2\frac{\langle H^{-2}\rangle}{\langle H^{-1}\rangle}((\overline{q}^2)_x - \overline{\eta q}_t)+ \\
+\delta^3\Big(- \frac{\mu}{c^2}\overline{q}_{ttt} - \frac{\alpha_2}{c^2}\overline{q}^2\overline{q}_t - \frac{\langle H^{-3}\rangle}{\langle H^{-1}\rangle}\eta(2(\overline{q}^2)_x - \overline{\eta q}_t)\Big)+ \\
+\delta^4\Big(\frac{\hat{\alpha}_4}{c^2}\overline{q}^3\overline{q}_x + \frac{\hat{\alpha}_6}{c^2}\overline{q}^2\overline{\eta q}_t - 2\frac{\gamma}{c^2}(2\overline{q}_x\overline{q}_{tt} - \overline{\eta q}_{ttt}) + \frac{\langle H^{-4}\rangle}{\langle H^{-1}\rangle}\overline{\eta}^2(3(\overline{q}^2)_x - \overline{\eta q}_t)\Big)+ \\
+\delta^5\Big(\frac{\nu_1}{c^4}\overline{q}_{ttttt} + \frac{\nu_2}{c^4}\overline{q}_{xxttt} + F(\overline{\eta},\overline{q})\Big) = O(\delta^6)
\end{gathered}
\tag{6.12}
$$

where the coefficients and the explicit form of $F(\overline{\eta},\overline{q})$ can be found in the appendix A.

The formulation of the homogenized system (6.11)-(6.12) presents two main drawbacks. Firstly, it involves high-order time derivatives, whereas the original shallow water equations were first-order in time. One possible way to address this issue is to introduce additional variables to represent the time derivatives of the dependent quantities. However, a more critical problem arises from the fact that these equations develop a linear instability at low wavenumbers. Indeed, let us consider the system (6.11)-(6.12) up to $O(\delta^5)$ linearized around $(0,0)$ given by

$$\eta_t + q_x = 0 \tag{6.13}$$

$$q_t + c^2\eta_x - \frac{\hat{\mu}}{c^2}q_{ttt} = 0 \tag{6.14}$$

where we drop the average sign for simplicity and where $\hat{\mu} = \delta^2\mu$. Looking for solutions of the system of the kind $\eta = \hat{\eta}\exp(i(kx-\omega t)), q = \hat{q}\exp(i(kx-\omega t))$ we obtain, inserting this ansatz, the following:

$$-i\omega\hat{\eta} + ik\hat{q} = 0 \tag{6.15}$$

$$-i\omega\hat{q} + ikc^2\hat{\eta} - i\frac{\hat{\mu}}{c^2}\omega^3\hat{q} = 0. \tag{6.16}$$

It is possible to determine non-trivial solution for this system if the determinant of the coefficient matrix is zero, which corresponds to *dispersion relation*

$$\omega^2 + \frac{\hat{\mu}}{c^2}\omega^4 - c^2k^2 = 0,$$

that becomes, once we divide by c^2k^2,

$$\Omega^2 + K^2\Omega^4 - 1 = 0, \quad \Omega = \frac{\omega}{ck}, \quad K = k\sqrt{\hat{\mu}}.$$

The four roots of this equation are given by

$$\Omega^2 = Z_\pm, \quad Z_\pm = \frac{1 \pm \sqrt{1+4K^2}}{2K^2},$$

therefore we have two real roots and two imaginary roots, and the existence of a root with positive coefficient of the imaginary part indicates that the initial value problem for system (6.13)-(6.14) is ill-posed.

One way to remedy this is to convert all higher-order time derivatives to space derivatives. This is accomplished by differentiating the equations and using equality of mixed partial 8 derivatives, keeping again only terms up to the desired order in δ:

$$\delta(\overline{\eta} + \overline{q}) = O(\delta^6) \tag{6.17}$$

$$\begin{gathered}
\delta(\overline{q}_t + c^2\overline{\eta}_x) + \delta^2 \frac{\langle H^{-2}\rangle}{\langle H^{-1}\rangle}((\overline{q}^2)_x + c^2\overline{\eta}\ \overline{\eta}_x) + \\
+\delta^3\Big(c^2\mu\overline{\eta}_{xxx} + \alpha_1\overline{q\eta q}_x - \alpha_2\overline{q}^2\overline{\eta} + g\alpha_3\overline{\eta}^2\overline{\eta}_x\Big) + \\
+\delta^4\Big(\frac{\hat{\alpha}_4}{c^2}\overline{q}^3\overline{q}_x + \alpha_5\overline{q\eta}^2\overline{q}_x + \alpha_6\overline{q}^2\overline{\eta\eta}_x + g\alpha_7\overline{\eta}^3\overline{\eta}_x + \hat{\alpha}_8\overline{q}_x\overline{q}_{xx} + g\hat{\alpha}\overline{\eta}_x\overline{\eta}_{xx} + \\
+g\hat{\alpha}_{10}\overline{\eta\eta}_{xxx} + \hat{\alpha}_{11}\overline{qq}_{xxx}\Big) = O(\delta^5),
\end{gathered} \tag{6.18}$$

where again coefficients are provided in A. The system is linearly stable for small wavenum bers. However, it is unstable for sufficiently large wavenumbers. Let us show it. Proceeding as done for (6.11)-(6.12), linearizing the system (6.17)-(6.18) up to $O(\delta^3)$ we obtain

$$\eta_t + q_x = 0 \tag{6.19}$$

$$q_t + c^2\eta_x + c^2\hat{\mu}\eta_{xxx} = 0, \tag{6.20}$$

which dispersion relation is

$$\omega^2 - c^2k^2 + \hat{\mu}c^2k^4 = 0,$$

that dividing by c^2k^2 becomes

$$\Omega^2 + K^2 = 1.$$

Solution are

$$\Omega_{\pm} = \pm\sqrt{1 - K^2}$$

As a consequence, the dispersion relation associated with the second form of the system imposes a restriction on the wavenumbers, allowing only bounded values. Specifically, in its non-dimensional form, it requires $|K| \leq 1$, which translates to

$$|k| \leq \frac{1}{\delta\sqrt{\mu}}.$$

This issue is less critical, and the solutions of this system remain in good agreement with those of the original shallow water equations. Nevertheless, it is possible to derive a system that maintains linear stability for all wavenumbers while remaining equivalent to the previous formulations up to the given order in δ. This can be achieved by expressing the linear term involving q_{ttt} in terms of q_{xxt}, making use of the equality of mixed partial derivatives.

The new system is:

$$\delta(\overline{\eta} + \overline{q}) = O(\delta^6) \tag{6.21}$$

$$\begin{gathered}
\delta(\overline{q}_t + c^2\overline{\eta}_x) + \delta^2\frac{\langle H^{-2}\rangle}{\langle H^{-1}\rangle}((\overline{q}^2)_x + c^2\overline{\eta}\;\overline{\eta}_x) + \\
+\delta^3\Big(- \mu\overline{q}_{xxt} + \alpha_1\overline{q\eta q}_x - \alpha_2\overline{q}^2\overline{\eta} + g\alpha_3\overline{\eta}^2\overline{\eta}_x\Big) + \\
+\delta^4\Big(\frac{\alpha_4}{g}\overline{q}^3\overline{q}_x + \alpha_5\overline{q\eta}^2\overline{q}_x + \alpha_6\overline{q}^2\overline{\eta\eta}_x + \\
+g\alpha_7\overline{\eta}^3\overline{\eta}_x + \alpha_8(2\overline{q}_x\overline{q}_{xx} + c^2\overline{\eta\eta}_{xxx}) + \alpha_9(5c^2\overline{\eta}_x\overline{\eta}_x x + 2\overline{qq}_{xxx})\Big) + \\
+\delta^5(\nu_1 + \nu_2 - \mu^2)\overline{q}_{xxxxt} + F(\overline{\eta}, \overline{q}) = O(\delta^6).
\end{gathered} \tag{6.22}$$

Let us see the dispersion relation for this new system. Linearizing (6.21)-(6.22) up to $0(\delta^3)$ around $(0, 0)$ we obtain

$$\eta_t + q_x = 0 \tag{6.23}$$

$$q_t + c^2\eta_x - \hat{\mu}q_{xxt} = 0, \tag{6.24}$$

which leads to the dispersion relation

$$\omega^2 - c^2k^2 + \hat{\mu}\omega^2k^2 = 0,$$

which dividing by c^2k^2 becomes

$$\Omega^2(1+K^2)-1=0,$$

that is solved by

$$\Omega_\pm = \pm\frac{1}{\sqrt{1+K^2}},$$

therefore there are no unstable modes. The same can be shown if we linearize the system including the $O(\delta^5)$ terms, as one can see in [7]. Let us know show, through numerical computation, how good is the approximation we are doing. In figure 6.2 we show the comparison between the numerical solution of the original shallow water system, computed through a method involving Lax-Wendroff scheme with limiters and then a 5-th order WENO reconstruction in space and a 4-th order Runge-Kutta integration in time, and the numerical solution of the homogenized systems (6.21)-(6.22) up to $0(\delta^3)$ and $O(\delta^5)$ terms, solved with a Fourier pseudo-spectral method. More details on the application of this method, in addition to the bathymetry and the initial conditions chosen for this example, can be found on [7]. We can see that the homogenized system is a good approximation at least at early times, while for later times it becomes less accurate. Moreover, as expected, the more terms in the expansion we consider, the more accurate the approximation is. In this kind of approximation, a compromise between accuracy and computation cost is needed. What is important is that the homogenized system can be considered a good approximation, and its dispersive nature and properties can be studied properly.

6.2.1 Travelling waves

Let us focus on the computation of travelling waves, i.e. solutions of the system that does not depend on x and t in separated ways but depend on a unique variable $\xi = x - Vt$, where V is the velocity of the travelling wave. For our purpose, we consider the last version of the system, given by (6.21)-(6.22).

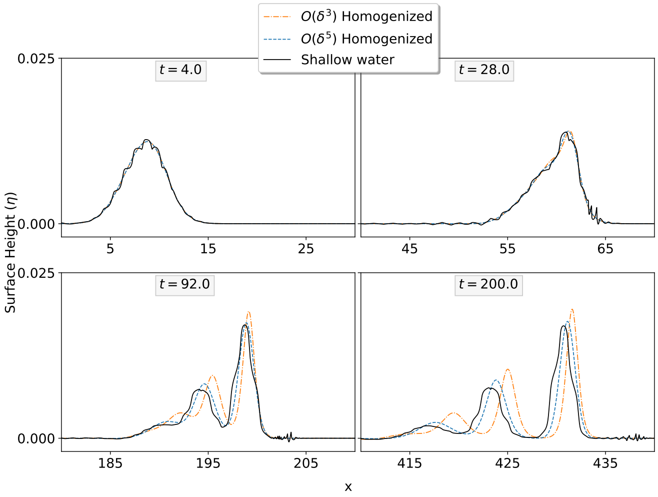


Figure 6.2: Comparison between homogenized and direct solutions over a piece-wise periodic bathymetry. This figure is reproduced from [7] with the permission of the authors.

6.2.2 Travelling waves to $O(\delta^3)$

Let us consider the case where we account for terms up to $O(\delta^3)$.

$$\eta_t + q_x = 0 \tag{6.25}$$

$$q_t + c^2\eta_x + \hat{\beta}_1\eta\eta_x + \hat{\beta}_2(q^2)_x - \hat{\beta}_3 q\eta q_x - \hat{\beta}_4 q^2\eta_x - \hat{\beta}_5\eta^2\eta_x - \hat{\mu}q_{xxt} = 0, \tag{6.26}$$

where the coefficients can be derived from (6.21)-(6.22). Looking for solutions that depends on ξ, the first equation becomes

$$-V\eta' + q' = 0,$$

which is solved by

$$q = q_0 + V(\eta - \eta_0).$$

We suppose $q_0 = 0, \eta_0 = 0$, so we have $q = V\eta$. Replacing it in the second equation we obtain

$$-V^2\eta' + c^2\eta' + \hat{\beta}_1\eta\eta' + V^2\hat{\beta}_2(\eta^2)' - V^2\hat{\beta}_3\eta^2\eta' - \hat{\beta}_4V^2\eta^2\eta' - \hat{\beta}_5\eta^2\eta' + V^2\hat{\mu}\eta''' = 0,$$

that, observing $\eta\eta' = \frac{1}{2}(\eta^2)'$ and $\eta^2\eta' = \frac{1}{3}(\eta^3)'$, can be written as

$$\frac{d}{d\xi}(\gamma_1\eta - \gamma_2\eta^2 + \gamma_3\eta^3 - V^2\hat{\mu}\eta'') = 0,$$

where the coefficients can be again deducted from the previous equation. Integrating we obtain

$$\eta'' = F(\eta), \quad F(\eta) = (\gamma_1\eta - \gamma_2\eta^2 + \gamma_3\eta^3 + A)/(\hat{\mu}V^2),$$

where A is the integration constant, which value is $A = 0$ since we suppose that $\eta \to 0$ when $x \to \infty$. This equation admits a first integral, which plays the role of total energy: multiplying the equation by η' and integrating we obtain

$$\frac{1}{2}(\eta')^2 + U(\eta) = E, \quad U(\xi) = (-\frac{1}{2}\gamma_1\eta^2 + \frac{1}{3}\gamma_2\eta^3 - \frac{1}{4}\gamma_3\eta^4)/(\hat{\mu}V^2).$$

This can be solved now using the approach described in Section 4.4. The results are shown in Figure 6.3.

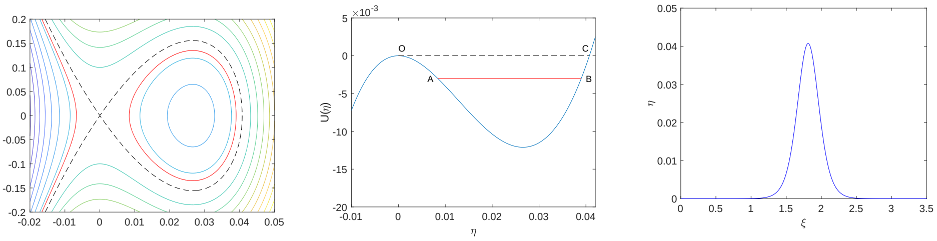


Figure 6.3: From the left to the right: The phase plane analysis, where lines have the same potential and the red line is the separatrix; The potential from where the trajectories of the first panel are derived; The travelling wave solution. These figures are reproduced from [7] with the permission of the authors.

6.2.3 Travelling waves to $O(\delta^5)$

Let us consider now the complete system (6.21)-(6.22). Using the same reasoning as the previous case, we obtain $q = V\eta$, that replaced in the second equation let us obtain the following:

$$-\gamma_1\eta' + 2\gamma_2\eta\eta' - 3\gamma_3\eta^2\eta' - 4\gamma_4\eta^3\eta' + 2\gamma_5\eta'\eta'' + 2\gamma_6\eta'\eta''' + \hat{\mu}\eta''' - \hat{\nu}\eta^{(5)} = 0.$$

Integrating we obtain

$$-\gamma_1\eta + \gamma_2\eta^2 - \gamma_3\eta^3 - \gamma_4\eta^4 + \gamma_5(\eta')^2 + \gamma_6(2\eta'\eta'' - (\eta')^2) + \hat{\mu}\eta'' - \hat{\nu}\eta^{(4)} = A.$$

where A is the constant of integration, which value is again $A = 0$ from the behaviour of η at infinity. In this case, the usual analogy with mechanics can not be done, since no integral prime seem to appear in this equation. In [7], a different approach has been used to solve numerically, and Figure 6.4 shows the comparison between the solitons of the 1D shallow water system (6.3)-(6.4) and the ones of the homogenized system (6.21)-(6.22) up to $O(\delta^3)$ and $O(\delta^5)$. It is possible to notice that the $O(\delta^5)$ solitary wave is in much

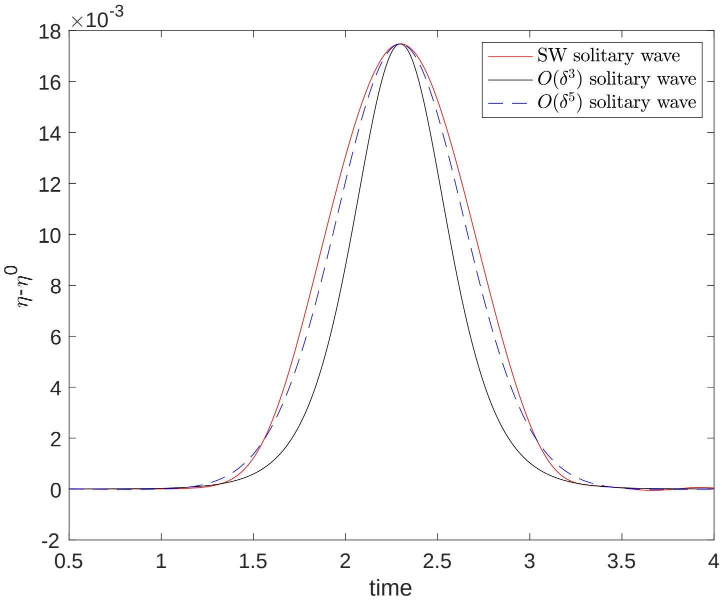


Figure 6.4: Comparison between the direct solution of the shallow water equations and the travelling wave solutions of the homogenized equations. This figure is reproduced from [7] with the permission of the authors.

better agreement with the true solitary wave, compared to the $O(\delta^3)$ one.

