Scuola Internazionale Superiore di Studi Avanzati - Trieste

Doctoral Thesis

Numerical study of dispersive shock waves in the KPI Equation

---

Author:
Giuseppe PITTON

Supervisors:
Prof Tamara GRAVA
Prof Luca HELTAI

PhD course in Mathematical Analysis, Modelling and Applications

Academic year 2017-2018

SISSA - Via Bonomea 265 - 34136 TRIESTE - ITALY

Abstract

Since its introduction in 1970 in the context of two-dimensional hydrodynamics, the Kadomtsev–Petviashvili (KP) Equation has been rediscovered in several, seemingly unrelated, areas of pure and applied Mathematics; in this respect the KP Equation shares the interdisciplinary nature of its one-dimensional counterpart, the Korteweg–De Vries Equation.

Despite the growing interest towards the KP Equation, little is known about the evolution of a generic, localised initial datum, particularly in the small dispersion limit.

This thesis aims at addressing precisely this topic. To this end, the long-time behaviour of the KPI Equation is studied by means of high-quality numerical computations; the numerical results show that generically the positive part of KP's dispersive waves are subject to a focusing instability phenomenon, and that this instability leads to the formation of a number of two-dimensional solitons, which arrange in a triangular lattice, and whose height and position scales with the dispersion coefficient.

We give a precise mathematical description of the numerical observations by deriving the modulation equations for the KP case by means of Whitham's averaging procedure. Whitham's theory is applied both in the general case of KP's traveling waves and in the special case of the line soliton solution, that is considered here as an approximation of the leading wave front in a packet of dispersive shock waves. We infer from Whitham's equations for the line soliton the existence of a caustic singularity that is responsible for the wave breaking observed numerically.

A further analysis of the modulation system for the line soliton leads us to point out an analogy between the KP Equation and the semiclassical limit of the focusing nonlinear Schrödinger Equation. Following this observation, we propose an asymptotic expression for the location of the KP solitons, corroborated by extensive numerical computations.

The numerical computations involved in this thesis are a particularly delicate task, due to the high quality results required in the small dispersion case and to the stiff, non-local, linear operator. For this reason, particular care is devoted to the study of a suitable time advancement method and to the crafting of an ad-hoc, parallel, numerical code.

Acknowledgements

Devising an accurate, reliable approximation for a nonlinear dispersive PDE is a painstaking endeavour. I would like to thank Christian Klein for sharing his experience and scrupulous attention to every detail of a proper approximation task.

I wish to thank Roberto Camassa for his hospitality at the University of North Carolina at Chapel Hill and for the many discussions and his insights on the theory of nonlinear wave equations. For my visit at UNC, I would like to thank the support received from the H2020-MSCA-RISE-2017 PROJECT No. 778010 IPaDEGAN.

I would like to thank Alessandro Michelangeli for his constant support and teaching throughout the course of my PhD studies. I am grateful to Evgenii Ferapontov for his remarks on the integrability of Whitham's Equations; to Martin Kronbichler and Giovanni Ortenzi for reading carefully this thesis and suggesting appropriate amendments; to Marco De Paoli for his valuable comments on early drafts of the present manuscript.

The large-scale computations required for this work would not have been possible without the resources provided through the Iscra C project "SCALEKPI", id: HP10CS4FHA, by CINECA, Bologna, Italy, which I acknowledge gratefully. I acknowledge as well the computational resources provided by SISSA.

Contents

Chapter 1

Introduction

The Kadomtsev–Petviashvili (KP) Equation is one of the most popular models for two-dimensional wave propagation. This equation describes the time evolution of a two-dimensional disturbance over the surface of a shallow basin by specifying a scalar function $u(x, y, t)$ that can be interpreted as the displacement of the surface with respect to a reference level. If the initial disturbance is prescribed by the scalar function $u(x, y, 0) = u_0(x, y)$, with u_0 a known function, then in the KP model the propagation of such disturbance is held by the following partial differential equation:

$$\left(u_t + uu_x + \epsilon^2 u_{xxx}\right)_x + \alpha u_{yy} = 0, \tag{1.1}$$

where x and y represent the two space coordinates, t is the time coordinate, ϵ is a dispersion/scaling coefficient, and $\alpha = \pm 1$ will be discussed in the next section.

It is known that a generic, localised initial datum evolves under the KP flow by forming a region with high gradients that leads to the formation in finite time of an oscillatory region; and that the oscillations' frequency is inversely proportional to the dispersion coefficient as ϵ approaches zero. These high frequency oscillations are called dispersive shock waves.

This thesis is concerned with the study of the formation, propagation and stability of dispersive shock waves in the KP Equation over long times.

The structure of this exposition is the following. In the first section of the current chapter, the generic structure of wave formation and propagation is sketched, presenting known and new results that will be covered in greater detail in other parts of the thesis. The rest of this chapter is devoted to the introduction of the KP Equation and to recall some of the many mathematical properties enjoyed by these equations.

In Chapter 2, Whitham's theory is recalled and applied to the KP Equation, leading to the explicit formulation of a modulation system for the line soliton and the lump solutions of the KPI Equation.

Then, a numerical approach for obtaining high quality approximations to the KP wave propagation is discussed in Chapter 3. The dispersive, non-local, stiff nature of the linear part of the KPI Equation makes its numerical approximation a particularly difficult task, and the efforts to provide an accurate, yet efficient, numerical code led us to the development of a Composite Runge–Kutta method with a new, ad-hoc, low-fast splitting. The detailed derivation of this numerical integrator, as well as a comparison with other highly efficient schemes are the main topic of Chapter 3.

The low-level details on the practical implementation of a computer code and on its parallel algorithms are presented in Chapter 4.

Finally, a detailed numerical study of the small dispersion KPI Equation is the main theme of Chapter 5. Besides, in Chapter 5 the scaling laws for lump position and height are determined numerically and compared with the known scalings for the semiclassical limit of the focusing nonlinear Schrödinger Equation.

Part of the results of this thesis can be found in published form in the joint work of the author with Tamara Grava and Christian Klein [57].

1.1 Nonlinear waves and the KP Equation

For over two centuries, many simplified models have been proposed for describing the motion of surface waves of an incompressible fluid. Here the word “simplified” is used to qualify models that are much simpler than the full, general, hydrodynamic equations (such as the Euler or the Navier–Stokes Equations) that could be adopted for describing a great deal of flow configurations, including wave motion. Such simplified models originate from several hypotheses that aim at describing wave motion only for a specific propagation regime. A frequent kinematic assumption is that of waves having an essentially monodimensional direction of propagation, neglecting to some extent the inherently three dimensional motion of fluids. Often this leads to the introduction of a single scalar quantity[1], u, that denotes both the height of the fluid surface above an initially unperturbed level, and the average speed of the fluid in a certain region of space. Another recurring assumption is the dimensionality of the wave propagation problem, that can be reduced to a one-dimensional motion if the disturbance can be considered “very long” with respect to the channel’s depth, or two-dimensional if such “long wave” hypothesis does not hold, but there is no mixing between fluid layers at different depth. A third assumption that characterises wave propagation regimes is related to the depth of the channel or basin in which the fluid is confined. We will not re-derive here any of the wave equations, limiting ourselves to a brief panoramic view of some of the most important models, and referring to [5] for a broader treatment.

One of the earliest and most widely known of the many wave models is undoubtedly the Korteweg–De Vries (KdV) Equation:

$$u_t + u_x + uu_x + u_{xxx} = 0. \tag{1.2}$$

The KdV Equation was derived for the description of essentially monodimensional wave propagation along a shallow channel, and it was immediately recognised that despite its simplicity can describe with very good accuracy the physics of shallow water waves propagating in one dimension. By considering the linear part of Equation (1.2):

$$u_t + u_x + u_{xxx} = 0, \tag{1.3}$$

and seeking for a traveling wave solution:

$$u(x,t) = a\mathrm{e}^{\mathrm{i}(kx-\omega t)}, \tag{1.4}$$

leads to an algebraic compatibility condition often called *dispersion relation*. For the KdV Equation, the dispersion relation has the form:

$$\omega = k - \frac{k^3}{3}, \tag{1.5}$$

which is a fourth-order accurate expansion of the nonlinear dispersion relation for wave propagation in shallow water:

$$\omega = \tanh k. \tag{1.6}$$

In this work, the KdV Equation is considered in a slightly different form:

$$u_t + uu_x + \epsilon^2 u_{xxx} = 0, \tag{1.7}$$

which follows from (1.2) after a change of reference frame and the introduction of a rescaling dependent on $\epsilon > 0$. More precisely, by introducing a reference frame moving with constant unitary speed in the x-direction, whose coordinates are:

$$X = x - t, \qquad\qquad T = t, \tag{1.8}$$

---

[1] There exist many other wave models whose formulation consists of systems of scalar quantities, such as the Airy model for shallow waters, that incorporates the mean velocity and the surface height.

then the solution u, written in terms of the new coordinate frame becomes:

$$U(X,T) = u(x,t), \tag{1.9}$$

and the time derivative transforms as:

$$u_t = U_T \frac{\partial T}{\partial t} + U_X \frac{\partial X}{\partial t} = U_T - U_X. \tag{1.10}$$

The space derivatives are not affected by the coordinate change (1.8), so the KdV Equation (1.2) in the new variables becomes:

$$U_T + UU_X + U_{XXX} = 0, \tag{1.11}$$

so that after a renaming, Equation (1.11) is equivalent to the (1.7) with $\epsilon = 1$.