6.3 2D Shallow water

In this section, we study the system

$$h_t + (hu)_x + (hv)_y = 0 \tag{6.27}$$

$$(hu)_t + (hu^2 + \frac{1}{2}gh^2)_x + (huv)_y = -ghb_x \tag{6.28}$$

$$(hv)_t + (hv^2 + \frac{1}{2}gh^2)_y + (huv)_x = -ghb_y \tag{6.29}$$

where g is the gravitational acceleration, $h(x, y, t)$ is the depth, $u(x, y, t)$ and $v(x, y, t)$ are the horizontal velocity components and $b(x, y)$ is the bottom elevation. We focus on the behaviour of waves propagating over bathymetry that does not depend on x, and it is periodic in y with period δ:

$$b(y + \delta) = b(y).$$

Considering the propagation of initially planar waves moving parallel to the x-axis, we impose the initial conditions:

$$\eta(x, y, 0) = \eta_0(x), \quad u(x, y, 0) = u_0(x), \quad v(x, y, 0) = 0.$$

where $\eta = h + b$ is again the surface elevation. In [18],[22] it has been shown that linear planar waves travelling parallel to variations exhibit effective dispersion, and this can lead to the formation of solitary waves even though the equations themselves are non-dispersive. So what we do is to develop, through multiple-scale analysis inspired again by [16], a model that can described the dispersiveness of the waves. What we will obtain is a Boussinesq-type system, similar to the one described in the chapter about dispersive equations, with dispersive coefficient depending on the bathymetry.

For our purpose, let us rewrite our equation in a different way, giving up on the conservative form (it is not a problem since we are interested in classical solutions and not in the weak ones). The quantities involved in the new form of the equations are $\eta = h + b, p = hv, u$, and the new system is the following

$$\eta_t + (u(\eta - b))_x + (p)_y = 0 \tag{6.30}$$

$$u_t + uu_x + g\eta_x + \frac{p}{\eta - b}u_y = 0 \tag{6.31}$$

$$p_t + \Big(\frac{p^2}{\eta - b}\Big)_y + g(\eta - b)\eta_y + (pu)_x = 0. \tag{6.32}$$

Now, in order to develop the multiple scale method, we need to understand how the variables scale and how we can redefine the variables in order to detect this behaviour. Since the bathymetry is periodic with respect to y of period δ, and assuming that that the wavelength of the typical waves we are interested in is long relative to the period δ of the bathymetry, the change of variables we should perform is $(x, t, y) \to (x, t, \tilde{y})$, where $\tilde{y} = y/\delta$. In this way, $\frac{\partial}{\partial y} = \delta^{-1}\frac{\partial}{\partial \tilde{y}}$ and b is a $1-$periodic function of $\tilde{y}$. The system in the variables, suppressing the tilde for simplicity of notations, becomes:

$$\eta_t + (u(\eta - b))_x + \delta^{-1}(p)_y = 0 \tag{6.33}$$

$$u_t + uu_x + g\eta_x + \delta^{-1}\frac{p}{\eta - b}u_y = 0 \tag{6.34}$$

$$p_t + \delta^{-1}\Big(\frac{p^2}{\eta - b}\Big)_y + \delta^{-1}g(\eta - b)\eta_y + (pu)_x = 0 \tag{6.35}$$

Now we look for solutions that are small perturbations from the state $(\eta, u, p) = (\eta_0, 0, 0)$, i.e. solutions that can be written as power series of δ in the form

$$\eta - \eta^0 = \delta\eta^1(x, y, t) + \delta^2\eta^2(x, y, t) + \dots \tag{6.36}$$

$$u = \delta u^1(x, y, t) + \delta^2 u^2(x, y, t) + \dots \tag{6.37}$$

$$p = \delta p^1(x, y, t) + \delta^2 p^2(x, y, t) + \dots, \tag{6.38}$$

where we assume that all the terms of the expansion are $1-$periodic with in the y variable. Substituting the expansions in the system (6.33)-(6.34)-(6.35) and imposing that each coefficient of powers of δ has to be zero we obtain:

- $O(\delta^0)$: the expansion of (6.31) does not contain terms of this kind, while for the other two equations we have

$$p^1_y = 0 \tag{6.39}$$

$$gH(y)\eta^1_y = 0, \tag{6.40}$$

so from these relations we deduce that p^1, η^1 do not depend on y.

- $O(\delta^1)$: collecting terms proportional to δ^1 we obtain

$$\eta^1_t + Hu^1_x + p^2_y = 0 \tag{6.41}$$

$$u^1_t + g\eta^1_x = 0 \tag{6.42}$$

$$p^1_t - \frac{(p^1)^2}{H^2}H'(y) + gH\eta^2_y = 0 \tag{6.43}$$

Averaging these equations with respect to y, we obtain the following facts:

- Equation (6.42) implies, since η^1 is independent from y, that u^1_t is also independent from y, and from the initial condition we deduce that also u^1 does not depend on y, obtaining

$$\langle u^!_t \rangle + g\langle \eta^1_x \rangle = 0;$$

- Solving (6.43) for η^2_y and averaging we have

$$\frac{\langle H^{-1} \rangle}{g} p^1_t,$$

that implies p^1_t, that implies $p^1 = 0$ and this means, from (6.43), $\eta^2_y = 0$, so $\eta^2 = \langle \eta^2 \rangle$;

- Averaging (6.41) we have

$$\langle \eta^1_t \rangle + \langle H \rangle \langle u^1_x \rangle = 0.$$

Taking together these averaged equations, we have the system

$$\langle u^!_t \rangle + g\langle \eta^1_x \rangle = 0; \tag{6.44}$$

$$\langle \eta^1_t \rangle + \langle H \rangle \langle u^1_x \rangle = 0. \tag{6.45}$$

- $O(\delta^2)$: collecting term proportional to δ^2, and remembering that some quantities are independent from y, we obtain

$$\eta_t^2 + Hu_x^2 + (\langle\eta^1\rangle\langle u^1\rangle)_x = -p_y^3 \tag{6.46}$$

$$u_t^2 + \langle u^1\rangle\langle u_x^1\rangle + g\langle\eta_x^2\rangle = 0 \tag{6.47}$$

$$\frac{1}{H}(p_t^2 + g\langle\eta^1\rangle\langle\eta_y^2\rangle) = -g\eta_y^3 \tag{6.48}$$

Since every other term in (6.47) is independent from y, also u^2 has to be independent from it, so this equation becomes

$$\langle u_t^2\rangle + \langle u^1\rangle\langle u_x^1\rangle + g\langle\eta_x^2\rangle = 0. \tag{6.49}$$

Averaging (6.48), we obtain $\langle p_t^2\rangle$ and as a consequence $\langle p^2\rangle$.

Based on what we have determined up to this point, we can write the series expansions more simply as

$$\eta - \eta^0 = \delta\eta^1(x,t) + \delta^2\eta^2(x,t) + \dots \tag{6.50}$$

$$u = \delta u^1(x,t) + \delta^2 u^2(x,t) + \dots \tag{6.51}$$

$$p = \delta^2 p^2(x,t) + \dots. \tag{6.52}$$

- $O(\delta^3)$: collecting terms proportional to δ^3, we obtain

$$\eta_t^3 + Hu_x^3 + (\eta^1 u^2 + \eta^2 u^1)_x = -p_y^4 \tag{6.53}$$

$$u_t^3 + (u^1 u^2)_x + [[H^{-1}[[H]]]]u_{xtx}^1 + g\eta_x^3 = 0 \tag{6.54}$$

$$\frac{1}{H}(p_t^3 - H^{-2}(p^2)^2 H' + 2H^{-1}p^2 p_y^2 + g\eta^1\eta_y^3 + (p^2 u^1)_x) = -g\eta_y^4, \tag{6.55}$$

where the averaging operator is described in A. Averaging (6.53) yields

$$\langle\eta^3\rangle_t + \langle Hu^3\rangle_x + (\langle\eta^1\rangle\langle u^2\rangle + \langle\eta^2\rangle\langle u^1\rangle)_x = 0. \tag{6.56}$$

Since u^3 depends on y, we must work directly with $\langle Hu^3\rangle$. Multiplying the (6.54) by H and averaging we obtain

$$\begin{gathered}\langle Hu^3\rangle_t + \langle H\rangle(\langle u^1\rangle\langle u^2\rangle)_x - \mu\langle u_{xxt}^1\rangle + g\langle H\rangle\langle\eta_{xt}^3\rangle = 0,\\ \mu = -\langle H[[H^{-1}[[H]]]]\rangle.\end{gathered} \tag{6.57}$$

Introducing the quantity q^j such that $\langle q^j \rangle = \langle H u^j \rangle$ and observing that $\langle f \rangle \langle g \rangle = \langle fg \rangle$ in the case that at least one of the two quantities does not depend on y, the equation (6.57) becomes

$$\langle q^3 \rangle_t + \langle H^{-1} \rangle (\langle q^1 \rangle \langle q^2 \rangle)_x - \langle H^{-1} \rangle \mu \langle q^1_{xxt} \rangle + g \langle H \rangle \langle \eta^3_x \rangle = 0. \qquad (6.58)$$

Finally, averaging (6.55) we obtain $\langle p^3 \rangle = 0$.

Now we can finally construct out Boussinesq type system. Let $\overline{q} = \langle q^1 \rangle + \delta \langle q^2 \rangle + \dots, \overline{\eta} = \langle \eta^1 \rangle + \delta \langle \eta^2 \rangle + \dots$. The formula

$$\delta \langle H \rangle \times (6.45) + \delta^2 \langle H \rangle \times (6.49) + \delta^3 \times (6.58)$$

gives

$$\delta(\overline{q}_t + g \langle H \rangle \overline{\eta}_x) + \delta^2 \langle H \rangle^{-1} \overline{q}\, \overline{q}_x - \delta^3 \langle H \rangle^{-1} \mu \overline{q}_{xxt} = O(\delta^4), \qquad (6.59)$$

while the formula

$$\delta \times (6.44) + \delta^2 \langle H \rangle \times (6.49) + \delta^3 \times (6.56)$$

results in

$$\delta(\overline{\eta}_t + \overline{q}_x) + \delta^2 \langle H \rangle^{-1} (\overline{\eta}\, \overline{q})_x = O(\delta^4). \qquad (6.60)$$

Dividing by δ each equation, considering terms up to $O(\delta^3)$ and dropping the overline bar for simplicity of notation we obtain the following system:

$$q_t + a_1 \eta_x + a_2 q q_x - \tilde{\mu} q_{xxt} = 0 \qquad (6.61)$$

$$\eta_t + q_x + a_2 (\eta q)_x = 0, \qquad (6.62)$$

where $a_1 := g \langle H \rangle$, $a_2 := \delta / \langle H \rangle$, $\tilde{\mu} = \delta^2 \mu / \langle H \rangle$. This system is a generalization of the Boussinesq system studied in section 4.3.2: indeed, if $g = \langle H \rangle = 1, \mu = 1/3$ we obtain the classical Boussinesq system. As done for the 1D shallow water system, let's solve numerically both the original system (6.27)-(6.28)-(6.29) and the homogenized system (6.61)-(6.62). The comparison is shown in Figure 6.5, where a piecewise-constant periodic bathymetry is considered. For the homogenized system, the solution has been computed through a Fourier pseudo-spectral discretization in space and an explicit

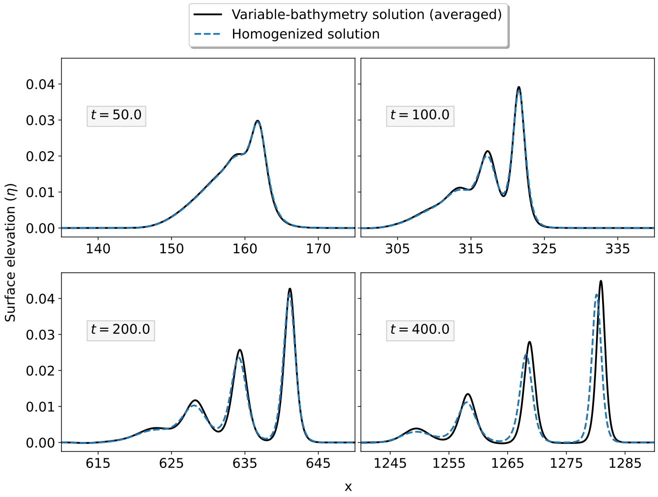


Figure 6.5: Comparison between homogenized and direct solutions over a piece-wise periodic bathymetry. This figure is reproduced from [8] with the permission of the authors.

3-stage 3rd order SSP Runge-Kutta integration in time, while for the starting hyperbolic system (6.27)-(6.28)-(6.29), a method based on a 5-th order WENO reconstruction in space and a 4-th order Runge-Kutta integration in time has been used. For more details on the methods, the bathymetry and the initial condition see [8]. What one can observe from Figure 6.5 is that there is a very small difference between the two solutions, that shows up at late times. This implies that the homogenized system is a very good approximation. Similar observation can be done in the case of smooth bathymetry.

6.3.1 Travelling wave

Let us now focus on the computation of the travelling wave for the system (6.61)-(6.62), i.e. a solution that depends on $\xi = x - Vt$ propagating on a lake at rest, so that the unperturbed state is q_0, η_0. Assuming η, q are functions of ξ we obtain the following system of PDEs:

$$-Vq' + a_1\eta' + a_2qq' + \tilde{\mu}Vq''' = 0 \tag{6.63}$$

$$-V\eta' + q' + a_2(\eta q)' = 0, \tag{6.64}$$

that can be written as

$$\frac{d}{d\xi}(-Vq + a_1\eta + a_{\frac{1}{2}}a_2q^2 + \tilde{\mu}Vq'') = 0 \tag{6.65}$$

$$\frac{d}{d\xi}(-V\eta + q + a_2\eta q) = 0, \tag{6.66}$$

which gives

$$-Vq + a_1\eta + a_{\frac{1}{2}}a_2q^2 + \tilde{\mu}Vq'' = C_1 \tag{6.67}$$

$$-V\eta + q + a_2\eta q = C_2. \tag{6.68}$$

Since we have that the perturbation of the lake at rest is 0, then the two integration constants are both zero. From (6.68) we can express η in function of q, obtaining

$$\eta = \frac{q}{V - a_2q}, \tag{6.69}$$

and substituting it in (6.67) we obtain the ODE

$$q'' = \frac{Vq - \frac{a_1q}{V - a_2q} - \frac{1}{2}a_2q^2}{\tilde{\mu}V} \tag{6.70}$$

that is of the form $q'' = F(q)$. Multiplying by q' and integrating we obtain

$$\frac{d}{d\xi}\Big(\frac{1}{2}(q')^2 + U(q)\Big) = 0$$

where

$$U(q) = -\int F(q)dq = \Big(\frac{1}{6}a_2q^3 - \frac{1}{2}Vq^2 - \frac{a_1}{a_2} - \frac{a_1}{a_2^2}Vlog(1 - a_2q/V)\Big)/(\tilde{\mu}V)$$

is the potential of F. Integrating we have

$$\frac{1}{2}(q')^2 + U(q) = E,$$

the analogue of total energy conservation. The results are shown in Figure 6.6

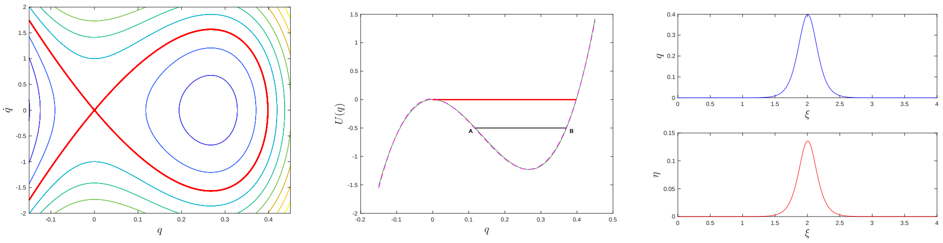


Figure 6.6: From the left to the right: The phase plane analysis, where lines have the same potential and the red line is the separatrix; The potential corresponding to $V = 3/10$; The travelling wave solutions. This figure is reproduced from [8] with the permission of the authors.

6.4 1D Euler Equations

Let us work now on the one-dimensional Euler equations (2.92)-(2.93)-(2.94) written in the following form:

$$\rho_t + (\rho u)_\chi = 0, \tag{6.71}$$

$$(\rho u)_t + (\rho u^2 + p)_\chi = 0, \tag{6.72}$$

$$\left(\frac{1}{2}\rho u^2 + \rho e\right)_t + \left(\frac{1}{2}\rho u^3 + \rho e u + up\right)_\chi = 0 \tag{6.73}$$

that, as said, represents the conservation of mass, momentum and energy. Here $\rho(\chi, t)$ is the mass per unit volume, $u(\chi, t)$ the gas velocity, $p(\chi, t)$ the gas pressure and e is the internal energy per unit mass. We consider a polytropic gas, for which

$$e = \frac{1}{\gamma - 1}\frac{p}{\rho},$$

where $\gamma = c_p/c_v$ is the polytropic constant, given by the ratio of specific heats at respectively constant pressure and constant volume. Let us see how this model can fall into our case study. Here, the "periodic media" where the "wave" propagates is given by the density, that varies periodically: indeed, considered the background density

$$\hat{\rho}(\chi)(= \begin{cases} 1/4 & \text{for} \quad 0 < \chi - \lfloor \chi \rfloor < 1/2, \\ 7/4 & \text{for} \quad 1/2 < \chi - \lfloor \chi \rfloor < 1, \end{cases} \tag{6.74}$$

and the initial data

$$\begin{aligned} p(x,0) &= 1 + \frac{3}{20}\exp(-x^2/25) \\ \rho(x,0) &= p(x,0)^{1/\gamma}\hat{\rho}(\chi(x)) \\ u(x,0) &= 0, \end{aligned} \tag{6.75}$$

where $x = \int \rho(\chi)d\chi$ is the material coordinate, we have a initial density that varies periodically. We can see in Figure 6.7 the emergence of a series of solitary waves, reminiscent of the behaviour of dispersive nonlinear wave, differently from the what happens in for example in Figure 6.8 where $\hat{\rho}(x) = 1/2$ so the disturbance propagates through a constant entropy field.

Essentially, the behaviour of the solution of this system depends on the background density in the same way the behaviour of 1D and 2D shallow water systems depended on the bathymetry: if those are periodic, then the solution will be a KdV-type solution, with a formation of a series of solitary wave; on the other hand, if those quantities are not periodic, then the solution will have the typical behaviour of solutions of hyperbolic systems of conservation laws described in Chapter 2. Makes sense to study this model as the previous ones.

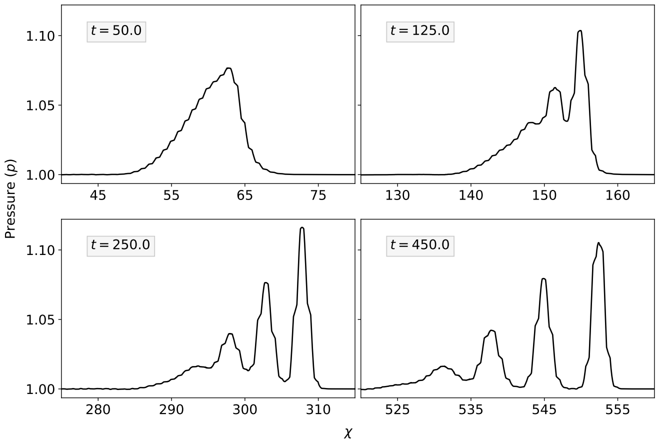


Figure 6.7: Solution of the Euler equations for a disturbance propagating through with an periodic initial density. This figure is reproduced from [9] with the permission of the authors.

We can rewrite the equations (6.71)-(6.72)-(6.73) in Lagrangian form by

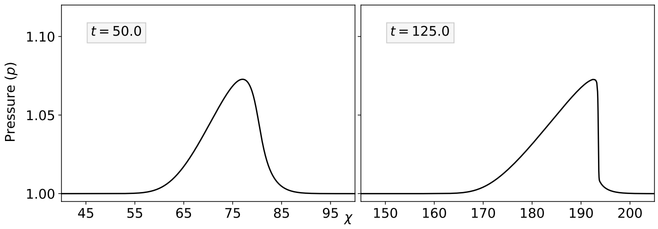


Figure 6.8: Solution of the Euler equations for a disturbance propagating through with an constant initial density. This figure is reproduced from [9] with the permission of the authors.

adopting the mass coordinate

$$x = \int_{\chi_0}^{\chi} \rho(\tilde{\chi}, t) d\tilde{\chi},$$

obtaining the system written in the new form

$$v_t + u_x = 0 \tag{6.76}$$

$$u_t + p_x = 0 \tag{6.77}$$

$$\Big(\frac{1}{2}u^2 + e\Big)_t + (up)_x = 0, \tag{6.78}$$

where $v = 1/\rho$ is the specific volume. In Lagrangian form, the equation (2.101) can be seen as

$$\frac{dS}{dt} = 0,$$

so the entropy density $s = S/cv$ does not depend on time. This condition can replace the energy equation. The entropy density for a polytropic gas is given by

$$s = \log(pv^\gamma) + const, \tag{6.79}$$

which can be written as

$$s - s_* = \log\Big(\frac{p}{p_*}\Big(\frac{v}{v_*}\Big)^\gamma\Big), \tag{6.80}$$

where p_*, v_* denote a particular reference state and s_* the corresponding entropy density, that without loss of generality we will suppose to be equal to zero. Solving for p, this functional relation becomes

$$p = p_* e^{s-s_*}(v_*/v)^\gamma, \tag{6.81}$$

where $s = s(x)$ is determined by considering an initial condition which is a perturbation of a stationary state

$$u_0 = 0, p_0 = p_*, v_0 = v_0(x),$$

so that

$$e^s = (v_0(x)/v_*)^\gamma, \quad s = \gamma \log(v_0(x)/v_*).$$

The system (6.76)-(6.77)-(6.78) can be now reduced two a 2 × 2 system

$$v_t - u_x = 0, \tag{6.82}$$

$$u_t + p(v, s(x))_x = 0. \tag{6.83}$$

We consider an unperturbed situation in which the initial density $v_0(x)$ is a periodic function of x, with period δ; since the pressure p is expected to be much less oscillatory than v we prefer to use p in place of v as dependent variable. We reformulate the problem as

$$\left(\frac{\partial v}{\partial p}\right)_{x=const} p_t - u_x = 0, \tag{6.84}$$

$$u_t + p_x = 0, \tag{6.85}$$

where the coefficient is given by the following relations

$$\left(\frac{\partial v}{\partial p}\right)^{-1}_{x=const} = \left(\frac{\partial p}{\partial v}\right)_{x=const} = \left(\frac{\partial p}{\partial v}\right)_{s=const} = -\frac{\gamma p}{v} = -c^2$$

where c^2 denotes the sound speed in Lagrangian coordinates, related to the Eulerian sound spedd c_E by the relation $c_E = vc$. Finally, the 2 × 2 system can be written as

$$p_t + c^2 u_x = 0, \tag{6.86}$$

$$u_t + p_x = 0, \tag{6.87}$$

where

$$c^2 = \frac{\gamma p}{v} = c_*^2 \Big(\frac{p}{p_*}\Big)^{1+1/\gamma} K(x), \quad K(x) = e^{-s(x)/\gamma}.$$

This system is usually known as *p-system*, and can be seen as the Lagrangian version of the Euler system. The p-system (6.86)-(6.87) is a special case of the model

$$\sigma_t - K(x)G(\sigma)u_x = 0 \tag{6.88}$$

$$\rho(x)u_t - \sigma_x = 0 \tag{6.89}$$

studied in [16] by R.LeVeque and D.H.Yong, where

$$\rho(x) \leftrightarrow 1 \tag{6.90}$$

$$u(x,t) \leftrightarrow u(x,t) \tag{6.91}$$

$$\sigma(x,t) \leftrightarrow -p(x,t) \tag{6.92}$$

$$K(x) \leftrightarrow e^{-s(x)/\gamma} := K(x) \tag{6.93}$$

$$G(\sigma) \leftrightarrow c_*^2 \Big(\frac{p}{p_*}\Big)^{1+1/\gamma} =: G(p) \tag{6.94}$$

Again, we proceed as [16] conducting the multiple-scale analysis. To do so, we introduce a fast spatial variable $y = x/\delta$, formally independent of x, and the partial derivative with respect to x becomes

$$\frac{\partial}{\partial x} \longrightarrow \frac{\partial}{\partial x} + \delta^{-1}\frac{\partial}{\partial y}.$$

Moreover, we suppose there exist a power series for p, u in terms of δ, expanding in power series $G(p)$ too:

$$p(x,y,t) = p^0(x,t) + \delta p^1(x,y,t) + \delta^2 p^2(x,y,t) + \dots \tag{6.95}$$

$$u(x,y,t) = u^0(x,t) + \delta u^1(x,y,t) + \delta^2 u^2(x,y,t) + \dots \tag{6.96}$$

$$\begin{aligned} G(p) &= G(p^0) + G'(p^0)(p - p^0) + \frac{1}{2}G''(p^0)(p-p^0)^2 + \dots \\ &= \frac{c_*^2}{p_*^{1+1/\gamma}}\Big((p^0)^{1+1/\gamma} + (1+1/\gamma)(p^0)^{1/\gamma} + \frac{1+1/\gamma}{2\gamma}(p^0)^{1/\gamma-1}(p-p^0)^2 + \dots\Big) \end{aligned} \tag{6.97}$$

Following the same reasoning as the previous sections and, in particular, the one in [16], we obtain

$$\overline{u}_t + \overline{p}_x = 0, \tag{6.98}$$

$$\begin{aligned}
&\overline{p}_t + \frac{G}{\langle K^{-1}\rangle}\overline{u}_x + \\
&\delta^2 \frac{\mu\langle K^{-1}\rangle}{G}\Big(\frac{4\overline{p}_t\overline{p}_{tt}G'}{G} + \overline{p}_t^3\Big(\frac{G''}{G} - \frac{3G'^2}{G^2}\Big) - \overline{p}_{ttt}\Big) + \\
&\delta^4 \frac{\zeta}{\langle K^{-1}\rangle}\Big(\alpha_1\overline{p}_{tttt}\overline{p}_t - \alpha_2\overline{p}_{tt}\overline{p}_{ttt} + \alpha_3\overline{p}_t\overline{p}_{tt}^2 + \\
&\alpha_4\overline{p}_t^2\overline{p}_{tt} + \alpha_5\overline{p}_t^3\overline{p}_{tt} + \alpha_6\overline{p}_t^5 + \alpha_7\overline{p}_{ttttt}\Big) = O(\delta^5),
\end{aligned} \tag{6.99}$$

where $G = G(\overline{p})$, the coefficients can be found in the Appendix B and the averaging operator is the same introduced previously.

System (6.98)-(6.99) contains higher order derivatives in time. This version of the system presents some inconveniences due to the presence of all these time derivatives. One problem, for example, is that this system is linear unstable, and the reason is precisely the presence of this type of derivatives, as in the case (6.11)-(6.12) in the 1D shallow water section. Taking inspiration from the case just mentioned, we convert almost all the $t-$derivatives, keeping one of them in the higher-order linear terms, obtaining

$$\overline{u}_t + \overline{p}_x = 0 \tag{6.100}$$

$$\begin{aligned}
\overline{p}_t + \frac{G(\overline{p})}{\langle K^{-1}\rangle}\overline{u}_x - \delta^2\mu\Big(\overline{p}_{xxt} + \frac{G'(\overline{p})}{\langle K^{-1}\rangle}\overline{p}_{xx}\overline{u}_x\Big) + \delta^4\Big(\frac{\zeta}{\langle K^{-1}\rangle^3} - \mu^2\Big)\overline{p}_{xxxxt} = \\
= \delta^4 N(\overline{p}, \overline{u}) + O(\delta^5),
\end{aligned} \tag{6.101}$$

where N is a function of $\overline{p}, \overline{u}$ where every term in nonlinear.

Let us study the linear stability of the system (6.100)-(6.101). Linearizing around the equilibrium configuration $(\overline{u}, \overline{p}) = (0, p_*)$ and neglecting terms of $O(\delta^5)$, we obtain the linearized system

$$\overline{u}_t + \overline{p}_x = 0 \tag{6.102}$$

$$\overline{p}_t + c^2\overline{u}_x - \delta^2\mu\overline{p}_{xxt} + \delta^4\nu\overline{p}_{xxxxt} = 0, \tag{6.103}$$

where

$$c^2 = \frac{G(p_*)}{\langle K^{-1}\rangle}, \quad \nu = \frac{\zeta}{\langle K^{-1}\rangle^3} - \mu^2.$$

We look for solutions of the linearized system of the form

$$\overline{u} = \hat{u}e^{i(kx-\omega t)},\ \overline{p} = \hat{p}e^{i(kx-\omega t)}.$$

Plugging this ansatz in the equations we obtain the following system for $\hat{u}, \hat{p}$:

$$-i\omega\hat{u} + ik\hat{p} = 0 \tag{6.104}$$

$$-i\omega\hat{p} + ic^2k\hat{u} - i\delta^2k^2\omega\mu\hat{p} - i\delta^4\nu k^4\omega\hat{p} = 0, \tag{6.105}$$

which has non-trivial solutions when the determinant of the coefficient matrix is zero. This yields to the dispersion relation

$$c^2k^2 - \omega^2(1 + \mu\delta^2k^2 + \nu\delta^4k^4) = 0,$$

that is solved by

$$\omega = \pm\frac{ck}{\sqrt{1 + \mu\delta^2k^2 + \nu\delta^4k^4}},$$

that means that, provided $\nu > 0$, the system admits linearly dispersive waves for all wave number $k \in \mathbb{R}$. Let us now, as done for the 1D and 2D shallow water systems, compare the solutions obtained by solving the initial Euler system (6.71)-(6.72)-(6.73), the p-system (6.86)-(6.87) and the homogenized system. The Euler equations and the p-system have been solved using a method based on a 5-th order WENO reconstruction and SSPRK of order 4 time integration, while a Fourier pseudospectral collocation method has been used for the computation of the solution of the homogenized equations. The bathymetry and the initial condition used are the ones in (6.75). The results are shown in Figure 6.9: it is possible to observe that the solutions of Euler equations and p-system are indistinguishable, while the homogenized solution shows close agreement at early times, with increasingly noticeable differences at later times.

6.4.1 Travelling wave to $O(\delta^2)$

Let us now try to compute the travelling wave for the system (6.100)-(6.101), looking for solutions that depends on $\xi = x - Vt$. To simplify the notation, we omit the overline bar. We start considering the case where we neglect terms of $O(\delta^4)$. The system of ODEs we obtain inserting the ansatz we

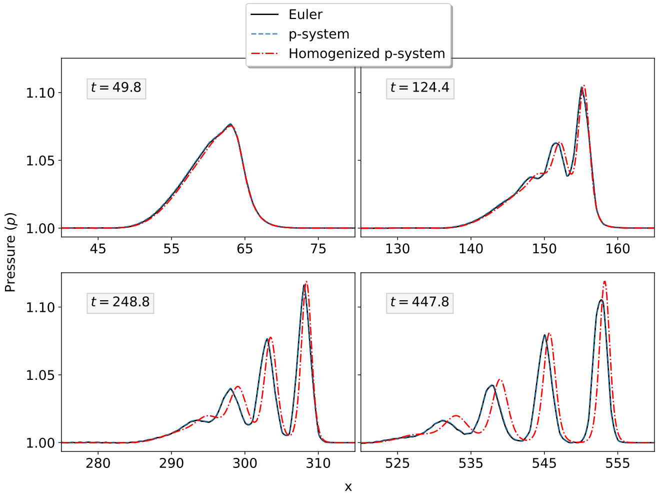


Figure 6.9: Comparison between the direct solution of the Euler equations, the p-system and the homogenized equations. This figure is reproduced from [9] with the permission of the authors.

desire is given by

$$-Vp' + \frac{G(p)}{\langle K^{-1}\rangle}u' - \delta^2\mu\Big(Vp''' + \frac{G'(p)}{\langle K^{-1}\rangle}p''u'\Big) = 0 \tag{6.106}$$

$$-Vu' + p' = 0. \tag{6.107}$$

From 6.107 we deduce $p = p_* + Vu$, since we consider travelling waves propagating on a unperturbed state given by $p = p_*, u = 0$. Inserting the dependence from u of p we obtain the following ODEs:

$$-V^2u' + \frac{G(p(u))}{\langle K^{-1}\rangle}u' - \delta^2\mu\Big(- V^2u''' + \frac{G'(p(u))}{\langle K^{-1}\rangle}Vu''u'\Big) = 0. \tag{6.108}$$

This is a third-order equation for u, and unfortunately this can not be written in general as a total derivative, that is what we desire in order to proceed as in the previous cases, even if we make explicit the form of G. We need to accept the need of some kind of approximation. Since the only "problem" for the integrability of this equation is the term G', let us approximate this term with $G'(p_*)$:

$$-V^2u' + \frac{G(p(u))}{\langle K^{-1}\rangle}u' - \delta^2\mu\Big(- V^2u''' + \frac{G'(p_*)}{\langle K^{-1}\rangle}Vu''u'\Big) = 0. \tag{6.109}$$

Let $\mathcal{G}$ be a primitive of G, we have that $d\mathcal{G}/d\xi = G(p(u))p' = G(p(u))Vu$, so we can write the last equation in total derivative form as

$$\frac{d}{d\xi}\Big[-V^2u + \frac{\mathcal{G}(p(u))}{V\langle K^{-1}\rangle} + \delta^2V^2\mu u'' - \delta^2\mu\frac{G'(p_*)}{2\langle K^{-1}\rangle}V(u')^2\Big] = 0, \tag{6.110}$$

which means, integrating with respect to ξ,

$$V^2u + \frac{\mathcal{G}(p(u)) - \mathcal{G}(p_*)}{V\langle K^{-1}\rangle} + \delta^2V^2\mu u'' - \delta^2\mu\frac{G'(p_*)}{2\langle K^{-1}\rangle}V(u')^2 = C. \tag{6.111}$$

Since $u(\xi)$ vanishes at infinity, we have $C = 0$ and the second order equation can be written in normal form as

$$u'' = \frac{G'(p_*)}{2V\langle K^{-1}\rangle}(u')^2 - \frac{\mathcal{G}(p(u)) - \mathcal{G}(p_*)}{\delta^2\mu V^3\langle K^{-1}\rangle} + \frac{u}{\delta^2\mu}. \tag{6.112}$$

This second order equation can be written as a first order system of the form

$$u' = v, \quad v' = F(u, v), \tag{6.113}$$

which has $(0, 0)$ as equilibrium point and which linearization around this point reads

$$u' = v, v' = \beta u, \quad \text{with } \beta = \Big(1 - \frac{G(p_*)}{V^2\langle K^{-1}\rangle}\Big)(\delta^2\mu)^{-1}.$$

When $\beta > 0$, there are two real roots $\lambda_\pm = \pm\sqrt{\beta}$, and the origin is a saddle point. From this, one can find a good approximation of a travelling wave integrating the system (6.113) with linear conditions really close to the origin and aligned with one of the eigenvectors. The result is shown in Figure 6.10.