After the groundbreaking work of Korteweg and De Vries, several other models for one-dimensional wave propagation have appeared, among which we mention the following.

Benjamin–Bona–Mahony (BBM). Introduced in [17] has the form:

$$u_t + u_x + uu_x - u_{xxt} = 0, \tag{1.12}$$

and was derived under the same assumptions of the KdV Equation, but with a different expansion for the dispersion relation. Specifically, the dispersion relation for Equation (1.12) is given by:

$$\omega = \frac{k}{1+k^2}, \tag{1.13}$$

which comes from a Padé expansion of order $(2,2)$ of the nonlinear dispersion relation (1.6).

Benjamin–Ono (BO). Introduced in [16] and further discussed in [104], is a model for the study of internal wave propagation in channels of infinite depth, and has the form:

$$u_t + u_x + uu_x + Hu_{xx} = 0,$$

where H is the Hilbert transform, defined as:

$$Hu(x) = \frac{1}{\pi} \fint_{-\infty}^{\infty} \frac{u(y)}{y-x}\, \mathrm{d}y. \tag{1.14}$$

In this case, there is no immediate analogy with the KdV and the BBM models, since in the deep wave regime the dispersion relation is different than Equation (1.6).

Camassa–Holm (CH).

$$u_t + u_x + uu_x - u_{xxt} = (u(u-u_{xx}))_x \tag{1.15}$$

This model was derived in [26] as a higher order extension of the BBM Equation, and indeed Equation (1.15) shares the dispersion relation (1.13) of the BBM model. Remarkably, the fully nonlinear dispersion allows for the description of wilder scenarios, such as wave breaking. The authors of the original paper [26] immediately recognised the existence of soliton solutions with discontinuous first derivative, sometimes called “peakons”, with expression:

$$u(x,t) = a e^{-|x-at|}. \tag{1.16}$$

Wave modelling is still a very active research field, and over the years countless generalisations or perturbations of the four main classes described above have been proposed. These Equations, despite being introduced for hydrodynamics applications, have many remarkable properties which makes their study an interesting mathematical pursuit regardless of their physical origin.

In an effort to generalise the KdV Equation, Kadomtsev and Petviashvili [75] derived what is probably the most popular model of two-dimensional dispersive wave propagation. The resulting equation, that now bears their name and will be abbreviated with KP in the following, was derived for the description of wave propagation on a two dimensional surface under nonlinear dispersion, with the hypothesis that the direction of propagation is essentially monodimensional, with weak dependence on the orthogonal direction. The form of the KP Equation chosen in the present work is:

$$\left(u_t + uu_x + \epsilon^2 u_{xxx}\right)_x + \alpha u_{yy} = 0. \tag{1.17}$$

The case of strong surface tension is characterised by $\alpha = -1$ and is called the KPI Equation, while the case with weak surface tension corresponds to $\alpha = 1$ and is called the KPII Equation. For more details on this characterisation, we refer to [4, § 1.2]. It is often convenient to recast Equation (1.17) in evolutionary form:

$$u_t + uu_x + \epsilon^2 u_{xxx} + \alpha \partial_x^{-1} u_{yy} = 0, \tag{1.18}$$

where the antiderivative can be defined in Fourier components:

$$\partial_x^{-1} u = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\mathrm{i}\xi_1} \widehat{u}(\xi_1, \xi_2, t) \mathrm{e}^{\mathrm{i}\xi_1 x} \mathrm{e}^{\mathrm{i}\xi_2 y} \, \mathrm{d}\xi_1 \, \mathrm{d}\xi_2, \tag{1.19}$$

with

$$\widehat{u}(\xi_1, \xi_2, t) := \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x, y, t) \mathrm{e}^{-\mathrm{i}\xi_1 x} \mathrm{e}^{-\mathrm{i}\xi_2 y} \, \mathrm{d}x \, \mathrm{d}y, \tag{1.20}$$

or alternatively with the integral:

$$\partial_x^{-1} u = \int_{-\infty}^{x} u(s, y, t) \, \mathrm{d}s, \tag{1.21}$$

but see the discussion in Section 1.3.4 on the definition of this term.

Since its introduction in 1970, the KP Equation has appeared in many seemingly unrelated branches of Pure Mathematics, see [103] for a review on the connections of the KP Equation with higher structures in Algebra and Geometry. The small dispersion ($\epsilon \ll 1$) regime is particularly important in the study of initial data for the Cauchy problem:

$$(u_t + uu_x + u_{xxx})_x + \alpha u_{yy} = 0 \qquad\qquad u(x, y, 0) = u_0(x, y) \tag{1.22}$$

in the case when u_0 is a slowly-changing function of x and y. In this case, it is expected that the space derivatives are small, and this in turn implies that the time derivative is small too. As a result, the time evolution can be expected to be slow. A possible workaround to study the time evolution while keeping $t \simeq O(1)$ is to consider the scaling:

$$U(X, Y, T) = u(\epsilon x, \epsilon y, \epsilon t), \tag{1.23}$$

so that Equation (1.22) for $u(x, y, t)$ becomes:

$$\left(U_T + UU_X + \epsilon^2 U_{XXX}\right)_X + \alpha U_{YY} = 0 \tag{1.24}$$

for the rescaled unknown $U(X, Y, T)$.

1.2 Shock wave formation in the KPI Equation

In this section we sketch the general mechanism by which dispersive shock waves form and propagate in the KPI Equation under small dispersion, $\epsilon \ll 1$. This process is generically characterised by two singular points. A first singularity appears in the solution to the dispersionless KP Equation, that can be solved explicitly by a deformation of the characteristics method, and is related to the onset of oscillations in the dispersive equation. This first singular behaviour was discovered in [42], further investigated in [58], and studied numerically in [82]. After the dispersive shock waves have developed, these start propagating along a parabolic wavefront, until for a sufficiently long time the leading positive wavefront manifests a focusing instability leading to the development of a very high, fast-moving, and localised peak, which is asymptotically approximated by a lump solution. For even longer times, more lumps emerge from the focusing of the secondary wavefronts, which after a complex interaction with the dispersive shock waves arrange in a triangular lattice. This generic behaviour was completely unknown for the KP equation and it was discovered by performing high quality numerics.

For negative initial data, the first singularity described by the dispersionless limit does still develop, leading to the formation of dispersive shock waves of negative mean value, but the second singularity does not appear, at least for the time scales considered in this work. This could be related to the fact that numerically the dispersive shock waves with negative mean seem to be stable [83]. In the article [58], it is shown that two singular points appear in the solution: one in the negative part and one in the positive part. Both singular points are responsible for the formation of two fronts of dispersive shock waves, appearing respectively in the negative and in the positive part of the solution. For the KPI Equation, the shock in the positive part of the solution is stronger, and develops earlier than in the negative part of the solution; conversely, for the KPII Equation the negative shock is stronger, and takes place at an earlier time than in the positive part of the solution.