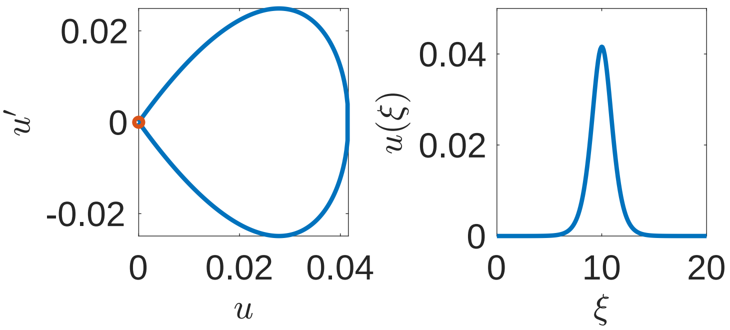


Figure 6.10: Computation of the travelling wave of the second order homogenized Euler equations. This figure is reproduced from [9] with the permission of the authors.

However, this same approach is not useful for the computation of the travelling wave for fourth order equation that derives from (6.100)-(6.101). This is due the fact that, following this reasoning, one should solve a first-order system of four equations, finding again, imposing some conditions, that $(0, 0, 0, 0)$ is a saddle point, but having some issues on the imposition of the initial condition aligned with the eigenvectors. We will see a new way to find an approximate travelling wave for this equation in the next chapter.

Chapter 7

Alternative way for the computation of travelling waves

7.1 Introduction

In this chapter, we will try to solve the problem of the (numerical) computation of travelling waves when the ODEs we have to solve do not have the properties needed for the application of the method described in the section dedicated to this topic. This problem arises from the previous chapter, when we tried to compute the travelling wave for the homogenized Euler system of the fifth-order: while for the third-order one we can integrate, obtaining a second order equation, and then apply the phase plane analysis, for the fifth order system this approach does not work since the equation we obtain is of order four and we should impose four initial condition aligned with the eigenvectors of the system matrix around the point with respect we are linearizing, that is way more difficult to do than the second order case.

We will introduce class of fixed point algorithms called *Petviashvili's methods*, first proposed by Petviashvili in 1976 and then improved in the last almost fifty years. These Petviashvili-type methods often converge fast and they are easy to implement. However, they can only converge to the ground states of nonlinear wave equations and would diverge for excited states. For our purpose, we will use a family of Petviashvili type methods that has been defined and which convergence has been studied in [2].

Then, we will show the results of the application of this method on a classical example (KdV).

In the end, we will finally use this method to solve the problem of the computation of the fifth-order travelling wave for the homogenized Euler equation, comparing the results obtained to the second order travelling wave and the direct solution of the initial system. The results of this chapter are original discoveries by the writer.

7.2 Petviashivili type methods

In this section, we introduce a class of fixed point algorithms for the numerical resolution of non linear systems of the form

$$Lu = N(u), \quad u \in \mathbb{R}^m, m > 1 \tag{7.1}$$

where L is a nonsingular $m \times m$ real matrix and $N\colon \mathbb{R}^m \to \mathbb{R}^m$ is an homogeneus function of the components of u with degree p, $|p| > 1$. This kind of systems arises in many applications, and we will see that one of them is the numerical generation of travelling waves in nonlinear dispersive systems. Before seeing the actual class of methods we are going to use, let us see how the original Petviashvili method works. Denoted by u^* a solution of (7.1), that is

$$Lu^* = N(u^*),$$

the classical fixed point algorithm is given, at the $(n+1)$th iteration, by the recurrence

$$Lu_{n+1} = N(u_n), \quad n = 0, 1, \ldots, \tag{7.2}$$

where $u_0 \neq 0$ is a initial value. This method is usually not convergent, and usually when convergent it goes to zero. In order to solve this problem, what is usually done is to introduce a stabilizing factor in the iteration in order to pump up the result of the iteration when it decays and to suppress it when it grows. The stabilizing factor introduced by Petviashvili in [21] and the consequent iterative method is the following

$$Lu_{n+1} = m(u_n)^\gamma N(u_n), \quad m(u_n) = \frac{\langle Lu_n, u_n \rangle}{\langle N(u_n), u_n \rangle}, \quad \gamma \in \mathbb{R}, \tag{7.3}$$

where $\langle \cdot, \cdot \rangle$ is the usual Euclidean inner product. The method introduced in [2] and used here is slightly different from (7.3) and it is of the form

$$Lu_{n+1} = s(u_n)N(u_n), \quad n = 0, 1, \dots \tag{7.4}$$

where $s \colon \mathbb{R}^m \to \mathbb{R}$ is a C^1 function satisfying the following properties:

- A set of fixed points of the iteration operator

$$F(u) = s(u)L^{-1}N(u), \tag{7.5}$$

  coincides with a set of fixed points of (7.1). This means that:

  - if u^* is a solution of (7.1) then $s(u^*) = 1$;
  - inversely, if the sequence generated by (7.4) u_n converges to y then $s(y) = 1$.

- s is homogeneous with degree such that $|p + q| < 1$.

In [2] different examples of s have been shown. In this work, the s that has been used is the following

$$s_r(u) = \Big(\frac{||Lu||_r}{||N(u)||_r}\Big)^\gamma, \quad q = \gamma(1-p), \quad |p+q| < 1, \tag{7.6}$$

where $|| \cdot ||_r$ is the usual r-norm with $1 \leq r \leq \infty$. We can see s as a generalization of the stabilizing factor of the original method (7.3).

7.2.1 Analysis of the convergence

Now, we introduce some theoretical result about the convergence of the method (7.4). We will just give the results without proving them. In order to see the proof, see [2]. Let us consider the Jacobian of the iteration operator (7.5) at the fixed point u^* solution of $Lu = N(u)$ given by

$$F'(u^*) = S + u^*(\nabla s(u^*)), \tag{7.7}$$

where $S = L^{-1}N'(u^*)$ is the iteration matrix at u^* of the recurrence (7.2) (where $N'(u)$ denotes the Jacobian on N at u). The first purpose is to relate the spectrum of (7.7) with that of S. The first result in this sense is the following

Lemma 7.1. *Let u^* be a solution of* (7.1), *$s\colon \mathbb{R}^m \to \mathbb{R}$ a C^1 function satisfying the two conditions required by the method for s. Then the matrix*

$$P(u^*) = u^*(\nabla s(u^*))$$

satisfies

$$P(u^*)^2 = qP(u^*).$$

Its spectrum consists of the eigenvalues $\lambda = q$, which is simple and its eigenspace is spanned by u^*, and $\lambda = 0$.

Lemma 7.1 means that $F'(u^*) - S = P(u^*)$ is like a nonorthogonal projection. Its structure allows to relate the spectrum of $F'(u^*)$ and S.

Lemma 7.2. *Under the hypotesis of Lemma 7.1 and assumed that p is a simple eigenvalue of S, we have:*

- *$\lambda^* = p + q$ is an eigenvalue of $F'(u^*)$ and u^* is the associated eigenvector;*
- *$\lambda = p$ can not be an eigenvalue of $F'(u)$;*
- *the part of the spectrum of S different from p is the spectrum of $F'(u^*)$.*

Consequently, the spectrum of $F'(u^*)$ consists of the spectrum of S except the eigenvalue p, which is substituted by $\lambda^* = p + q$. Note that λ may appear in the spectrum of S initially, independently of its formation as substitute of p; in that case, this implies that λ^* would not be simple as eigenvalue of $F'(u^*)$. The properties described by this last lemma imply that the study of convergence for the family of methods (7.4) can be oriented by the part of the spectrum of S different from p. For example, the Contraction Mapping Theorem can be used to obtain the following classical local convergence result:

Theorem 7.3. *Assume that*

1. *there exists $R > 0$ such that u^* is the unique fixed point of* (7.1) *in $B(u^*, R) = \{u \in \mathbb{R}^m \colon ||u - u^*|| < R\}$;*
2. *p, as eigenvalue of S, is simple;*
3. *the rest of the eigenvalues of λ of S satisfies $|\lambda| < 1$.*

Then if $u_0 \in B(u^*, R), u_0 \neq 0$, the sequence $\{u_n\}_{n\in\mathbb{N}}$ generated by (7.4) *converges to u^*.*

7.3 Application on models: the Kdv equation

Let us now see how we can apply this new method to equations or systems we saw in the previous chapters. Let us start with an easy case, of which we know almost everything and that is easy to work on, the KdV equation. Recall that the equation is

$$u_t + 6uu_x + u_{xxx} = 0.$$

We saw in 4.4.1 that the ODE for the travelling wave of the KdV is

$$-vu' + 6uu' + u''' = 0, \tag{7.8}$$

that integrated (and imposed that u vanishes at infinity) becomes

$$-vu + 3u^2 + u'' = 0. \tag{7.9}$$

Since the method (7.4) is a method for algebraic systems, we need to transform this equation in an algebraic system. For simplicity, we will suppose that u is defined in an interval $[a, b]$ and we will impose periodic boundary conditions. First of all, we need to discretize u, in order to see it as a vector $u = (u_1, \dots, u_n)$. Now, writing the equation in the form

$$3u^2 = vu - u'',$$

we observe that the left-hand side is an homogeneous term, so it will be our $N(u)$ in the method. Finally, we need to rewrite the right-hand side in order to have a discrete form of it. What we do is to use finite difference to discretize the differential operator $v - \frac{d^2}{d\xi^2}$, obtaining a differentiation matrix $vI - D_2$, where I is the identity matrix and D_2 is the matrix that approximate the second derivatives through finite difference given by

$$D_2 = \frac{1}{\Delta x^2}\begin{bmatrix} -2 & 1 & 0 & \cdots & 0 & 1 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & 0 & 1 & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & -2 & 1 \\ 1 & 0 & \cdots & 0 & 1 & -2 \end{bmatrix}. \tag{7.10}$$

The final matrix L is now given by

$$L = vI - \frac{1}{\Delta x^2}D =$$

$$v\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & 0 & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 1 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix} - \frac{1}{\Delta x^2}\begin{bmatrix} -2 & 1 & 0 & \cdots & 0 & 1 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & 0 & 1 & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & -2 & 1 \\ 1 & 0 & \cdots & 0 & 1 & -2 \end{bmatrix}. \tag{7.11}$$

Now the equation is in the form

$$Lu = N(u)$$

required by the method (7.4). In Figure 7.1 we can see the travelling wave for the KdV equation computed using the method just described. In Appendix C the MATLAB code used for the computation is shown.

7.4 Application on models: the Homogenized Euler System

Finally, in this section we describe how to compute numerically the travelling wave for the system (6.100)-(6.101).

7.4.1 Travelling waves to $O(\delta^2)$

First, we consider the approximate system in which we neglect terms of $O(\delta^4)$. We look for solutions which depends only on the variable $\xi = x - Vt$,

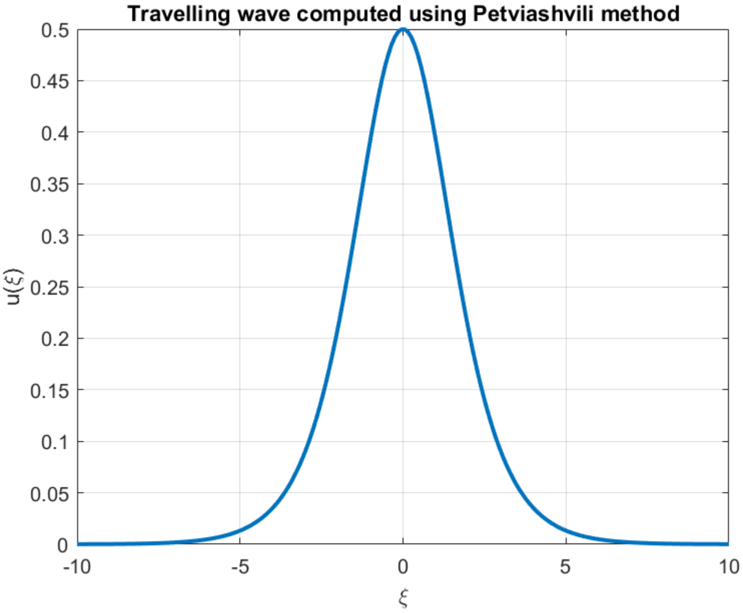


Figure 7.1: Travelling wave computed using the Petviashvili method

where V is the travelling speed of the wave, i.e. we look for the solutions of the form $p = p(\xi)$ and $u = u(\xi)$, and, to simplify the notation, we omit the overbar on top of the variables. From now, we will assume u defined in an interval $[a, b]$ and periodic, assumption already done in the KdV case. Inserting the ansatz for u and p in system (21), we obtain the system of ODEs

$$\begin{aligned} & -Vp' + \frac{G(p)}{\langle K^{-1}\rangle}u' - \delta^2\mu(-Vp''' + \frac{G'(p)}{\langle K^{-1}\rangle}p''u') = 0 && (7.12) \\ & -Vu' + p' = 0 && (7.13) \end{aligned}$$

From the second equation we deduce that $p' = Vu'$, and therefore $p = p_* + Vu$, since we consider propagation of travelling waves on an unperturbated state given by $p = p_*$ and $u = 0$. Inserting the dependence of p from u, we obtain the following third order ODE for $u'(\xi)$:

$$-V^2u' + \frac{G(p(u))}{\langle K^{-1}\rangle}u' - \delta^2\mu\Big(- V^2u''' + \frac{G'(p(u))}{\langle K^{-1}\rangle}Vu''u'\Big) = 0. \qquad (7.14)$$

In general, the left hand side of the resulting expression can not be written as a total derivative. We need to proceed as done in section 6.4.1, trying to approximate G and/or G' in order to obtain an easier equation. What we do in this case is to use the definition of G, for relating G and G' and the approximate the value of $G'(p)$. First of all, we recall that

$$G(p) = \frac{c_*^2}{p_*^{1+\frac{1}{\gamma}}} p^{1+\frac{1}{\gamma}},$$

so we have that

$$G'(p) = \frac{c_*^2}{p_*^{1+\frac{1}{\gamma}}}(1+\frac{1}{\gamma})p^{\frac{1}{\gamma}} = \frac{G(p)}{p}(1+\frac{1}{\gamma}) \Longrightarrow G(p) = G'(p)p\frac{1}{(1+\frac{1}{\gamma})}.$$

Consequently, substituting this new form of $G(p)$ in the equation and replacing p with $p_* + Vu$ we obtain

$$\begin{aligned}&-V^2u' + \frac{G'(p(u))p_*}{\langle K^{-1}\rangle(1+\frac{1}{\gamma})}u' + \frac{G'(p(u))Vu}{\langle K^{-1}\rangle(1+\frac{1}{\gamma})}u' + \\ &-\delta^2\mu(-V^2u''' + \frac{G'(p(u))}{\langle K^{-1}\rangle}Vu''u') = 0.\end{aligned}$$

In general, the left hand side of the resulting expression can not be written as a total derivative. However, if we approximate $G'(p)$ by $G'(p_*)$, we can integrate the equation and reduce the problem to a second order ODE, which is more amenable to treat. The equation for u becomes

$$\begin{aligned}&-V^2u' + \frac{G'(p_*)p_*}{\langle K^{-1}\rangle(1+\frac{1}{\gamma})}u' + \frac{G'(p_*)Vu}{\langle K^{-1}\rangle(1+\frac{1}{\gamma})}u' - \\ &\delta^2\mu(-V^2u''' + \frac{G'(p_*)}{\langle K^{-1}\rangle}Vu''u') = 0,\end{aligned} \tag{7.15}$$

that can be written as

$$\begin{aligned}&\frac{d}{d\xi}\Big[\Big(-V^2 + \frac{G'(p_*)p_*}{\langle K^{-1}\rangle(1+\frac{1}{\gamma})}\Big)u + \frac{G'(p_*)V}{\langle K^{-1}\rangle(1+\frac{1}{\gamma})}\frac{u^2}{2} + \\ &+ V^2\mu\delta^2u'' - \delta^2\mu\frac{G'(p_*)V}{\langle K^{-1}\rangle}\frac{(u')^2}{2}\Big] = 0,\end{aligned}$$

that indicates that the quantity in square brackets is constant. Given that at infinity $u(\xi)$ vanishes with its derivatives, integrating we obtain the second order equation for u, which can be written in normal form as

$$\Big(\frac{1}{\mu\delta^2}-\frac{G'(p_*)p_*}{\langle K^{-1}\rangle(1+\frac{1}{\gamma})V^2\mu\delta^2}\Big)u-\frac{G'(p_*)}{\langle K^{-1}\rangle(1+\frac{1}{\gamma})V\mu\delta^2}\frac{u^2}{2}+\frac{G'(p_*)}{\langle K^{-1}\rangle V}\frac{(u')^2}{2}=u''$$

The second order equation can be written as a first order system of the form

$$u'=v, v'=F(u,v).$$

Since (0,0) is an equilibrium point for this system, linearizing the system around the origin we obtan

$$u'=v, v'=\beta u$$

with

$$\beta=\Big(\frac{1}{\mu\delta^2}-\frac{G'(p_*)p_*}{\langle K^{-1}\rangle(1+\frac{1}{\gamma})V^2\mu\delta^2}\Big)=\frac{1}{\delta^2\mu}\left(1-\frac{G(p_*)}{V^2\langle K^{-1}\rangle}\right),$$

where we used in the last equation the fact that of

$$G(p_*)=G'(p_*)p_*\frac{1}{(1+\frac{1}{\gamma})}.$$

When $\beta>0$, there are two real roots, $\lambda=\pm\sqrt{\beta}$, and the origin is a saddle point. Integrating the system with an initial condition very close to the origin, aligned with the eigenvector corresponding to the positive eigenvalue, one obtains a good approximation of a travelling wave. We can observe that the eigenvalues found are the same of the one founded in section 6.4.1. This means that if we apply the phase plane analysis to this system, the solution we will find will be very similar to one found previously.

Let us now try to use the Petviashvili method (7.4) introduced earlier to this equation. First of all, we need the equation to be written in the "form" (7.1), where L has to interpreted now has a differential operator: indeed, if this happens, then we can, as done in the KdV case, discretize over the interval, so that u becomes a vector $(u_1,\ldots,u_n)$, transform the differential operator into a differentiation matrix and now finally apply the method (7.4). Unfortunately, the equation 7.4.1 does not have properly the form required by the method described by Petviashvili, but it has a really similar form:

indeed, we can write 7.4.1 as

$$Lu = N(u, u'),$$

where

$$L = -\beta + \frac{d^2}{d\xi^2}$$
$$N(u, u') = -\frac{G'(p_*)}{\langle K^{-1}\rangle(1+\frac{1}{\gamma})V\mu\delta^2}\frac{u^2}{2} + \frac{G'(p_*)}{\langle K^{-1}\rangle V}\frac{(u')^2}{2}.$$

It is obvious that N is homogeneous of degree 2. In order to obtain an homogeneous function of u alone, we approximate u' using the central difference, i.e.

$$u'(x) \approx \frac{u(x+h) - u(x-h)}{2h}, \quad h \in \mathbb{R},$$

so that we can finally write

$$N(u, u') \approx N(u, \frac{u(x+h) - u(x-h)}{2h}) := \tilde{N}(u),$$

and since $\tilde{N}(u)$ is homogeneous of degree 2, after a spatial discretization we obtain a system of the type

$$\tilde{L}u = \tilde{N}(u),$$

where

$$\tilde{L} = -\beta I + D_2 =$$
$$-\beta \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & 0 & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 1 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix} + \frac{1}{\Delta x^2} \begin{bmatrix} -2 & 1 & 0 & \cdots & 0 & 1 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & 0 & 1 & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & -2 & 1 \\ 1 & 0 & \cdots & 0 & 1 & -2 \end{bmatrix} \tag{7.16}$$

is the non singular matrix required by the method, and in particular it is the differentiation matrix that approximates the differential operator $L = -\beta + \frac{d^2}{d\xi^2}$, and $u = (u_1, \ldots, u_n)$ is a discretization of u. On this system we can finally apply a Petviashvili type method (7.4) in order to generate the travelling wave.

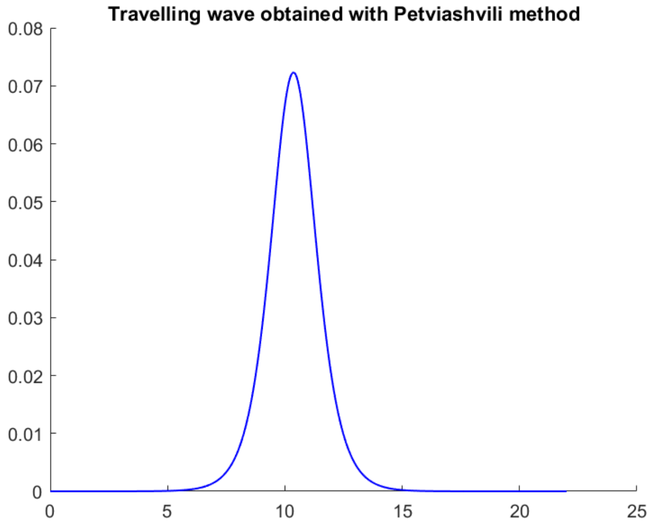


Figure 7.2: Travelling wave obtained through Petviashvili method for the second order equation

In figure 7.2, the travelling wave constructed using this new method is shown. In Appendix C the MATLAB code used to develop this travelling wave is shown and explained, here we just say that the initial condition we gave is the travelling wave shown in 6.10. With this choice, the convergence is very rapid, but doing experiments with the code it is possible to obtain the convergence to the solution shown in the figure even with very bad initial conditions in a relatively small amount of time.

In Figure 7.3, it is possible to see the plot of both the solution obtained with a classic approach and the one obtained with the new method. The fact that there is no difference to the naked eye is because the two curves are practically the same: indeed, in Figure 7.4, the difference between the two solutions is shown, and the order of this difference is 10^{-6} at its maximum value. This comparison confirms that the method works and that it is possible to build numerical solutions that are admissible proceeding in this way.

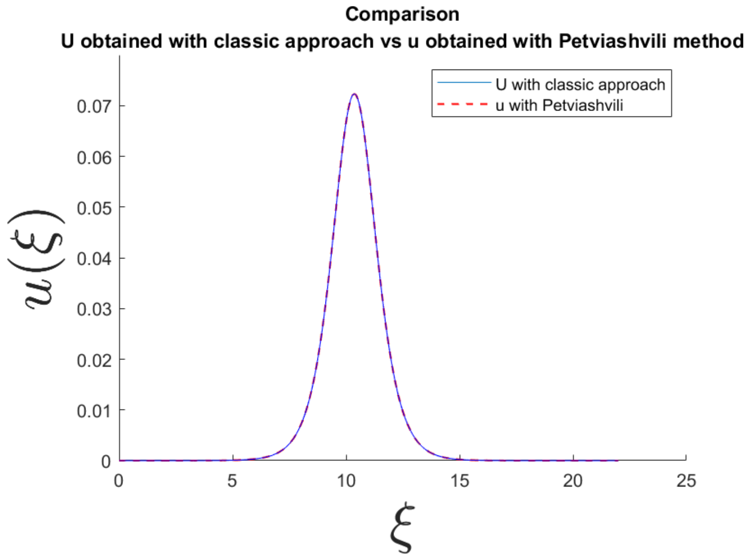


Figure 7.3: Comparison between the travelling wave in Figure 6.10 and the one obtained using the Petviashvili method applied on the second order equation

7.4.2 Travelling waves to $O(\delta^4)$

We now look for travelling waves for the system of order five (6.100)-(6.101), keeping the term p_{xxxxt} and neglecting the nonlinear terms multiplied by δ^4. Differently from the third-order case, where we were able to find the travelling wave even without using a Petviashvili type method, in this case the application of this method will allow us to find a solution of a problem that we were not able to solve with the previous approach. Let ν be the coefficient of p_{xxxxt}, looking for solutions of the type $p(\xi), u(\xi)$, with $\xi = x - Vt$, we obtain the system of ODEs given by

$$-Vp' + \frac{G(p)}{\langle K^{-1}\rangle}u' - \delta^2\mu(-Vp''' + \frac{G'(p)}{\langle K^{-1}\rangle}p''u') - V\nu p^{(5)} = 0 \tag{7.17}$$

$$-Vu' + p' = 0. \tag{7.18}$$

As before, from the second equation we have $p = p_* + Vu$, since again we are considering propagation of travelling waves on an unperturbated state given by $p = p_*, u = 0$. Substituting it in the first equation we obtain the

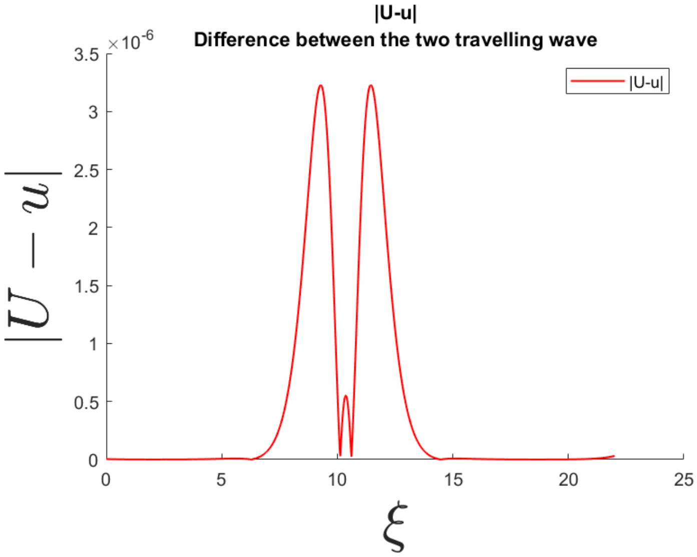


Figure 7.4: Computation of the absolute value of the difference between the two travelling waves obtained with different methods in the second order equation

following ODE for u:

$$-V^2u' + \frac{G(p(u))}{\langle K^{-1}\rangle}u' - \delta^2\mu\Big(-V^2u''' + \frac{G'(p(u))}{\langle K^{-1}\rangle}Vu''u'\Big) - V^2\nu u^{(5)} = 0. \tag{7.19}$$

Again, this equation is generally non expressible as a total derivate, so proceeding as section 7.4.1, we use the expression of G in order to write it in term of G' and then approximate $G'(p)$ with $G'(p_*)$, obtaining in the end the following ODE in normal form:

$$u^{(4)} = (\frac{1}{\nu} + \frac{G(p_*)}{\langle K^{-1}})u + \frac{G'(p_*)}{\langle K^{-1}\rangle(1+\frac{1}{\gamma})V\nu}\frac{1}{2}u^2 - \frac{\delta^2\mu}{\nu}u'' + \frac{G'(p_*)}{\langle K^{-1}\rangle V\nu}\frac{1}{2}(u')^2. \tag{7.20}$$

We rewrite (7.20) as

$$u^{(4)} - (\frac{1}{\nu} + \frac{G(p_*)}{\langle K^{-1}\rangle})u + \frac{\delta^2\mu}{\nu}u'' = \frac{G'(p_*)}{\langle K^{-1}\rangle(1+\frac{1}{\gamma})V\nu}\frac{1}{2}u^2 + \frac{G'(p_*)}{\langle K^{-1}\rangle V\nu}\frac{1}{2}(u')^2, \tag{7.21}$$

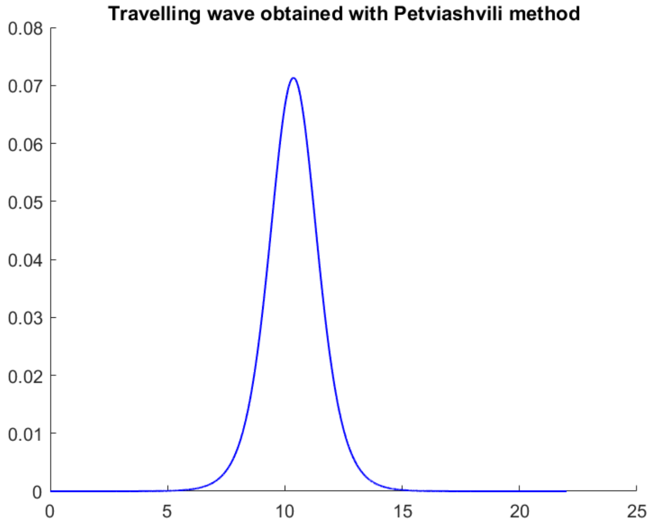


Figure 7.5: Travelling wave obtained through Petviashvili method for the fourth order equation

in order to visualize it in the form required by the Petviashvli method (7.4), even if this is not already the form we seek. Indeed, we know that the form required by the method is

$$Lu = N(u),$$

where, in the continuous framework, L is a differential operator and N is an homogeneous function, but instead our equation is in the form, as in the previous case,

$$Lu = N(u, u').$$

Proceeding as before, we first approximate u' using central difference, and after that we finally obtain an equation of the form

$$Lu = \tilde{N}(u)$$

with $\tilde{N}$ homogeneous function of degree two. Now we can finally apply (7.4), discretizing over the interval and construction the differentiation matrix for

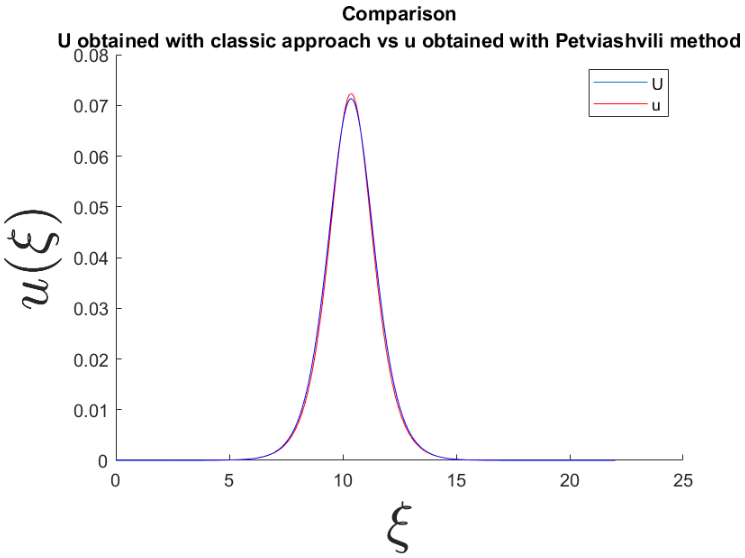


Figure 7.6: Comparison between the travelling wave in Figure 6.10 and the one obtained using the Petviashvili method applied on the fourth order equation

this equation as

$$\tilde{L} = \frac{\delta^2\mu}{\nu} I - \left(\frac{1}{\nu} + \frac{G(p_*)}{\langle K^{-1}\rangle}\right) D_2 + D_4 =$$

$$= \frac{\delta^2\mu}{\nu}\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix} - \frac{1}{\Delta x^2}\alpha \begin{bmatrix} -2 & 1 & 0 & \cdots & 0 & 1 \\ 1 & -2 & 1 & \cdots & 0 & 0 \\ 0 & 1 & -2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -2 & 1 \\ 1 & 0 & 0 & \cdots & 1 & -2 \end{bmatrix} +$$

$$+\frac{1}{\Delta x^4}\begin{bmatrix} 6 & -4 & 1 & 0 & \cdots & 0 & 1 & -4 \\ -4 & 6 & -4 & 1 & \cdots & 0 & 0 & 1 \\ 1 & -4 & 6 & -4 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -4 & 6 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 6 & -4 & 1 \\ 1 & 0 & 0 & 0 & \cdots & -4 & 6 & -4 \\ -4 & 1 & 0 & 0 & \cdots & 1 & -4 & 6 \end{bmatrix}, \tag{7.22}$$

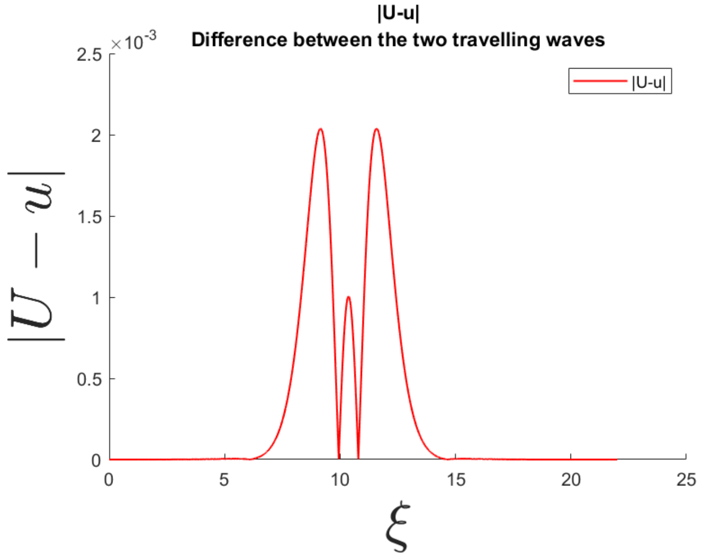


Figure 7.7: Computation of the absolute value of the difference between the travelling wave in Figure 6.10 and the one obtained using the Petviashvili method applied on the fourth order equation

where $\alpha = 1/\nu + G(p_*)/\langle K^{-1}\rangle$, I is the identity matrix, D_2 is the differentiation matrix that allows us to approximate the second derivate and D_4 is the differentiation matrix for the derivative of order four. In Figure 7.5, the travelling wave for the fourth order equation built up using the method just described is shown. Again, the MATLAB code used for the computation of the travelling wave is shown in Appendix C, here we just underline that again the initial condition used is the "classic" travelling wave for to $O(\delta^2$, in order to have fast convergence, but one can do just some experiment with the code in order to see that the method will converge quickly to this solution even if a worse initial condition is inserted.In Figure 7.6, we can see the comparison between the travelling wave for the second order equation obtained with a classic approach and the fourth-order travelling wave computed using the Petviashvili method. In this case, a small difference can be noticed even with the naked eye: indeed, as shown in Figure 7.7, the order of the difference between these two solutions at their peak is 10^{-3}, that is reasonable since they are solutions of equations derived from two different homogenized systems, and one is a better approximation than the other.

Chapter 8

Conserved quantities for our models: discovery and numerical validation

8.1 Introduction

In this chapter, we focus on the conserved quantities for the homogenized models we derived in Chapter 6. As said in the dedicated section, conserved quantities play an important role in the study of differential equations, and since the models introduced in Chapter 6 has been derived recently and there is no study on this topic in the original papers, what we do in this chapter is to try to fill this gap investigating the possible presence of conserved quantities. Different approaches and techniques have been used in each system, in order to take advantage of the peculiarities of each model, and interesting results have been found. Moreover, each discovery is accompanied by a numerical validation, as a confirmation of the theoretical calculations and discoveries.