1.2.1 Shock wave formation

A valid tool for studying the onset of a shock in the small dispersion regime is given by the dispersionless KP (dKP) Equation, obtained by setting $\epsilon = 0$ in Equation (1.17):

$$(u_t + uu_x)_x + u_{yy} = 0. \tag{1.25}$$

Note that despite its name, Equation (1.25) is actually an equation of dispersive nature. By considering a change of variables in the form:

$$\begin{aligned} u &\to -u, \\ t &\to -t \end{aligned} \tag{1.26}$$

it is possible to obtain from the (1.25) its counterpart for the KPI Equation:

$$(u_t + uu_x)_x - u_{yy} = 0. \tag{1.27}$$

This equation provides one of the simplest models for the development and propagation of shock waves in two space dimensions. In analogy with the Hopf Equation in 1D:

$$u_t + uu_x = 0, \tag{1.28}$$

Equation (1.25) does develop a singularity at a single point (x_c, t_c) in finite time t_c, where the solution remains bounded and takes the value $u_c = u(x_c, y_c, t_c)$.

A useful tool developed in [96, 97, 58] for solving explicitly Equation (1.25) is a two-dimensional deformation of the characteristics method, where the solution is given by:

$$u(x, y, t) = F(\eta, y, t), \tag{1.29}$$

and η is defined implicitly as:

$$\begin{cases} x = tF(\eta, y, t) + \eta \\ F(x, y, 0) = u(x, y, 0). \end{cases} \tag{1.30}$$

Grava, Klein and Eggers proceed showing that F is a solution of the following PDE:

$$\left(\frac{F_t + tF_y^2}{1 + tF_\eta} \right)_\eta = \alpha F_{yy}, \tag{1.31}$$

that can also be written in the evolutionary form:

$$F_t = -\alpha \partial_\eta^{-1} F_{yy} + t \left(F_\eta \partial_\eta^{-1} F_{tt} - F_y^2 \right). \tag{1.32}$$

The nonlinear part of Equation (1.32) is multiplied by t, so for small times the equation for F is "less nonlinear" than the dKP Equation. As a result the PDE (1.31) remains regular for times longer than t_c; this allows to detect the singular point as the point at which the transformation (1.30) is no longer invertible for η, which mimics the crossing of characteristics in the Hopf Equation. The authors are also able to show that generically the singularity of the dKP Equation unfolds as a cusp catastrophe.

In [58] the singular behaviour of the dKP Equation is further studied by means of multiple scale expansions close to the singularity point; this approach allows the authors to find the region where the solution given by characteristics is multiple valued, and to track the shock front. They also show that the behaviour close to the shock front follows a universal scaling, which is generically a 3/2 cusp. A similar parabolic profile for the leading front of the multivalued region is found in [111], even for a more general class of modified KP Equations.

In the work [42], the authors obtain a small-ϵ expansion for the KP Equation in a neighbourhood of the singular point:

$$u(x,y,t) = u_c + 6 \left(\frac{\epsilon^2}{\kappa^2} \right)^{\frac{1}{7}} U(\xi, \tau) + \beta(y - y_c) + O(\epsilon^{\frac{4}{7}}), \tag{1.33}$$

where

$$\begin{gathered} X = (x - x_c) = u_c(t - t_c) + c_1(t - t_c)(y - y_c) + c_2(y - y_c) + c_3(y - y_c)^2 + c_4(y - y_c)^3 \\ T = (t - t_c) + b(y - y_c) \\ \xi = \frac{X}{(\kappa \epsilon^6)^{1/7}} \\ \tau = \frac{T}{(\kappa^3 \epsilon^4)^{1/7}}, \end{gathered} \tag{1.34}$$

for $c_1, c_2, c_3, c_4, b, \beta, \kappa$ constants, and the function U is a solution of the Painlevé I2 Equation (see Section 1.3.7 for self-similar solutions given in terms of Painlevé Transcendents):

$$\xi = 6\tau U - \left(U^3 + \frac{1}{2} U_\xi^2 + U U_{\xi\xi} + \frac{1}{10} U_{\xi\xi\xi\xi} \right), \tag{1.35}$$

complemented by suitable asymptotic conditions. The fact that these oscillations are described by transcendent functions reveals the inherently nonlinear nature of shock formation. This could be expected, given the large energies involved with shock waves.

1.2.2 Shock propagation and instability

We conclude the discussion by illustrating the behaviour described above with the help of a numerical computation. We consider the initial datum:

$$u(x,y,0) = -6 \partial_x \operatorname{sech}^2 \sqrt{x^2 + y^2}, \tag{1.36}$$

shown in Figure 1.1 (left), where it can be seen that the initial datum consists of two peaks, a negative and a positive one, oriented along the x axis. The evolution of the initial datum under the flow of the KPI Equation can be inferred from the sequence of snapshots of Figure 1.1. Initially, since the dispersion coefficient is small, the evolution is dominated by the nonlinear term, which induces a steepening of both the positive and negative parts of the solution, thus forming a shock front. As discussed above, when the solution approaches the breakup time of the dKP Equation, the wavefront gradients become strong enough that the dispersive term $\epsilon^2 u_{xxx}$, despite the small constant, can compensate the focusing effect of the nonlinearity, dispersing away some energy in the form of high frequency oscillations, shown in Figure 1.1

(center). In [58], the exact location of the singularity is given both for the positive and the negative part of the solution, and it is found that the positive part develops a singularity at $t = 0.222$, and the negative part at $t = 0.300$. Note that the plot in the center of Figure 1.1 corresponds to a time ($t = 0.77$) when both singularities have already appeared, leading to the development of the two fronts of dispersive shock waves. For even longer times, a second breaking in the form of wave instability takes place in the positive part of the solution. Such instability consists in a wave focusing leading to the appearance of very high, localised peaks which arrange in a triangular lattice (shown in Figure 1.1 (right)) and propagate away at high speed, while keeping their original shape. This orderly arrangement of lumps is the result of a complex interaction between lumps and dispersive shock waves.

A detailed numerical investigation of dispersive shock waves in the KP Equation is available in [7]. In the cited work, a step-like initial datum with parabolic front is evolved numerically, showing the emergence of line solitons with parabolic front. The results for the KP Equation are then compared with a two-dimensional version of the Benjamin–Ono Equation.

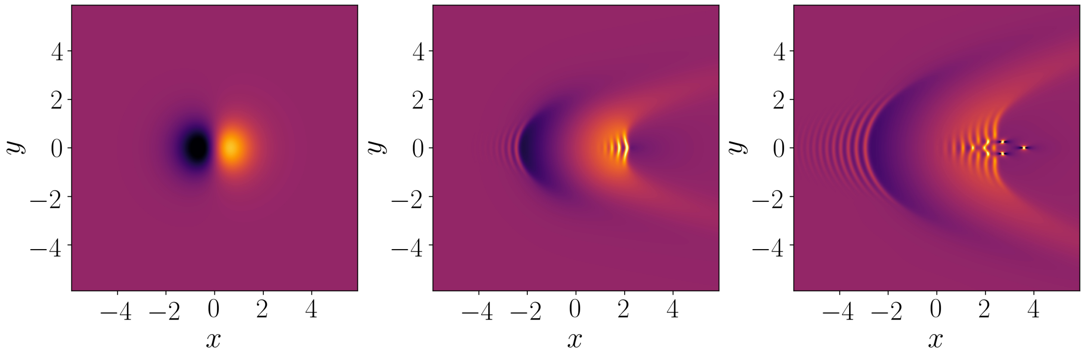


FIGURE 1.1: Snapshots taken from the time evolution of the initial datum (1.36) under the KPI flow with $\epsilon = 0.03$. For the meaning of the colours, we refer to the colorbar in Figure 1.2. On the left, the initial datum at $t = 0$. At the center ($t = 0.42$), the dispersive shock waves forming after the first breakup time are clearly visible both in the positive and in the negative part of the solution. On the right ($t = 0.77$), after the second breakup time the first lumps appear and interact with the dispersive shock waves arranging in a triangular lattice.

This behaviour is typical of the KPI Equation. For the KPII Equation, a highly oscillatory region still develops, but with an opposite curvature of the wavefront, and the wavefronts themselves are stable. A comparison of the KPI and KPII flow for the same value of ϵ and for the same initial datum is available in Figure 1.2 as pseudocolor plot and in Figure 1.3 as a surface plot.

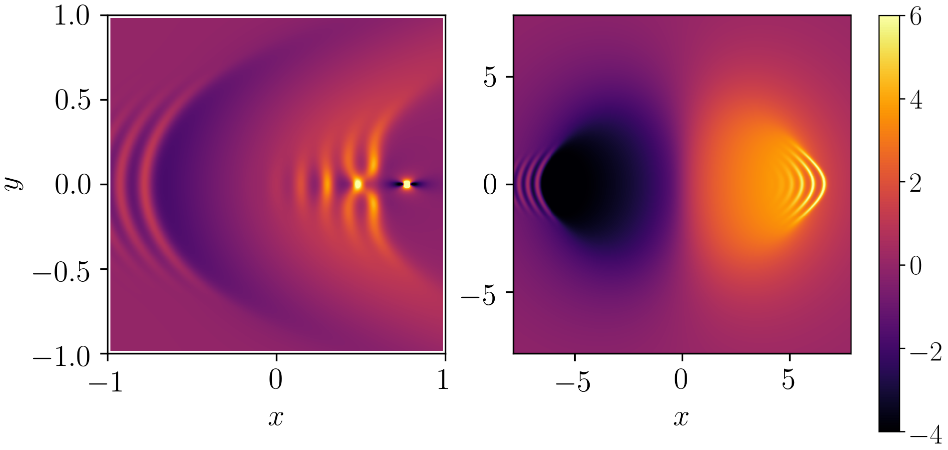


FIGURE 1.2: Evolution of the initial datum (1.36) for $\epsilon = 0.1$ at $t = 0.8$ for the KPI (left) and KPII (right) Equations. For a three-dimensional visualisation of these same plots, see Figure 1.3.

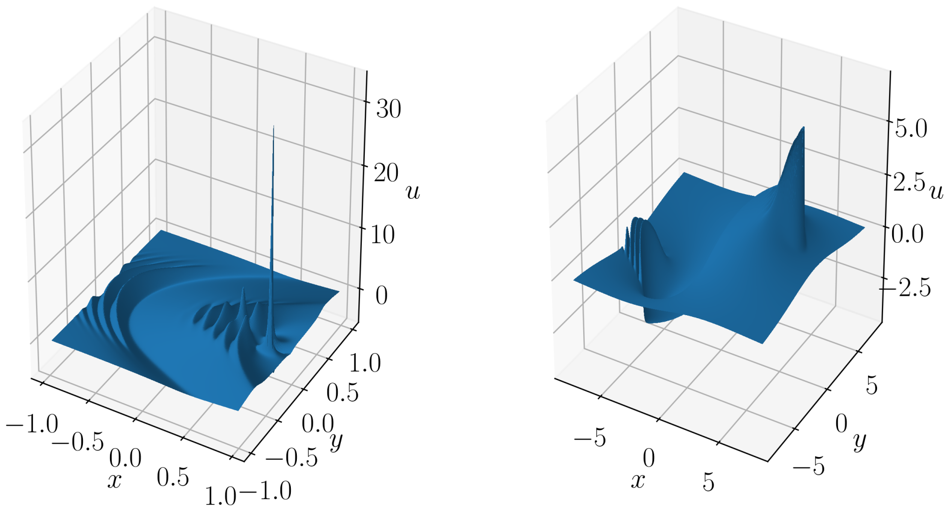


Figure 1.3: Snapshots of the solution to the KPI (left) and KPII (right) Equations at time $t = 0.8$ for the initial datum (1.36) and $\epsilon = 0.1$. In the KPI Equation the lump is clearly visible in the form of a very high and localised peak. This figure displays the same information shown with a pseudo-color plot in Figure 1.2.

1.2.3 New contributions

The purpose of this work is to study the propagation of dispersive shock waves in the KPI Equation for times longer than those considered in the existing literature. A possible interpretation of the long-time behaviour must rely on the analysis of the two main competing effects of dispersion and nonlinearity. After the dispersive shock waves have formed according to the mechanism summarised in Section 1.2.2, these may keep propagating in purely dispersive fashion, transporting energy away from the initial disturbance, thus leading to a motionless state for $t \to \infty$, or it may be the case that the nonlinear effects balance the dispersion by creating stable structures such as solitons in the Korteweg-De Vries Equation. A third possibility, studied in [84] for generalised KP Equations with a higher order nonlinearity is that the nonlinear term concentrates mass in a certain region of space sufficiently fast that the linear operator is not able to disperse away this buildup of mass, leading to blowup in finite time. Blowup is however ruled out for the standard KP Equation, since well posedness has been proved for all times, as discussed in Section 1.3.8. The generic nature of lump formation, and their arrangement in a orderly structure is however new and is the main result of this work.

The numerical results are complemented by the derivation of the modulation equations for the line soliton and lump solutions of the KP Equation. The modulation equations for the line solitons are viewed as an approximation for the much more complex modulation equations for the packet of dispersive shock waves, and the study of their properties justifies theoretically the breaking phenomenon. This result has been discovered by means of high quality numerical computations, and later described precisely by applying Whitham's method of wave modulations. Whitham Modulation Theory allows to describe the wavefronts appearing on the positive part of the solution after the first breakup as modulated line solitons. The modulation equations for the line soliton are elliptic, and it is known from the theory of elliptic systems that these generically develop a caustic singularity, which represents the focusing effect up to a critical time (second breakup time) when the lump detaches from the line soliton. Modulation equations are derived for the lump too, and the resulting system of conservation laws is also in this case elliptic. We refer to Section 2.6 for a possible interpretation of this fact.

For times longer than the second breakup time, Whitham's theory does not seem to be an effective tool. For this reason, we study numerically the formation of the triangular lattice of lumps, and we propose an analogy with the focusing nonlinear Schrödinger Equation, which allows to extend some rigorous asymptotic results to the KP Equation.

1.3 Properties of the KP Equation

We conclude the chapter by recalling some of the many important properties of the KP Equation. As a matter of fact, the KP Equation has a rich mathematical structure, so much so that it is impossible to sketch in a few pages, and the choice of topics that we succinctly treat here is to some extent a matter of personal taste. The topics briefly summarised in the following aim mainly towards two important properties of the KP Equation: its Lagrangian form and the existence of exact solutions. These two properties are at the basis of the development of modulation theory in Chapter 2. Some readable introductions to the general theory of the KP Equation are [36, 78] and [54, Ch. 1].

1.3.1 Reductions

The KP Equation can be interpreted as a generalisation of three important equations of Mathematical Physics: the Korteweg–de Vries (KdV) Equation, the Boussinesq Equation and the Zakharov–Kuznetsov Equation. In addition, its dispersionless version, sometimes also called Zabolotskaya–Khokhlov Equation, has received much attention in recent years.

Considering the KP Equation with unit dispersion in its evolutionary form:

$$u_t + uu_x + u_{xxx} + \alpha\partial_x^{-1}u_{yy} = 0, \tag{1.37}$$

it is simple to check that a solution not depending on y solves the KdV Equation:

$$u_t + uu_x + u_{xxx} = 0, \tag{1.38}$$

so that conversely any solution of the KdV Equation is also a solution of both the KPI and KPII Equations.

A stationary solution of the KPII Equation, $u_t = 0$, satisfies:

$$u_{yy} = -\left(\frac{u^2}{2} + u_{xx}\right)_{xx}, \tag{1.39}$$

and interpreting the y variable as a time, Equation (1.39) coincides with the celebrated Boussinesq Equation.

The dispersionless KP Equation (obtained by setting $\epsilon = 0$) is sometimes called the Zabolotskaya–Khokhlov Equation [115, 33]:

$$(u_t + uu_x)_x = u_{yy}. \tag{1.40}$$

This equation finds application as a model for wave propagation in nonlinear acoustics. As mentioned in Section 1.2.1, we remark that despite its name, Equation (1.40) actually is a dispersive equation.

Lastly, by considering the dispersionless KPI Equation and considering the left-moving wave (which amounts to changing the sign of the nonlinear term, as shown in [4, § 1.2]), yields the Zakharov–Kuznetsov Equation [121]:

$$u_{tx} = \left(\frac{u^2}{2}\right)_{xx} + u_{yy}, \tag{1.41}$$

which was introduced as a model for nonlinear acoustic waves in a plasma.