8.2 Conserved quantities for the homogenized 1D shallow water equations

Let us investigate the possible presence of conserved quantities in the homogenized shallow water equations. Since from the linear case we can have some hints for the conserved quantities in nonlinear case, we start with the linearized systems that we found in section 6.2.

8.2.1 Linear case up to $O(\delta^3)$ terms

Let us consider the first of the many type of third order equations that we can derive from the homogenization of the shallow water given by

$$\eta_t + q_x = 0 \tag{8.1}$$
$$q_t + c^2\eta_x - \frac{\hat{\mu}}{c^2}q_{ttt} = 0. \tag{8.2}$$

In order to find a conserved quantity, we procede in the following way: we multiply the first equation by $c^2\eta$, the second by q and we add the two new equations. In this way, we obtain the equation

$$c^2\eta\eta_t + qq_t + c^2\eta q_x + c^2\eta_x q + \hat{\mu}qq_{ttt} = 0. \tag{8.3}$$

We now prove that this equation can be written in the form

$$f(\eta, q)_t + g(\eta, q)_x = 0. \tag{8.4}$$

For this purpose, we show how to integrate the term qq_{ttt}: we have

$$qq_{ttt} = (qq_{tt})_t - q_t q_{tt} = (qq_{tt})_t - \frac{1}{2}(q_t^2)_t. \tag{8.5}$$

So in the end we obtain the energy conservation law

$$\Big(\frac{c^2}{2}\eta^2 + \frac{1}{2}q^2 - \hat{\mu}qq_{tt} + \frac{1}{2}q_t^2\Big)_t + c^2(\eta q)_x = 0. \tag{8.6}$$

The quantity

$$\frac{c^2}{2}\eta^2 + \frac{1}{2}q^2 - \hat{\mu}qq_{tt} + \frac{1}{2}q_t^2 \tag{8.7}$$

is a conserved quantity for the previous system. Now, let us consider the second version of this system, the one with the third order term with only x-derivatives

$$\eta_t + q_x = 0 \tag{8.8}$$
$$q_t + c^2\eta_x + \hat{\mu}c^2\eta_{xxx} = 0. \tag{8.9}$$

We procede similarly as the previous case, but in this case we multiply the first equation by q, the second by η and we add the results, obtaining

$$q\eta_t + \eta q_t + qq_x + c^2\eta\eta_x + \hat{\mu}c^2\eta\eta_{xxx} = 0. \tag{8.10}$$

Observing that

$$\eta\eta_{xxx} = (\eta\eta_{xx})_x - \eta_x\eta_{xx} = (\eta\eta_{xx}) - \frac{1}{2}(\eta_x)_x, \tag{8.11}$$

we find from the equation (8.10) the energy conservation law

$$(q\eta)_t + \Big(\frac{1}{2}q^2 + \frac{1}{2}c^2\eta^2 + \hat{\mu}c^2\eta\eta_{xx} - \frac{1}{2}\eta_x\Big)_x = 0. \tag{8.12}$$

In this case, the conserved quantity for this system is $q\eta$. Finally, we analyze the version of this system where the term q_{xxt} appears, i.e.

$$\eta_t + q_x = 0 \tag{8.13}$$

$$q_t + c^2\eta_x - \hat{\mu}q_{xxt} = 0. \tag{8.14}$$

In this case, we procede as in the first case, multiplying the first equation by $c^2\eta$, the second by q and adding the results, obtaining

$$c^2\eta\eta_t + c^2\eta q_x + qq_t + c^2q\eta_x - \hat{\mu}qq_{xxt} = 0. \tag{8.15}$$

Let us see how to integrate this last term: we have

$$qq_{xxt} = qq_{xtx} = (qq_{xt})_x - q_xq_{xt} = (qq_{xt})_x - \eta_t\eta_{tt} = qq_{xt})_x - \frac{1}{2}(\eta_t^2)_t. \tag{8.16}$$

So from the equation (8.15) we obtain the conservation law

$$\Big(\frac{c^2}{2}\eta^2 + \frac{1}{2}q^2 + \frac{\hat{\mu}}{2}\eta_t^2\Big)_t + (c^2\eta q + qq_{xt})_x = 0. \tag{8.17}$$

In this case the conserved quantity is

$$\frac{c^2}{2}\eta^2 + \frac{1}{2}q^2 + \frac{\hat{\mu}}{2}\eta_t^2. \tag{8.18}$$

8.2.2 Linear case up to $O(\delta^5)$ terms

Now we do the same thing with the fifth order system that can be obtained considering only the linear terms up to $O(\delta^5)$. The system in question is

$$\eta_t + q_x = 0 \tag{8.19}$$
$$q_t + c^2\eta_x - \hat{\mu} q_{xxt} + (\hat{v}^2 - \hat{\mu}^2) q_{xxxxt} = 0. \tag{8.20}$$

Proceeding as in the third order case with mixed derivatives, we obtain the equation

$$c^2\eta\eta_t + c^2\eta q_x + qq_t + c^2 q\eta_x - \hat{\mu} qq_{xxt} + (\hat{v}^2 - \hat{\mu}^2) qq_{xxxxt} = 0. \tag{8.21}$$

Let us see how we can integrate the last term. We have

$$\begin{aligned} qq_{xxxxt} = qq_{xxxtx} = (qq_{xxxt})_x - q_x q_{xxxt} = (qq_{xxxt})_x - \eta_t\eta_{ttxx} = \\ = (qq_{xxxt})_x - (\eta_t\eta_{ttx})_x + \eta_{xt}\eta_{ttx} = (qq_{xxxt})_x - (\eta_t\eta_{ttx})_x + \frac{1}{2}(\eta_{xt}^2)_t, \end{aligned}$$

so we obtain the conservation law

$$\begin{aligned} \Big(\frac{c^2}{2}\eta^2 + \frac{1}{2}q^2 + \frac{\hat{\mu}}{2}\eta_t^2 + (\hat{v}^2 - \hat{\mu}^2)\frac{1}{2}\eta_{xt}^2\Big)_t + \\ (c^2\eta q + qq_{xt} + (\hat{v}^2 - \hat{\mu}^2)(qq_{xxxt} - \eta_t\eta_{ttx}))_x = 0. \end{aligned} \tag{8.22}$$

The conserved quantity in this case is

$$\frac{c^2}{2}\eta^2 + \frac{1}{2}q^2 + \frac{\hat{\mu}}{2}\eta_t^2 + (\hat{v}^2 - \hat{\mu}^2)\frac{1}{2}\eta_{xt}^2. \tag{8.23}$$

8.2.3 Nonlinear case

Let us study now the way less trivial case of the third order system where we include the nonlinear terms, namely

$$\eta_t + q_x = 0 \tag{8.24}$$
$$q_t + c^2\eta_x + \hat{\beta}_1\eta\eta_x + \hat{\beta}_2(q^2)_x - \hat{\beta}_3 q\eta q_x - \hat{\beta}_4 q^2\eta - \hat{\beta}_5\eta^2\eta_x - \hat{\mu} q_{xxt} = 0 \tag{8.25}$$

Let us try to procede in the same way as done in linear case 8.13, multiplying the first equation by $c^2\eta$, the second by q and add them. We obtain, doing the same calculations, the following equation:

$$\begin{gathered}c^2\eta\eta_t + c^2\eta q_x + qq_t + c^2 q\eta_x + \hat{\beta}_1 q\eta\eta_x + \hat{\beta}_2 q(q^2)_x - \\ \hat{\beta}_3 q^2\eta q_x - \hat{\beta}_4 q^3\eta_x - \hat{\beta}_5 q\eta^2\eta_x - \hat{\mu} q q_{xxt} = 0.\end{gathered} \tag{8.26}$$

Let us see how we can integrate some of these terms:

- $q\eta\eta_x = (\frac{1}{2}q\eta^2)_x - \frac{1}{2}q_x\eta^2 = (\frac{1}{2}q\eta^2)_x + \frac{1}{2}\eta_t\eta^2 = (\frac{1}{2}q\eta^2)_x + \frac{1}{6}(\eta^3)_t$;
- $q(q^2)_x = (q^3)_x - q_x q^2 = (q^3)_x - \frac{1}{3}(q^3)_x = \frac{2}{3}(q^3)_x$;
- $q\eta^2\eta_x = \frac{1}{3}(q\eta^3)_x - \frac{1}{3}q_x\eta^3 = \frac{1}{3}(q\eta^3)_x + \frac{1}{3}\eta_t\eta^3 = \frac{1}{3}(q\eta^3)_x + \frac{1}{12}(\eta^4)_t$.

While for these terms we found a way to express them as a "total" derivative, for the terms $q^2\eta q_x$ and $q^3\eta_x$ is difficult to find such form and at this very moment we do not have this kind of stuff that would allow us to find the energy conservation law that we seek. Accepting this problem, we can do two things, observing that this two terms have something in common. Indeed,

- we can obtain, integrating one term, the other one plus a derivative term ($q^2\eta q_x = \frac{1}{3}(q^3\eta)_x - \frac{1}{3}q^3\eta_x$ and "viceversa");
- imposing $3\hat{\beta}_3 = \hat{\beta}_4$, we can observe that this sum is the derivative of $q^3\eta$.

Let us follow the second approach. In this case, we obtain

$$-\hat{\beta}_4(3q^2\eta q_x + q^3\eta_x) = -\hat{\beta}_4(q^3\eta)_x, \tag{8.27}$$

so now that every term can be expressed as sum of x and t derivatives, we obtain, from what we have developed in the linear case and what we have done now, the energy conservation law

$$\begin{gathered}\Big(\frac{c^2}{2}\eta^2 + \frac{1}{2}q^2 + \frac{\hat{\mu}}{2}\eta_t^2 + \frac{1}{6}\hat{\beta}_1\eta^3 - \frac{1}{12}\hat{\beta}_5\eta^4\Big)_t + \\ +\Big(c^2\eta q + qq_{xt} - \hat{\beta}_4 q^3\eta + \frac{1}{2}\hat{\beta}_1 q\eta^2 + \frac{2}{3}\hat{\beta}_2 q^3 - \hat{\beta}_4 q^3\eta - \frac{1}{3}\hat{\beta}_5 q\eta^3\Big)_x = 0.\end{gathered} \tag{8.28}$$

The conserved quantity for this non linear system is

$$\frac{c^2}{2}\eta^2 + \frac{1}{2}q^2 + \frac{\hat{\mu}}{2}q_x^2 + \frac{1}{6}\hat{\beta}_1\eta^3 - \frac{1}{12}\hat{\beta}_5\eta^4, \tag{8.29}$$

where we replaced η_t with $-q_x$ from the first equation. We need now to investigate if it is possible this relation between the coefficients $\hat{\beta}_3, \hat{\beta}_4$. Since $\hat{\beta}_3 = -\delta^2\alpha_1, \hat{\beta}_4 = -\delta^2\alpha_1$, our relation becomes $3\alpha_1 = \alpha_2$. From the Appendix A, we know that

$$\alpha_1 = \frac{2\langle H^{-2}\rangle^2 - 4\langle H^{-3}\rangle\langle H^{-1}\rangle}{\langle H^{-2}\rangle^2}; \tag{8.30}$$

$$\alpha_2 = \frac{3\langle H^{-2}\rangle^2 - 2\langle H^{-3}\rangle\langle H^{-1}\rangle - 3\langle H^{-4}\rangle}{2\langle H^{-2}\rangle^2}, \tag{8.31}$$

where $\langle\cdot\rangle$ indicates the average of a function in the interval $[0, 1]$. The condition $3\alpha_1 = \alpha_2$ becomes

$$\begin{gathered}\frac{9}{2}\langle H^{-2}\rangle^2 - 11\langle H^{-3}\rangle\langle H^{-1}\rangle + \frac{3}{2}\langle H^{-4}\rangle = 0 \Longrightarrow \\ 9\langle H^{-2}\rangle^2 - 22\langle H^{-3}\rangle\langle H^{-1}\rangle + 3\langle H^{-4}\rangle = 0.\end{gathered} \tag{8.32}$$

Let us try to find a possible bathimetry b periodic in $[0, 1]$ that allows $H = \eta^0 - b(x)$ to verify the previous identity. Let us consider

$$b(y) = \begin{cases} -\frac{1}{a_1} & \text{for} \quad x \in [0, \frac{1}{2}[\\ -\frac{1}{a_2} & \text{for} \quad x \in [\frac{1}{2}, 1] \end{cases} \tag{8.33}$$

with $a_1, a_2 \neq 0$. For simplicity, we fix $\eta^0 = 0$, so in the end we have

$$H(y) = \begin{cases} \frac{1}{a_1} & \text{for} \quad x \in [0, \frac{1}{2}[\\ \frac{1}{a_2} & \text{for} \quad x \in [\frac{1}{2}, 1] \end{cases} . \tag{8.34}$$

Let us compute the quantities involved in the equation:

- $\langle H^{-2}\rangle^2 = ((\int_0^{\frac{1}{2}} a_1^2 + \int_{\frac{1}{2}}^1 a_2^2)^2 = (\frac{1}{2}(a_1^2 + a_2^2))^2 = \frac{1}{4}(a_1^4 + a_2^4 + 2a_1a_2);$
- $\langle H^{-3}\rangle = \int_0^{\frac{1}{2}} a_1^3 + \int_{\frac{1}{2}}^1 a_2^3 = \frac{1}{2}(a_1^3 + a_2^3);$
- $\langle H^{-1}\rangle = \int_0^{\frac{1}{2}} a_1 + \int_{\frac{1}{2}}^1 a_2 = \frac{1}{2}(a_1 + a_2);$
- $\langle H^{-4}\rangle = \int_0^{\frac{1}{2}} a_1^4 + \int_{\frac{1}{2}}^1 a_2^4 = \frac{1}{2}(a_1^4 + a_2^4).$

Putting everything together, in the end we obtain from 8.32 the order four algebraic equation

$$7a_1^4 + 7a_2^4 - 18a_1^2a_2^2 + 22a_1a_3^3 + 22a_1^3a_3 = 0. \tag{8.35}$$

Using appropriate tools for calculations, we find that, for any a_1 fixed we find two different real value of a_2 that solve the equation. Unfortunately, these two a_2s that we find have always opposite sign with respect to a_1, so in this case we should give up on the assumption $H(y) > 0$ (in the paper, the really needed assumption seem to be $H(y) \neq 0$ so we are not losing too much if we just ask this instead of $H(y) > 0$). We can also consider different piecewise continuous functions in $[0, 1]$ and add a new degree of freedom given by the way of dividing the interval (for example, we can divide it in 3 smaller interval having the same length or divide the interval in two subintervals having different length).

Since we aim to general results, it is not that interesting to find conserved quantities for just some particular bathymetry b, so we should look for other ways to approach the problem. An idea that one can try to implement is to operate on the scaling in order to modify the equations. Since we want to give to each term the right "importance", the homogenized equations that we obtain are not as good as we seek, since there are terms that should not considered of the same order that instead appear to be multiplied by the same power of δ. In order to adjust this thing, we modify a little the system (6.11)-(6.12), replacing η with $\delta\eta$, q with δq, and then dividing everything by δ^2. In this way, we obtain the system

$$\eta_t + q_x = 0; \tag{8.36}$$

$$q_t + c^2\eta_x + \delta^2\Big(\frac{\langle H^{-2}\rangle}{\langle H^{-1}\rangle}((q^2)_x + c^2\eta\eta_x - \mu q_{xxt}\Big) + \delta^4\Big(- \frac{\alpha_2}{c^2}q^2 q_t - \frac{\langle H^{-3}\rangle}{\langle H^{-1}\rangle}\eta(2(q^2)_x - \eta q_t) - 2\frac{\gamma}{c^2}(2q_x q_{tt} - \eta q_{ttt}) + \frac{\nu_1}{c^4}q_{ttttt} + \frac{\nu_2}{c^2}q_{xxttt}\Big) = O(\delta^6), \tag{8.37}$$

where we switched the derivatives up to δ^2 terms as done when we derived (6.21)-(6.22) from (6.11)-(6.12) in order to underline the similarities between this system up to δ^3 and the previous one.