1.3.2 Lagrangian Structure

It is not possible to find a Lagrangian for for the KP Equation in the original variables. A similar obstruction for the KdV Equation was pointed out directly by Whitham [116]. As for the KdV Equation, the workaround consists in introducing a potential function φ such that:

$$\varphi_x = u. \tag{1.42}$$

The KP Equation, rewritten for the potential function becomes:

$$\varphi_{xt} + \epsilon\varphi_x\varphi_{xx} + \epsilon^2\varphi_{xxxx} + \alpha\varphi_{yy} = 0. \tag{1.43}$$

A possible Lagrangian for Equation (1.43) is:

$$L = \epsilon^2\varphi_t\varphi_x + \frac{\epsilon^3}{3}\varphi_x^3 - \epsilon^4\varphi_{xx}^2 + \epsilon^2\alpha\varphi_y^2, \tag{1.44}$$

and indeed it can be checked that the Euler–Lagrange Equations for the Lagrangian (1.44), which generically have the form:

$$\frac{\partial}{\partial t}L_{\varphi_t} + \frac{\partial}{\partial x}L_{\varphi_x} + \frac{\partial}{\partial y}L_{\varphi_y} - \frac{\partial^2}{\partial x^2}L_{\varphi_{xx}} = 0 \tag{1.45}$$

give the correct Equation (1.43) for φ.

Furthermore, the existence of a Lagrangian allows, among other things, to check for the conservation laws by means of the Noether Theorem. Namely, if it is possible to find a continuous group of symmetries (sometimes also called *one-parameter group of transformations*) for the Lagrangian, then Noether's Theorem ensures the existence of a conserved quantity associated to the symmetry group. For a detailed treatment and a proof of the Noether Theorem in Lagrangian and Hamiltonian Mechanics see [14, 73, 65].

1.3.3 Hamiltonian Structure

For an introduction to the Hamiltonian and Multi-Hamiltonian theory of many integrable systems, including the KP Equation, we refer to [19].

The KP Equation is an Hamiltonian equation, namely can be written in Hamiltonian form as:

$$u_t = \frac{\partial}{\partial x}\left(\frac{\delta\mathcal{H}}{\delta u}\right), \tag{1.46}$$

where $\mathcal{H}$ is an Hamiltonian functional and $\frac{\delta}{\delta u}$ denotes the Fréchet derivative with respect to u. For the KP Equation, the Hamiltonian takes the form:

$$\mathcal{H} = -\int_{\mathbb{R}^2} h \,\mathrm{d}x\,\mathrm{d}y := -\int_{\mathbb{R}^2} \frac{u^3}{6} - \frac{u_x^2}{2} + \frac{\alpha}{2}\left(\partial_x^{-1}u_y\right)^2 \mathrm{d}x\,\mathrm{d}y, \tag{1.47}$$

h being the Hamiltonian density. Written explicitly, the Fréchet derivative in this case can be expanded as:

$$\frac{\delta\mathcal{H}}{\delta u} = \frac{\partial h}{\partial u} + \frac{\partial}{\partial x}\frac{\partial h}{\partial u_x} + \frac{\partial}{\partial^{-1}x\partial y}\frac{\partial h}{\partial\left(\partial_x^{-1}u_y\right)}, \tag{1.48}$$

So that the equivalence between (1.46) and (1.18) can be verified:

$$\begin{aligned} u_t &= -\frac{\partial}{\partial x}\left(\frac{u^2}{2} + \frac{\partial}{\partial x}u_x + \alpha\frac{\partial}{\partial^{-1}x\partial y}\partial_x^{-1}u_y\right) \\ &= -uu_x - u_{xxx} - \alpha\partial_x^{-1}u_{yy}. \end{aligned} \tag{1.49}$$

The existence of a Hamiltonian structure is particularly useful for studying the existence of conserved quantitites. Indeed, a function $F(u)$ of u is a conserved quantity if:

$$\frac{\mathrm{d}}{\mathrm{d}t}F = \{\mathcal{H}, F\} = 0, \tag{1.50}$$

where the Poisson bracket $\{\cdot\}$ is defined as:

$$\{\mathcal{H}, F\} = \left\langle \frac{\delta F}{\delta u}, \frac{\partial}{\partial x}\frac{\delta\mathcal{H}}{\delta u} \right\rangle, \tag{1.51}$$

and the angled parentheses denote an inner product. The first conserved quantity, T_1, can be obtained by writing the KP Equation in the conservative form:

$$\frac{\partial}{\partial t}(u) + \frac{\partial}{\partial x}\left(\frac{u^2}{2}\epsilon^2 u_{xx} + \alpha\partial_x^{-2}u_{yy}\right) = 0, \tag{1.52}$$

which has the general structure:

$$\frac{\partial T_1}{\partial t} + \frac{\partial X_1}{\partial x} = 0, \tag{1.53}$$

with

$$T_1 = u \qquad\qquad X_1 = \frac{u^2}{2}\epsilon^2 u_{xx} + \alpha\partial_x^{-2}u_{yy}. \tag{1.54}$$

Assuming $X_{1,x}$ to be integrable over the real plane, it is possible to integrate Equation (1.52) to:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{\mathbb{R}^2} T_1 \,\mathrm{d}x\,\mathrm{d}y\right) = 0, \tag{1.55}$$

showing that the integral:

$$\int_{\mathbb{R}^2} u \,\mathrm{d}x\,\mathrm{d}y \tag{1.56}$$

is a conserved quantity, and u is the relative conserved density. Sometimes the relation (1.55) is referred to as a *mass conservation* property.

Similarly, multiplying the KP Equation by u gives the conservation law:

$$\frac{\partial}{\partial t}\left(\frac{u^2}{2}\right) + \frac{\partial}{\partial x}\left(\frac{u^3}{3} + \epsilon^2 u u_{xx} - \epsilon^2\frac{u_x^2}{2} + \alpha\partial_x^{-2}u_{yy}\right) = 0 \tag{1.57}$$

which gives the conserved quantity:

$$\int_{\mathbb{R}^2} u^2 \,\mathrm{d}x\,\mathrm{d}y. \tag{1.58}$$

This property is called *momentum conservation* and is equivalent to the conservation of the $L^2(\mathbb{R}^2)$ norm of the solution.

In a similar fashion, it is possible to prove the *energy conservation* law:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{\mathbb{R}^2}\left(-\frac{u^3}{6} + \frac{u_x^2}{2} + \frac{\alpha}{2}\left(\partial_x^{-1}u_y\right)^2\right)\mathrm{d}x\,\mathrm{d}y\right) = 0. \tag{1.59}$$

Note that the conserved density in Equation (1.59) is the same as the Hamiltonian density in Equation (1.47).

1.3.4 Symmetries and constraints

Some physically relevant symmetries of the KPI Equation are the following.

Translation invariance. If $u(x, y, t)$ is a solution, then $u(x + x_0, y + y_0, t + t_0)$ is a solution for all $x_0, y_0, t_0 \in \mathbb{R}$.

Reflection invariance. If $u(x, y, t)$ is a solution, then $u(-x, y, t)$ and $u(-x, -y, t)$ are solutions.

Galilean invariance. If $u(x, y, t)$ is a solution, then $c/6 + u(x - ct, y, t)$ is also a solution.

Scaling invariance. If $u(x, y, t)$ is a solution, then $\lambda^2 u(\lambda x, \lambda^2 y, \lambda^3 t)$ is a solution for any $\lambda > 0$.

Rotational invariance. If $u(x, y, t)$ is a solution, then $u(x - cy - \alpha c^2 t, y - 2\alpha ct, t)$ is a solution for any $c \in \mathbb{R}$.

As mentioned briefly in Section 1.3.3, for each of these symmetries there exists a corresponding conserved quantity.

In addition to the symmetries described above, the KP flow imposes several constraints on the solution, first obtained in [93]. In the evolutionary form (1.18), the antiderivative has sense if and only if the following constraint holds for any time:

$$\int_{\mathbb{R}} u \,\mathrm{d}x = 0. \tag{1.60}$$

This result is proved in [93] under the hypothesis that u is smooth and that u, u_t, as well as all the derivatives of u with respect to x and y tend to zero as $|x| \to \infty$. A second constraint follows from taking the derivative of (1.60) with respect to t and replacing the original equation (1.18) for u_t:

$$\int_{\mathbb{R}} \partial_x^{-1} u_{yy} \,\mathrm{d}x = 0. \tag{1.61}$$

Repeating one more time the same procedure, namely differentiating Equation (1.61) with respect to t and making use of (1.18) gives a third constraint:

$$\int_{\mathbb{R}} \left(\partial_x^{-2} u_{yyyy} - \frac{1}{2}(u^2)_{yy} \right) \mathrm{d}x = 0. \tag{1.62}$$

Similarly, by iterating this procedure it is possible to find an infinite number of constraints for the solutions of the KP Equation. In [8] it is shown that the infinite number of constraints discussed above is not required if the nonlocal term:

$$\partial_x^{-1} u_{yy} := \int_{-\infty}^{x} u(s,y,t)_{yy} \,\mathrm{d}s \tag{1.63}$$

is replaced by the symmetric form:

$$\partial_x^{-1} u_{yy} := \frac{1}{2} \left(\int_{-\infty}^{x} u(s,y,t)_{yy} \,\mathrm{d}s - \int_{x}^{\infty} u(s,y,t) \,\mathrm{d}s \right). \tag{1.64}$$

In the form of Equation (1.64), however, x integration and the limit for $t \to 0$ cannot be exchanged due to a discontinuity of the solution at $t = 0$, as proved in [20].

Even if the initial datum does not satisfy the constraints, the solution does satisfy them for any $t > 0$. In this sense, in the KP Equation there is an infinite speed of propagation. Furthermore, the solution develops immediately algebraic tails, so that even if the initial datum is in the Schwartz space of rapidly decreasing functions, the solution is no longer in the Schwartz space for any $t > 0$. More precisely, in [20] it is proved that even for an initial datum in the Schwartz space, the solution decays immediately as:

$$u(x,y,t) = \frac{c_0}{\sqrt{tx}|x|} \int_{\mathbb{R}^2} u \,\mathrm{d}x \,\mathrm{d}y + o(|x|^{-3/2}), \tag{1.65}$$

for $x \to \infty$, where c_0 is a constant. This fact does not contradict the momentum conservation property (1.58), which requires only that $u \in L^2(\mathbb{R}^2)$.

1.3.5 Lax pair representation

Given a differential equation in evolutionary form:

$$u_t = F(u), \tag{1.66}$$

it is sometimes possible to find two operators L, M such that the original Equation (1.66) is equivalent to the "Lax pair" representation:

$$\mathrm{L}_t = [\mathrm{L}, \mathrm{M}], \tag{1.67}$$

where $[\mathrm{A},\mathrm{B}] := \mathrm{AB}-\mathrm{BA}$ is the commutator of A and B. This representation was the starting point for the discovery of the Inverse Scattering Transform and of the infinitely many conserved quantities for the KdV Equation. A simple way to point out the importance of the representation (1.67) is by considering a finite dimensional analogue, where L is a time-dependent, diagonisable matrix whose spectrum does not depend on time. If L is decomposed canonically as:

$$\mathrm{L}(t)=\mathrm{S}^{-1}(t)\Lambda\mathrm{S}(t), \tag{1.68}$$

where $\Lambda=\mathrm{diag}(\lambda_1,\ldots,\lambda_n)$ is the matrix with the eigenvalues of L, which do not depend on time, then it is possible to show that the time derivative of L verifies a relation in the form of (1.67), for an appropriate choice of M. To see this, it is convenient to compute explicitly the time derivative of Equation (1.68):

$$\mathrm{L}_t=\mathrm{S}_t^{-1}\Lambda\mathrm{S}+\mathrm{S}^{-1}\Lambda\mathrm{S}_t. \tag{1.69}$$

An explicit expression for S_t^{-1} can be obtained by differentiating the identity $\mathrm{S}^{-1}\mathrm{S}=\mathrm{I}$:

$$\mathrm{S}_t^{-1}=-\mathrm{S}^{-1}\mathrm{S}_t\mathrm{S}^{-1}. \tag{1.70}$$

Replacing the expression just found (1.70) into Equation (1.69), it follows that:

$$\begin{aligned}
\mathrm{L}_t &= -\mathrm{S}^{-1}\mathrm{S}_t\mathrm{S}^{-1}\Lambda\mathrm{S}+\mathrm{S}^{-1}\Lambda\mathrm{S}_t && (1.71)\\
&= -\mathrm{S}^{-1}\mathrm{S}_t\mathrm{S}^{-1}\Lambda\mathrm{S}+\mathrm{S}^{-1}\Lambda\mathrm{S}\mathrm{S}^{-1}\mathrm{S}_t && (1.72)\\
&= -\mathrm{ML}+\mathrm{LM} && (1.73)\\
&= [\mathrm{L},\mathrm{M}], && (1.74)
\end{aligned}$$

with $\mathrm{M}=\mathrm{S}^{-1}\mathrm{S}_t$. There are at least two reasons for which the Lax pair form is of great practical importance. The first one is that if a Lax pair can be found, then the eigenvalues of L:

$$\mathrm{L}\psi_m=\lambda_m\psi_m \qquad \text{for } m=1,\ldots,n \tag{1.75}$$

are constants of the motion, $\frac{\mathrm{d}}{\mathrm{d}t}\lambda_m=0$. The second is that the time evolution equation for the eigenfunctions ψ_m:

$$\psi_{m,t}=\mathrm{M}\psi_m \tag{1.76}$$

is linear (but note that M depends on time).

The Lax pair representation for the KP Equation was derived in [38], and has the form:

$$\mathrm{L}_t=\sqrt{\alpha}[\mathrm{L},\mathrm{M}], \tag{1.77}$$

with operators:

$$\begin{aligned}
\mathrm{L} &= \epsilon^2\partial_{xx}-\frac{\sqrt{\alpha}}{\sqrt{3}}\partial_y+\frac{1}{6}u && (1.78)\\
\mathrm{M} &= -\partial_{xxx}-\frac{1}{4}u\partial_x-\frac{1}{6\sqrt{3}}\partial_x^{-1}u_y. && (1.79)
\end{aligned}$$

It is important to reckon that the operator multiplications in the commutator of the Lax pair are understood in the sense of Ordinary Differential Operators. As an example, the product between the first term of L and the second term of M in Equation (1.79) is:

$$6\partial_{xx}\left(-\mathrm{i}u\partial_x\right)=-6\mathrm{i}\left(u_{xx}\partial_x+2u_x\partial_{xx}+u\partial_{xxx}\right). \tag{1.80}$$

An equivalent interpretation for the Lax pair is as a compatibility condition for a set of two linear equations. To see this, let us define the operator $\widetilde{\mathrm{L}}$ as:

$$\widetilde{\mathrm{L}}=\epsilon^2\partial_{xx}+\frac{1}{6}u, \tag{1.81}$$

and consider the two linear equations:

$$\frac{\sqrt{\alpha}}{\sqrt{3}}\frac{\partial}{\partial y}\psi = \widetilde{\mathrm{L}}\psi \tag{1.82}$$

$$\frac{1}{4}\frac{\partial}{\partial t}\psi = \mathrm{M}\psi. \tag{1.83}$$

The flow of the two equations (1.83) commutes if:

$$\psi_{yt} = \psi_{ty}, \tag{1.84}$$

which in turn implies the following commutativity relation for the Lax operators:

$$\left[\frac{\sqrt{\alpha}}{\sqrt{3}}\frac{\partial}{\partial y} - \widetilde{\mathrm{L}}, \frac{1}{4}\frac{\partial}{\partial t} - \mathrm{M}\right] = 0, \tag{1.85}$$

which is equivalent to (1.77).

1.3.6 Hirota Bilinear Form

Hirota [63] has shown a remarkable way of constructing exact solutions for several equations, among which the KP Equation, that is particularly useful for the description of line solitons interactions in the KPII Equation. The main component of Hirota's method is a set of "nicely weighted functions" (this is the terminology used in [78]). A *nicely weighted function* $\tau = \tau(x, y, t)$ is any function that verifies the relations:

$$\tau_t = \tau_{xxx} \qquad \text{and} \qquad \tau_{xx} = \tau_y. \tag{1.86}$$

Hirota proved that if τ is a nicely weighted function, then

$$u = 12\partial_{xx}\log\tau \tag{1.87}$$

is a solution of the KP Equation. This can be seen by replacing the expression (1.87) in the original form of the KP Equation (1.17):

$$\partial_{xt}\partial_{xx}\log\tau + 12\partial_x\left(\partial_{xx}\log\tau\,\partial_x\partial_{xx}\log\tau\right) + \partial_{xxxx}\partial_{xx}\log\tau + \alpha\partial_{yy}\partial_{xx}\log\tau, \tag{1.88}$$

which is equivalent to:

$$\partial_{xt}\log\tau + 12\left(\frac{1}{2}\frac{\tau_x^4}{\tau^4} - \frac{\tau_x^2\tau_{xx}}{\tau^3} + \frac{1}{2}\frac{\tau_{xx}^2}{\tau^2}\right) + \partial_{xxxx}\log\tau + \alpha\partial_{yy}\log\tau = 0, \tag{1.89}$$

since:

$$\partial_x\left(\partial_{xx}\log\tau\,\partial_{xxx}\log\tau\right) = \partial_{xx}\left(\frac{1}{2}\frac{\tau_x^4}{\tau^4} - \frac{\tau_x^2\tau_{xx}}{\tau^3} + \frac{1}{2}\frac{\tau_{xx}^2}{\tau^2}\right). \tag{1.90}$$

After some straightforward computations, it is possible to obtain the following bilinear equation for τ:

$$\tau\tau_{xt} - \tau_x\tau_t + 3\tau_{xx}^2 - 4\tau_x\tau_{xxx} + \tau\tau_{xxxx} + \alpha\left(\tau\tau_{yy} - \tau_y^2\right) = 0, \tag{1.91}$$

where two integration constants have been set to zero. It can be checked that if τ is a nicely weighted function, then it is a solution of the bilinear Equation (1.91).