We can see now that the "bad" terms that prevented us from integrating the nonlinear equation up to $O(\delta^3)$ without imposing conditions on the coefficient that have been now transformed in terms of greater order, so their weight in the homogenized equations is surely smaller. The equations now are

$$\eta_t + q_x = 0 \tag{8.38}$$

$$q_t + c^2\eta_x + \hat{\beta}_1\eta\eta_x + \hat{\beta}_2(q^2)_x - \hat{\mu}q_{xxt} = 0, \tag{8.39}$$

and this case is equivalent to (8.24) with $\beta_3 = \beta_4 = \beta_5 = 0$, so now we can construct a conserved quantity for the third order system without saying anything about the coefficients. What we obtain, proceeding as before, is the following conservation law:

$$\begin{gathered}\Big(\frac{c^2}{2}\eta^2 + \frac{1}{2}q^2 + \frac{\hat{\mu}}{2}q_x^2 + \frac{1}{6}\hat{\beta}_1\eta^3\Big)_t + \\ \Big(c^2\eta q + qq_{xt}\eta + \frac{1}{2}\hat{\beta}_1 q\eta^2 + \frac{2}{3}\hat{\beta}_2 q^3\Big)_x = 0,\end{gathered} \tag{8.40}$$

so

$$\frac{c^2}{2}\eta^2 + \frac{1}{2}q^2 + \frac{\hat{\mu}}{2}q_x^2 + \frac{1}{6}\hat{\beta}_1\eta^3 \tag{8.41}$$

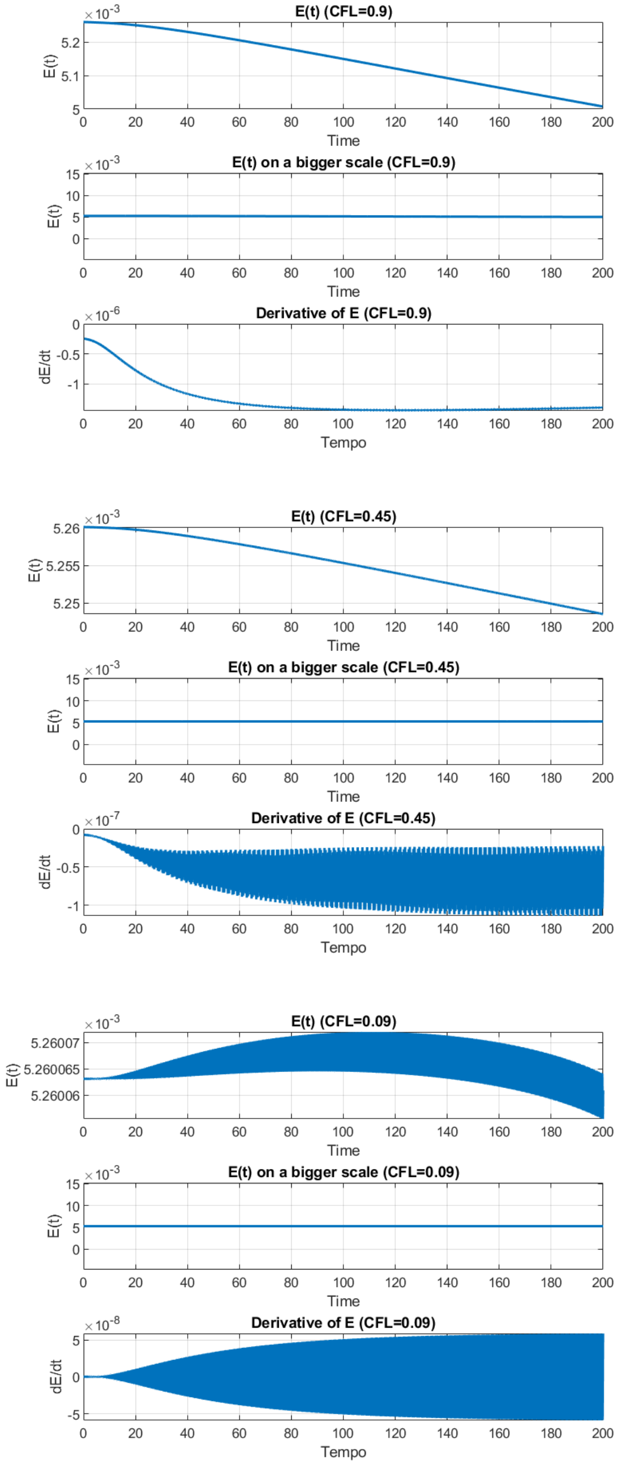
is a conserved quantity for the system (8.36)-(8.37). This means that the quantity

$$E(t) = \int \Big(\frac{c^2}{2}\eta^2 + \frac{1}{2}q^2 + \frac{\hat{\mu}}{2}q_x^2 + \frac{1}{6}\hat{\beta}_1\eta^3\Big)dx \tag{8.42}$$

is constant in time, i.e

$$\frac{dE(t)}{dt} = 0.$$

About the fifth-order system obtained dropping out terms multiplied by powers of δ greater than five, this new system apparently does not give any hint on how we can obtain conserved quantities. We could proceed as in the previous case, trying to impose conditions on coefficients, but again this is not really interesting. Further studies can be done in order to discover conserved quantities for this system. Let us now see on a numerical point of view how it is possible to confirm that the quantity is conserved. In Appendix C, the MATLAB code used for this numerical proof is shown. In order to compute our quantities η and q, a Fourier pseudo-spectral method in space and a Runge-Kutta 4 in time have been used, so the accuracy of our result is the same of the Runge-Kutta 4 method used in time.

 E(t) (CFL=0.9)

E(t) on a bigger scale (CFL=0.9)

Derivative of E (CFL=0.9)

E(t) (CFL=0.45)

E(t) on a bigger scale (CFL=0.45)

Derivative of E (CFL=0.45)

E(t) (CFL=0.09)

E(t) on a bigger scale (CFL=0.09)

Derivative of E (CFL=0.09)

Figure 8.1: From the top to the bottom: in each panel the plots of the conserved quantities, the same plot on a bigger scale and the plot of the derivative of $E(t)$ with CFL=0.9,0.45,0.09 are shown.

This means that, if we divide the time step Δt by a quantity h, the relative error defined as

$$\frac{|E(initial) - E(final)|}{E(final)},$$

should be reduced by a factor of h^4. Let us see that this happens. Since the time step of RK4 depends on the Courant–Friedrichs–Lewy convergence condition, that relates the time-step and the space-step in order to have a convergent method, we can reduce the time step working on this CFL condition. In Figure 8.1, the quantity E with $CFL = 0.9, 0.45, 0.09$ is shown. We can already see from the plot that the variation of E in time drops when the CFL is decreased. Computing the relative error of E for each CFL, we have

$$\text{Relative error for E with } CFL = 0.9 : 0.050384 \tag{8.43}$$

$$\text{Relative error for E with } CFL = 0.45 : 0.002190 \tag{8.44}$$

$$\text{Relative error for E with } CFL = 0.09 : 4.20053041 \cdot 10^{-7}. \tag{8.45}$$

We can see that the relative error decreases agreeing with the decreasing of the CFL and with the accuracy of the method used for the time discretization. This is a confirm that the quantity described is truly conserved.

8.3 Conserved Quantities for the homogenized 2D Shallow Water equations

Let us now look for conserved quantities the 2D shallow water homogenized system. The system considered is (6.61)-(6.62), that here we recall:

$$q_t + a_1\eta_x + a_2 q q_x - \tilde{\mu} q_{xxt} = 0 \tag{8.46}$$

$$\eta_t + q_x + a_2(\eta q)_x = 0. \tag{8.47}$$

Again, we recall that $a_1 := g\langle H\rangle$, $a_2 := \delta/\langle H\rangle$, $\tilde{\mu} = \delta^2\mu/\langle H\rangle$. We start observing that this system is similar to the Boussinesq System and the particular case of $a_1 = a_2 = 1$ falls into the class of more general systems of the form

$$q_t + \eta_x + qq_x + c\eta_{xxx} - dq_{xxt} = 0; \tag{8.48}$$

$$\eta_t + q_x + (q\eta)_x + aq_{xxx} - b\eta_{xxt} = 0. \tag{8.49}$$

considered in the paper [3]. In this very paper, we learn that the closest we can have in terms of non trivial conservation laws when $b \neq d$ (our case, where $b = 0, d = \tilde{\mu}$) is the formula

$$\frac{d}{dt}\int_{-\infty}^{\infty}(c\eta_x^2 + aq_x^2 - \eta^2 - q^2 - q^2\eta)dx = 2(b-d)\int_{-\infty}^{\infty}\eta_t q_{xt}, \tag{8.50}$$

so there does not seem to be much room for new interesting stuff or at least we can say that no particular energy conservation laws from this similar system can be adapted to our system. What one can try is to start doing some calculations and to hope that the peculiarities of this system will happen to make things better, even if usually things are supposed to be worse in more theoretical case. In particular, the difference between this system and the one considered in the paper is that the terms η_x and $qq_x, (\eta q)$ are multiplied respectively by "generic" coefficients a_1, a_2, where a_2 is proportional to δ. So, instead of looking for a utopian form of the kind

$$F_t + G_x = 0, \tag{8.51}$$

that does not seem to exist (at least apart from the trivial cases), we can try to construct something of the form

$$F_t + G_x = a_2^n H = O(\delta^n), \tag{8.52}$$

for some n, a positive integer, and H a function of q and η, so in this way we can truncate the equation at the order we want and find an almost-conserved quantity, where the residual part is as smaller as we want.

Let us try to build up this idea: multiplying the first equation by q, the second by $a_1\eta$ and summing them we obtain:

$$\frac{1}{2}(q^2)_t + \frac{a_1}{2}(\eta^2)_t + a_1(q\eta)_x + \frac{a_2}{3}(q^3)_x - \tilde{\mu}(qq_{xt})_x + \tilde{\mu}\frac{1}{2}(q_x^2)_t + a_1a_2\eta(\eta q)_x =$$
$$(F_0)_t + (G_0)_x + a_1a_2\eta(\eta q)_x = 0. \tag{8.53}$$

This last term will allows us to find the form of the equation that we seek. Indeed, we have

$$\begin{aligned}
&a_1a_2\eta(\eta q)_x = a_1a_2(\eta^2 q)_x - a_1a_2\eta_x\eta q = a_1a_2(\eta^2 q)_x - \frac{a_1a_2}{2}(\eta^2)_x q = \\
&= a_1a_2(\eta^2 q)_x - \frac{a_1a_2}{2}(\eta^2 q)_x + \frac{a_1a_2}{2}\eta^2 q_x = \\
&= \frac{a_1a_2}{2}(\eta^2 q)_x + \frac{a_1a_2}{2}\eta^2(-\eta_t - a_2(\eta q)_x) = \\
&= \frac{a_1a_2}{2}(\eta^2 q)_x - \frac{a_1a_2}{6}(\eta^3)_t - \frac{a_1a_2^2}{2}(\eta^2(\eta q)_x).
\end{aligned} \tag{8.54}$$

So, in a first iteration, we found that this term can be written in the form

$$\begin{aligned}
&a_1a_2\eta(\eta q)_x = \frac{a_1a_2}{2}(\eta^2 q)_x - \frac{a_1a_2}{6}(\eta^3)_t - \frac{a_1a_2^2}{2}(\eta^2(\eta q)_x) = \\
&(F_1)_t + (G_1)_x - \frac{a_1a_2^2}{2}\eta^2(\eta q)_x.
\end{aligned} \tag{8.55}$$

Let us now rewrite this last in order to confirm this kind of iteration rule. We have

$$\begin{aligned}
&-\frac{a_1a_2^2}{2}\eta^2(\eta q)_x = -\frac{a_1a_2^2}{2}(\eta^3 q)_x + \frac{a_1a_2^2}{2}(\eta^2)_x\eta q = \\
&= -\frac{a_1a_2^2}{2}(\eta^3 q)_x + a_1a_2^2\eta^2\eta_x q = -\frac{a_1a_2^2}{2}(\eta^3 q)_x + \frac{a_1a_2^2}{3}(\eta^3)_x q = \\
&= -\frac{a_1a_2^2}{6}(\eta^3 q)_x - \frac{a_1a_2^2}{3}(\eta^3)q_x = -\frac{a_1a_2^2}{6}(\eta^3 q)_x - \frac{a_1a_2^2}{3}\eta^3(-\eta_t - a_2(\eta q)_x) = \\
&= -\frac{a_1a_2^2}{6}(\eta^3 q)_x + \frac{a_1a_2^2}{12}(\eta^4)_t + \frac{a_1a_2^3}{3}(\eta^3(\eta q)_x).
\end{aligned} \tag{8.56}$$

So we obtain

$$\begin{aligned}
&-\frac{a_1a_2^2}{2}\eta^2(\eta q)_x = -\frac{a_1a_2^2}{6}(\eta^3 q)_x + \frac{a_1a_2^2}{12}(\eta^4)_t + \frac{a_1a_2^3}{3}(\eta^3(\eta q)_x) = \\
&(F_2)_t + (G_2)_x + \frac{a_1a_2^3}{3}\eta^3(\eta q)_x.
\end{aligned} \tag{8.57}$$

In the end, for each $n \in \mathbb{N}$ we find

$$\sum_{k=1}^{n}(F_k)_t + \sum_{k=1}^{n}(G_k)_x + a_2^n H_n = 0, \tag{8.58}$$

so we have for each n

$$\sum_{k=1}^{n}(F_k)_t + \sum_{k=1}^{n}(G_k)_x = \left(\sum^{n} F_k\right) + \left(\sum_{k=1}^{n} G_k\right)_x = O(\delta^n). \tag{8.59}$$

Since this procedure can be perpetuated for each n positive integer, we obtain a conservation law with quantities expressed as power expansion on δ, i.e. the following equation

$$F_t + G_x = 0 \tag{8.60}$$

holds with

$$F = F_0 + F_1 + F_2 + F_3 + \ldots = \tilde{F}_0 + \delta\tilde{F}_1 + \delta^2\tilde{F}_2 + \delta^3\tilde{F}_3 + \ldots \tag{8.61}$$
$$G = G_0 + G_1 + G_2 + G_3 + \ldots = \tilde{G}_0 + \delta\tilde{G}_1 + \delta^2\tilde{G}_2 + \delta^3\tilde{G}_3 + \ldots, \tag{8.62}$$

where the last-type expressions for F and G come from the fact that each term F_i, G_i is multiplied by a_2^i and a_2 contains in its definition a δ terms, so we can rewrite F_i as $\delta\tilde{F}_i$ and the same is possible for G_i. Moreover, we can observe that, from $i = 2$, the terms F_i, G_i can be written in the following way

$$F_2 = (-1)^i \frac{a_1 a_2^i}{(i+2)(i+1)}\eta^{i+2}; \tag{8.63}$$
$$G_2 = (-1)^{i-1} \frac{a_1 a_2^i}{(i+1)i}\eta^{i+1} q, \tag{8.64}$$

while the first two terms are defined as

$$F_0 = \frac{1}{2}q^2 + \frac{a_1}{2}\eta^2 - \frac{\tilde{\mu}}{2}q_x^2; \tag{8.65}$$
$$G_0 = a_1 q\eta - \tilde{\mu} q q_{xt}; \tag{8.66}$$
$$F_1 = -\frac{a_1 a_2}{6}\eta^3; \tag{8.67}$$
$$G_1 = \frac{a_2}{3}q^3 + \frac{a_1 a_2}{2}\eta^2 q. \tag{8.68}$$

Even if it was not our goal, we found a conservation law, and the very surprising thing is that the conserved quantity and the entropy flux are power expansions of δ. Now, for each n we can define the quantity

$$e_n = \sum_{i=0}^{n} F_i \tag{8.69}$$

and the quantity

$$E_n(t) = \int e_n dx. \tag{8.70}$$

We have that E_n is a an approximation of the conserved quantity F: the bigger is n, the more conserved we expect E_n to be. Let us see a numerical validation of both the conservation of F and the improvement of the approximation E_n as n increases. In Appendix C, the MATLAB code used for this numerical proof is shown. As done in the 1D shallow water case, in order to compute our quantities η and q, a Fourier pseudo-spectral method in space and a Runge-Kutta 4 in time has been used. In order to see that what done theoretically has a numerical confirm, we will work on the CFL, analysing how the relative error

$$\frac{|F(initial) - F(final)|}{F(final)},$$

decreases as the CFL decrease. Since F is a power series expansion, we will work with E_n, plotting the graph and computing the relative error for some of these quantities. In Figure 8.2, the quantity E_4 with $CFL = 0.9, 0.45, 0.09$ is plotted. We can already see from the plot that the variation of E_4 in time drops when the CFL is decreased. Let us compute the relative error of E_4 for each CFL: we have

$$\text{Relative error for } E_4 \text{ with } CFL = 0.9 : 0.047431 \tag{8.71}$$

$$\text{Relative error for } E_4 \text{ with } CFL = 0.45 : 0.002442 \tag{8.72}$$

$$\text{Relative error for } E_4 \text{ with } CFL = 0.09 : 8.343267188092798 \cdot 10^{-7}. \tag{8.73}$$

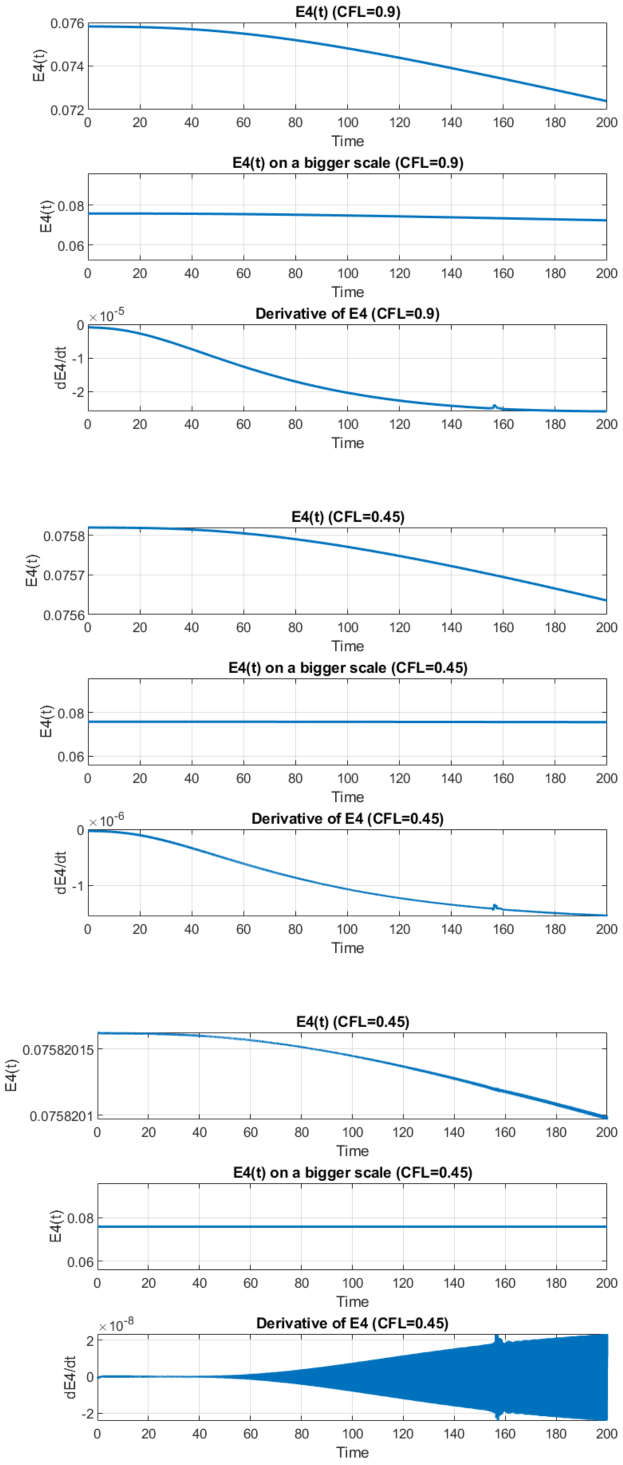


Figure 8.2: From the top to the bottom: in each panel the plots of the conserved quantities, the same plot on a bigger scale and the plot of the derivative of $E4(t)$ with CFL=0.9,0.45,0.09.