Moreover, any number of tau functions can be combined to produce a new solution of the KP Equation, via the "nonlinear superposition":

$$\tau_{\text{new}} = \mathrm{Wr}(\tau_1, \dots, \tau_n) = \det\begin{pmatrix} \tau_1 & \tau_2 & \dots & \tau_n \\ \partial_x\tau_1 & \partial_x\tau_2 & \dots & \partial_x\tau_n \\ \vdots & \vdots & \ddots & \vdots \\ \partial_x^{n-1}\tau_1 & \partial_x^{n-1}\tau_2 & \dots & \partial_x^{n-1}\tau_n \end{pmatrix}. \tag{1.92}$$

Hirota rewrote Equation (1.91) in a simpler way by introducing the *Hirota derivative* operator D:

$$\mathrm{D}_x^n(f \cdot g) = \left(\partial_x - \partial_{x'}\right)^n f(x)g(x')\Big|_{x'=x}, \tag{1.93}$$

and similarly for the other variables. Noting that:

$$\mathrm{D}_x\,\mathrm{D}_y(\tau \cdot \tau) = 2\left(\tau\tau_{xt} - \tau_x\tau_t\right) \tag{1.94}$$

$$\mathrm{D}_x^4(\tau \cdot \tau) = 2\left(3\tau_{xx}^2 - 4\tau_x\tau_{xxx} + \tau\tau_{xxxx}\right) \tag{1.95}$$

$$\mathrm{D}_y^2(\tau \cdot \tau) = 2\left(\tau\tau_{yy} - \tau_y^2\right), \tag{1.96}$$

then Equation (1.91) can be rewritten in the equivalent, but simpler looking form:

$$\left(\mathrm{D}_t\,\mathrm{D}_x + \mathrm{D}_x^4 + \alpha\,\mathrm{D}_y^2\right)(\tau \cdot \tau) = 0. \tag{1.97}$$

Hirota's bilinear operator enjoys some remarkable properties that considerably simplify the solution process for the Bilinear Equation. A typical application of the Hirota bilinear Equation is to find multi-soliton solutions to any equation (not necessarily integrable) which admits solitonic solutions. For an example of this approach to multisolitonic solutions of the KdV Equation see [36, Ch. 5].

1.3.7 Exact solutions

A typical feature of integrability is the possibility of exhibiting many exact solutions. The KPI Equation is no exception, and in this section we recall a few exact solutions, some of which will be useful in the following.

Line solitons

As discussed in Section 1.3.1, any solution of the KdV Equation that does not depend on y is also a solution of the KP Equation. A landmark solution of the KdV Equation is the nonlinear traveling wave solution, that consists in a localised perturbation that propagates without dispersing. By seeking for a traveling wave solution of the KdV Equation, namely a solution in the form:

$$u(x,t) = aF(bx + ct) \tag{1.98}$$

after some computations which are recalled in a more general setting in Section 2.2, gives the soliton solution:

$$u(x,t) = a\,\mathrm{sech}^2\left(\frac{\sqrt{a}}{2\sqrt{3}}x - \frac{a^{3/2}}{6\sqrt{3}}t\right), \tag{1.99}$$

which describes a right-moving positive soliton. Extending this approach to the KP Equation, with a slightly more general traveling wave ansatz:

$$u(x,y,t) = aF(bx + ly + ct), \tag{1.100}$$

yields to:

$$u(x,y,t) = a\,\mathrm{sech}^2\left(\frac{\sqrt{a}}{2\sqrt{3}}x + ly - \frac{a^2 + 36\alpha l^2}{6\sqrt{3a}}t\right). \tag{1.101}$$

Equation (1.101) is the expression of the *line soliton* for the KP Equation. We remark that Equation (1.101) can be obtained by applying the rotational transformation described in Section 1.3.4 to the "straight" soliton of Equation (1.99). From this equation, it is clear that the direction of propagation of the line soliton can be chosen freely by specifying l, however the speed of the soliton depends on the direction of propagation. For the KPI Equation ($\alpha = -1$), the more the direction of propagation aligns with x ($l \to 0$), the faster the soliton's speed. The converse statement holds for the KPII Equation ($\alpha = 1$). Solitons are well known for interacting in a way that preserves their shape and speed, affecting only their phase. Moreover, a generic localised initial datum will evolve in a certain number of solitons traveling to the right with different amplitudes (and thus speeds), plus some low-amplitude radiation traveling to the left. This behaviour was discovered in [118], and a similar result exists for line solitons in the KPII Equation [86]. As a matter of fact, both the KPI and the KPII Equations have multisoliton solutions that are not decaying at infinity for some direction in the (x, y) plane.

In [83], it is shown numerically that the line soliton in the KPI Equation is unstable with respect to transverse perturbations: for sufficiently long times there is a wave breaking leading to the formation of lump solutions. The instability observed numerically in [69, 70, 83] mimics the exact solution given in [107] in terms of tau functions. Other instability results for line solitons in the KPI Equation are shown in [13] and [122], where a focusing behaviour is shown in dimension 2 and 3, and in a more general water wave context in [110].

Lumps

Lump solutions were obtained first in [3], and further investigated in [112] and [98], where also the interaction between lumps is studied. Lumps are localised peaks which propagate conserving their shape, for this reason they are often called "two-dimensional solitons", and exist only for the KPI Equation. The KPII Equation does not seem to admit two-dimensional, localised, soliton solutions. The KPI Equation admits an infinite number of rational, two-dimensional soliton solutions. The simplest lump solution is a ratio between two polynomials in x, y, t of degree 2 and 4:

$$u(x,y,t) = 24\frac{-\left(x+ay+(a^2-3b^2)t\right)^2+3b^2\left(y+2at\right)^2+\frac{1}{b^2}}{\left(\left(x+ay+(a^2-3b^2)t\right)^2+3b^2\left(y+2at\right)^2+\frac{1}{b^2}\right)^2}, \tag{1.102}$$

with a and b arbitrary constants. A wireframe plot of this solitonic solution is shown in Figure 1.4 (left), where the localised nature of the lump is clearly visible. It is simple to check that the location of the maximum of a lump solution has coordinates:

$$x = (a^2+3b^2)t \qquad\qquad y = -2at. \tag{1.103}$$

A possible way of deriving lump solutions is by considering a traveling wave polynomial ansatz for the tau function:

$$\tau(x,y,t) = a_0(x+ct)^2 + a_1y^2 + a_2. \tag{1.104}$$

Inserting the ansatz (1.104) in the Hirota Equation (1.91) with $\alpha = -1$ gives:

$$\begin{aligned} 12a_0^2 - 2a_1a_2 + 2a_0a_2c - 2a_0c^2\left(a_1+c\right)t^2 - 4a_0c\left(a_1+a_0c\right)tx& \\ -2a_0\left(a_1+a_0c\right)x^2 + 2a_1\left(a_1+a_0c\right)y^2 = 0&, \end{aligned} \tag{1.105}$$

which in turn introduces the following constraints for a_0, a_1, a_2, c:

$$\begin{cases} 6a_0^2 - a_1a_2 + a_0a_2c = 0 \\ a_1 + c = 0 \\ a_1 + a_0c = 0. \end{cases} \tag{1.106}$$

Taking a_2 as a free variable, system (1.106) gives:

$$a_0 = 1 \qquad\qquad a_1 = \frac{3}{a_2} \qquad\qquad c = -\frac{3}{a_2}. \tag{1.107}$$

Renaming $a_2 = \frac{1}{b^2}$ gives the following expression for the tau function:

$$\tau(x,y,t) = (x-3b^2t)^2 + 3b^2y^2 + \frac{1}{b^2}, \tag{1.108}$$

which in turn leads to the lowest order lump solution for u:

$$u(x,y,t) = 12\partial_{xx}\log\tau(x,y,t) = 24\frac{-(x-3b^2t)^2+3b^2y^2+\frac{1}{b^2}}{\left((x-3b^2t)^2+3b^2y^2+\frac{1}{b^2}\right)^2}. \tag{1.109}$$

This is just a special case with $a = 0$ of the general lump solution of degree 2, Equation (1.102).