Moreover, we have a decreasing of the as n increase: indeed, we have, for $CFL = 0.9$:

$$\text{Relative error for } E_0 : 0.056897 \tag{8.74}$$

$$\text{Relative error for } E_2 : 0.047438 \tag{8.75}$$

$$\text{Relative error for } E_4 : 0.047431 \tag{8.76}$$

As supposed, we can see that there is an improvement of the relative error as number of terms involved in the approximation increase. It is possible to see the same holds for any CFL condition imposed and any number of term considered in the approximation. This ends our numerical validation of the conserved quantity determined previously.

Chapter 9

Conclusions and further developments

In this work a way to detect dispersive behaviour of solutions of hyperbolic systems in periodic media has been described, with a focus on the three models described in Chapter 6. In Chapters 2-5 the theoretical framework in which we are moving and the tools needed have been described, while, as said, Chapter 6 is dedicated to the application of what developed in the previous chapters on the 1D and 2D shallow water systems and the Euler system. The main original contributions of this work have been described in Chapters 7 and 8, which are a little improvement of what already done in works like [7],[8] and [9], that are the cornerstone of this thesis: while in Chapter 7 an alternative method for the computation of the travelling wave of the homogenized Euler systems has been described and developed, in Chapter 8 some conserved quantities for this models have been constructed and numerically validated. Further studies that can be conducted from this work, on both theoretical and applicative/numerical side, are:

- Generalization of Petviashvili method to the case of a sum of homogeneous terms with different degree;
- Discovery of different conserved quantities, especially for the homogenized Euler system;
- Development of conservative numerical methods, w here the conserved quantity in the construction of the method and not just for the validation.

Appendix A

1D shallow water equations

A.1 Averaging operators

In this section, we define the averaging operators used in Chapter 6. First of all, we define the *average operator* of a function f, and we indicate it as $\langle f\rangle \in \mathbb{R}$, is defined as

$$\langle f\rangle = \int_0^1 f(y)dy. \tag{A.1}$$

From this definition, we can define as the *fluctuating part* of f, and denote it as $\{f\}$, the quantity

$$\{f\}(y) = f(y) - \langle f\rangle. \tag{A.2}$$

Finally, we define the *fluctuating part of the antiderivative of the fluctuating part* as

$$[[f]] = \Big\{ \int_0^y \{f(\xi)\}d\xi \Big\}. \tag{A.3}$$

Clearly, we have

$$\langle\{f\}\rangle = 0 \text{ and } \langle[[f]]\rangle = 0.$$

A.2 Coefficients

In this section we provide the coefficients of the homogenized shallow water equations. The coefficients that appear in the system (6.11)-(6.12), (6.17)-

(6.18) and (6.21)-(6.22) are:

$$\mu = \frac{\langle [[H^{-1}]]^2 \rangle}{\langle H^{-1} \rangle^2} \tag{A.4}$$

$$\gamma = \frac{\langle [[H^{-1}]][[H^{-2}]] \rangle}{\langle H^{-1} \rangle^2} \tag{A.5}$$

$$\nu_1 = \frac{\langle H^{-1}([[[[H^{-1}]]]])^2 \rangle}{\langle H^{-1} \rangle^3} \tag{A.6}$$

$$\nu_2 = 3\frac{\langle ([[[[H^{-1}]]]])^2 \rangle}{\langle H^{-1} \rangle^2} \tag{A.7}$$

$$\alpha_1 = \frac{2}{\langle H^{-1} \rangle^2}\Big(\langle H^{-1} \rangle^2 - 2\langle H^{-3} \rangle \langle H^{-1} \rangle \Big) \tag{A.8}$$

$$\alpha_2 = \frac{3\langle H^{-2} \rangle^2 - 2\langle H^{-1} \rangle \langle H^{-3} \rangle - 3\langle H^{-4} \rangle}{2\langle H^{-1} \rangle^2} \tag{A.9}$$

$$\alpha_3 = \frac{1}{\langle H^{-1} \rangle^3}\Big(\langle H^{-2} \rangle^2 - \langle H^{-3} \rangle \langle H^{-1} \rangle \Big) \tag{A.10}$$

$$\alpha_4 = \frac{3\langle H^{-2} \rangle^3 - 4\langle H^{-1} \rangle \langle H^{-2} \rangle \langle H^{-3} \rangle - 3\langle H^{-2} \rangle \langle H^{-4} \rangle + 4\langle H^{-1} \rangle \langle H^{-5} \rangle}{\langle H^{-1} \rangle^3} \tag{A.11}$$

$$\alpha_5 = \frac{2\langle H^{-2} \rangle^3 - 6\langle H^{-1} \rangle \langle H^{-2} \rangle \langle H^{-3} \rangle + 6\langle H^{-1} \rangle^2 \langle H^{-4} \rangle}{\langle H^{-1} \rangle^3} \tag{A.12}$$

$$\alpha_6 = \frac{3\langle H^{-2} \rangle^3 - 7\langle H^{-1} \rangle \langle H^{-2} \rangle \langle H^{-3} \rangle - 3\langle H^{-1} \rangle \langle {}^2H^{-4} \rangle - 3\langle H^{-2} \rangle \langle H^{-4} \rangle + 6\langle H^{-1} \rangle \langle H^{-5} \rangle}{\langle H^{-1} \rangle^3} \tag{A.13}$$

$$\alpha_7 = \frac{\langle H^{-2} \rangle^3 - 2\langle H^{-1} \rangle \langle H^{-2} \rangle \langle H^{-3} \rangle + \langle H^{-1} \rangle^2 \langle H^{-4} \rangle}{\langle H^{-1} \rangle^3} \tag{A.14}$$

$$\alpha_8 = 2\Big(\mu\frac{\langle H^{-2} \rangle}{\langle H^{-1} \rangle} - \gamma \Big) \tag{A.15}$$

$$\alpha_9 = \mu\frac{\langle H^{-2} \rangle}{\langle H^{-1} \rangle}. \tag{A.16}$$

$$\hat{\alpha}_4 = \frac{4\langle H^{-5} \rangle - 2\langle H^{-2} \rangle \langle H^{-3} \rangle}{\langle H^{-1} \rangle^2} \tag{A.17}$$

$$\hat{\alpha}_6 = \frac{5\langle H^{-2} \rangle \langle H^{-3} \rangle - 3\langle H^{-1} \rangle \langle H^{-4} \rangle - 6\langle H^{-5} \rangle}{\langle H^{-1} \rangle^2} \tag{A.18}$$

$$\hat{\alpha}_8 = -4\gamma + 10\mu\frac{\langle H^{-2} \rangle}{\langle H^{-1} \rangle} \tag{A.19}$$

$$\hat{\alpha}_9 = 8\mu\frac{\langle H^{-2} \rangle}{\langle H^{-1} \rangle^2} \tag{A.20}$$

$$\hat{\alpha}_{10} = 3\mu\frac{\langle H^{-2} \rangle}{\langle H^{-1} \rangle^2} - \frac{2\gamma}{\langle H^{-1} \rangle} \tag{A.21}$$

$$\hat{\alpha}_{11} = 4\mu\frac{\langle H^{-2} \rangle}{\langle H^{-1} \rangle} \tag{A.22}$$

Appendix B

Euler equations

In this section we provide the coefficients of the homogenized shallow water equations. The coefficients that appear in the system (6.98)-(6.99) are:

$$\mu = \frac{\langle [[K^{-1}]]^2 \rangle}{\langle K^{-1} \rangle^2} \tag{B.1}$$

$$\zeta = \langle K^{-1}([[K^{-1}]])^2 \rangle \tag{B.2}$$

$$\alpha_1 = -9\frac{G'}{G} \tag{B.3}$$

$$\alpha_2 = -15\frac{G'}{G^3} \tag{B.4}$$

$$\alpha_3 = \frac{62G'^2 - \frac{37}{2}GG''}{G^4} \tag{B.5}$$

$$\alpha_4 = \frac{46G'^2 - 13GG''}{G^4} \tag{B.6}$$

$$\alpha_5 = \frac{-11G^2G''' - 160G'^3 + 108GG'G''}{G^5} \tag{B.7}$$

$$\alpha_6 = \frac{-G^3G^{(4)} + 9G^2G''^2 + 75G'^4 + 13G^2G'''G' - \frac{165}{2}GG'^2G''}{G^6} \tag{B.8}$$

$$\alpha_7 = \frac{1}{G^2} \tag{B.9}$$

Appendix C

Matlab codes

The codes used for the numerical computation of the results obtained in this work can be seen at the following link:
https://github.com/GiuseppeMinissale/CodesMasterThesis/tree/main.
In particular:

- Using the MATLAB code named *PetKdv.m* one can compute the travelling wave for the KdV equation using the Petviashvili-type method. The results have been shown in Section 7.3.

- Using the MATLAB code named *Petviashvili_Euler_order_3.m* one can compute an accurate approximation of the travelling wave for the homogezined Euler system of order three using the Petviashvili-type method. The results have been shown in Section 7.4.1.

- Using the MATLAB code named Petviashvili_Euler_order_5.m one can compute an accurate approximation of the travelling wave for the homogezined Euler system of order five using the Petviashvili-type method described. The results have been shown in Section 7.4.2.

- Using the MATLAB code named ConsSW.m one can compute the conserved quantity for the homogenized 1D Shallow water equations. The results have been shown in Section 8.2.3.

- Using the MATLAB code named ConsSW2D.m one can compute the conserved quantity for the homogenized 2D Shallow water equations. The results have been shown in Section 8.3.

Bibliography

[1] M. A. Ablowitz and P. A. Clarkson. *Solitons, Nonlinear Evolution Equations and Inverse Scattering.* London Mathematical Society Lecture Note Series. Cambridge University Press, 1991.

[2] J. Alvarez and A. Duran. *Petviashvili type methods for traveling wave computations: I. Analysis of convergence.* 2013. arXiv: `1311.2546 [math.NA]`. URL: `https://arxiv.org/abs/1311.2546`.

[3] J L Bona, M Chen, and J-C Saut. "Boussinesq equations and other systems for small-amplitude long waves in nonlinear dispersive media: II. The nonlinear theory". In: *Nonlinearity* 17.3 (Feb. 2004), p. 925. DOI: `10.1088/0951-7715/17/3/010`. URL: `https://dx.doi.org/10.1088/0951-7715/17/3/010`.

[4] Joseph Boussinesq. "Théorie des ondes et des remous qui se propagent le long d'un canal rectangulaire horizontal, en communiquant au liquide contenu dans ce canal des vitesses sensiblement pareilles de la surface au fond." In: *Journal de Mathématiques Pures et Appliquées* (), pp. 55–108. URL: `https://api.semanticscholar.org/CorpusID:125469048`.

[5] James Glimm. "Solutions in the large for nonlinear hyperbolic systems of equations". In: *Communications on Pure and Applied Mathematics* 18.4 (1965), pp. 697–715. DOI: `https://doi.org/10.1002/cpa.3160180408`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.3160180408`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.3160180408`.

[6] Reza N. Jazar. "Multiple Scale Method". In: *Perturbation Methods in Science and Engineering.* Cham: Springer International Publishing, 2021, pp. 527–550. ISBN: 978-3-030-73462-6. DOI: `10.1007/978-3-

030-73462-6_9. URL: https://doi.org/10.1007/978-3-030-73462-6_9.

[7] David I. Ketcheson, Lajos Lóczi, and Giovanni Russo. *A multiscale model for weakly nonlinear shallow water waves over periodic bathymetry.* 2023. arXiv: 2311.02603 [math.AP]. URL: https://arxiv.org/abs/2311.02603.

[8] David I. Ketcheson and Giovanni Russo. *A dispersive effective equation for transverse propagation of planar shallow water waves over periodic bathymetry.* 2024. arXiv: 2409.00076 [math.AP]. URL: https://arxiv.org/abs/2409.00076.

[9] David I. Ketcheson and Giovanni Russo. *Solitary wave formation in the compressible Euler equations.* 2024. arXiv: 2412.11086 [math.AP]. URL: https://arxiv.org/abs/2412.11086.

[10] D. J. Korteweg and G. de Vries. "XLI. On the change of form of long waves advancing in a rectangular canal, and on a new type of long stationary waves". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 39.240 (1895), pp. 422–443. DOI: 10.1080/14786449508620739. eprint: https://doi.org/10.1080/14786449508620739. URL: https://doi.org/10.1080/14786449508620739.

[11] Sukeyuki Kumei. "Applications of Lie Groups to Differential Equations (Peter J. Olver)". In: *SIAM Review* 30.2 (1988), pp. 336–337. DOI: 10.1137/1030074. eprint: https://doi.org/10.1137/1030074. URL: https://doi.org/10.1137/1030074.

[12] Peter D. Lax. *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves.* Society for Industrial and Applied Mathematics, 1973. DOI: 10.1137/1.9781611970562. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611970562. URL: https://epubs.siam.org/doi/abs/10.1137/1.9781611970562.

[13] Peter D. Lax. "The Formation and Decay of Shock Waves". In: *The American Mathematical Monthly* 79.3 (1972), pp. 227–241. DOI: 10.1080/00029890.1972.11993023. eprint: https://doi.org/10.1080/00029890.1972.11993023. URL: https://doi.org/10.1080/00029890.1972.11993023.

[14] Randall J. LeVeque. "Gas Dynamics and the Euler Equations". In: *Finite Volume Methods for Hyperbolic Problems.* Cambridge Texts in Applied Mathematics. Cambridge University Press, 2002, pp. 291–310.

[15] Randall J. LeVeque. *Numerical Methods for Conservation Laws.* 2nd. Basel: Birkhäuser, 1992.

[16] Randall J. Leveque and Darryl H. Yong. "Solitary Waves in Layered Nonlinear Media". In: *SIAM Journal on Applied Mathematics* 63.5 (2003), pp. 1539–1560. ISSN: 00361399. URL: `http://www.jstor.org/stable/4096050` (visited on 11/15/2024).

[17] Ruchika Lochab and Vivek Kumar. "A comparative study of high-resolution methods for nonlinear hyperbolic problems". In: *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik* 102.7 (2022), e202100462. DOI: `https://doi.org/10.1002/zamm.202100462`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/zamm.202100462`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/zamm.202100462`.

[18] MANUEL QUEZADA DE LUNA and DAVID I. KETCHESON. "TWO-DIMENSIONAL WAVE PROPAGATION IN LAYERED PERIODIC MEDIA". In: *SIAM Journal on Applied Mathematics* 74.6 (2014), pp. 1852–1869. ISSN: 00361399. URL: `http://www.jstor.org/stable/24511424` (visited on 11/16/2024).

[19] John W. Miles. "Linear and Nonlinear Waves. By G. B. WHITHAM. Wiley-Interscience, 1974". In: *Journal of Fluid Mechanics* 87.2 (1978). DOI: `10.1017/S0022112078211676`.

[20] R. E. O'malley. "Perturbation Methods. By A. H. NAYFEH. Wiley, 1973. 425 pp. £9 (hardback) or £3.75 (paperback)." In: *Journal of Fluid Mechanics* 63.3 (1974), pp. 623–623. DOI: `10.1017/S0022112074211820`.

[21] V. I. Petviashvili. "Equation of an extraordinary soliton". In: *Soviet Journal of Plasma Physics* 2 (May 1976), p. 257.

[22] Manuel Quezada de Luna and David I. Ketcheson. "Solitary water waves created by variations in bathymetry". In: *Journal of Fluid Mechanics* 917 (2021), A45. DOI: `10.1017/jfm.2021.267`.

[23] Fadil Santosa and William W. Symes. “A Dispersive Effective Medium for Wave Propagation in Periodic Composites”. In: *SIAM Journal on Applied Mathematics* 51.4 (1991), pp. 984–1005. DOI: 10.1137/0151049. eprint: https://doi.org/10.1137/0151049. URL: https://doi.org/10.1137/0151049.

[24] B. Shivamoggi. *Perturbation Methods for Differential Equations.* Perturbation Methods for Differential Equations. Birkhäuser Boston, 2002. ISBN: 9780817641894. URL: https://books.google.it/books?id=OgWA8qYcQlIC.