Zaitsev's solution

A traveling wave solution which represents a sort of hybrid between line solitons and lumps was discovered by Zaitsev in [119]. The expression of Zaitsev's solution is:

$$u(\xi, y) = 2\alpha^2 \frac{1 - \beta \cosh(\alpha\xi) \cos \delta y}{(\cosh \alpha\xi - \beta \cos \delta y)^2}, \tag{1.110}$$

with

$$\xi = x - ct \qquad c = \alpha^2 \frac{4 - \beta^2}{1 - \beta^2} \qquad \delta = \alpha^2 \sqrt{\frac{3}{1 - \beta^2}}. \tag{1.111}$$

As for the line soliton, the solution consists of an unbounded wavefront along a ray parallel to the y axis, rapidly decreasing along the direction orthogonal to the wavefront. The wave profile however changes along y in a periodic fashion, forming an ordered chain of lump-like crests. A visualisation of Zaitsev's solution is available in Figure 1.4 (right).

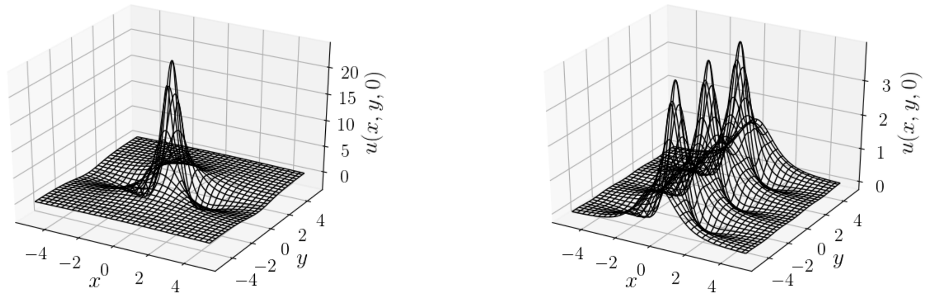


FIGURE 1.4: Wireframe plots of the lump solution (left) of Equation (1.102) with parameters $a = 0, b = 1$, and of Zaitsev's solution (right) of Equation (1.110) with parameters $\alpha = 1, \beta = 1/2$.

Theta functions

The algebro-geometric method for the production of exact solutions to integrable equations was shown first by Krichever in [89, 88], and for an introductory treatment we refer also to [40, 43].

Given a polynomial f in two variables $z, w \in \mathbb{C}$ with complex coefficients, consider the algebraic curve of the form:

$$\Gamma = \{(z, w) \in \mathbb{C}^2 \ : \ f(z, w) = 0\}. \tag{1.112}$$

The algebraic curve Γ is a one-complex dimensional analytic manifold, and as such a *Riemann Surface*. In the following, we will suppose that the Riemann Surface Γ has genus g. Let a_i and b_i, for $i = 1, \dots, g$ be a basis for the homology cycles of Γ (see Figure 1.5), and let ω_i be a basis for the holomorphic differentials (or differentials of the first kind) on Γ.

When Γ is a elliptic or hyperelliptic Riemann Surface,

$$f(z, w) = w^2 - P_{2g+1}(z), \tag{1.113}$$

with $P_{2g+1}(z)$ a polynomial of degree $2g+1$ in the z variable only, the holomorphic differentials have a particularly simple form in local coordinates:

$$\omega_i = \frac{z^{i-1}}{w} \mathrm{d}z = \frac{z^{i-1}}{\sqrt{P_{2g+1}(z)}} \mathrm{d}z \qquad \text{for } i = 1, \dots, g. \tag{1.114}$$

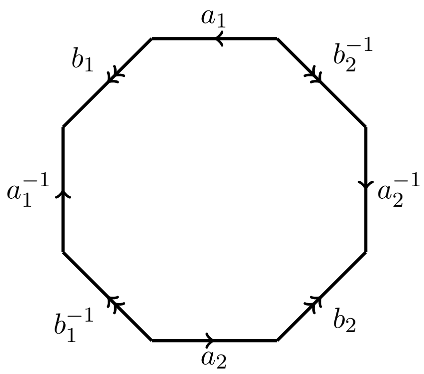


FIGURE 1.5: Basis cycles for the homology of a torus with genus 2.

We assume that the holomorphic differentials are normalised as:

$$\oint_{a_i} \omega_j = \delta_{ij}. \tag{1.115}$$

Let us denote with A and B the matrices defined as:

$$A_{ij} = \oint_{a_i} \omega_j \qquad\qquad B_{ij} = \oint_{b_i} \omega_j, \tag{1.116}$$

where B_{ij} is called the *period matrix* of Γ. This matrix has many important properties, notably is symmetric with negative definite real part (see e.g. [43] for a proof).

We now define the theta function associated to Γ by the following Fourier series:

$$\vartheta : \mathbb{C}^g \to \mathbb{C} \qquad\qquad \vartheta(\boldsymbol{z}|B) = \sum_{\boldsymbol{k}\in\mathbb{Z}^g} \exp(2\pi \mathrm{i}(\frac{1}{2}\boldsymbol{k}^T B\boldsymbol{k} + \boldsymbol{k}^T\boldsymbol{z})), \tag{1.117}$$

which has the following periodicity property:

$$\vartheta(\boldsymbol{z}+\boldsymbol{k}+B\boldsymbol{m}|B) = \exp(-2\pi \mathrm{i}(\frac{1}{2}\boldsymbol{m}^T B\boldsymbol{m} + \boldsymbol{m}^T\boldsymbol{z}))\vartheta(\boldsymbol{z}|B), \qquad \forall\, \boldsymbol{k},\, \boldsymbol{m},\, \boldsymbol{z} \in \mathbb{Z}^g. \tag{1.118}$$

Theta functions on Riemann Surfaces can be used to give explicit solutions to the KP Equation as follows:

$$u(x,y,t) = c + 2\partial_{xx}\log\vartheta(kx+ly+\omega t|B), \tag{1.119}$$

where c is an arbitrary constant, and k, l, ω can be reconstructed from the periods of the holomorphic differentials. In Equation (1.119), the wave speed ω should not be confused with the basis elements for the holomorphic differentials. We refer e.g. to [43, § 3.2] for the details on this construction.

Connection with Painlevé II

A class of self-similar solutions can be found [36] by imposing the ansatz:

$$u(x,y,t) = t^n F(\xi) \qquad\qquad \xi = xt^p + \lambda y^2 t^q. \tag{1.120}$$

Replacing the ansatz (1.120) in the KP Equation leads to the expression:

$$\begin{aligned}\left[(n+p)t^p + 2\alpha\lambda t^{q+1}\right] F_\xi + t^{n+2p+1}F_\xi^2 + \left[pt^{2p}x + q\lambda y^2 t^{p+q}\right] FF_{\xi\xi}\\ +4\alpha\lambda^2 y^2 t^{2q+1} FF_{\xi\xi} + t^{n+2p+1}FF_{\xi\xi} + t^{4p+1}F_{\xi\xi\xi\xi} = 0.\end{aligned} \tag{1.121}$$

For a similarity solution, all the terms in Equation (1.121) must scale equally with time, leading to the relations between the exponents:

$$p = q+1 = n+2p+1 = 4p+1, \tag{1.122}$$

with solution:

$$n = -\frac{2}{3} \qquad p = -\frac{1}{3} \qquad q = -\frac{4}{3}. \tag{1.123}$$

With the exponents known, Equation (1.121) takes the form:

$$(2\alpha\lambda - 1)F_\xi + F_\xi^2 + \left(-\frac{1}{3}\xi - \lambda y^2 t^{-4/3}\right) F_{\xi\xi} + 4\alpha\lambda^2 y^2 t^{-4/3} F_{\xi\xi} + F F_{\xi\xi} + F_{\xi\xi\xi\xi} = 0. \tag{1.124}$$

In Equation (1.124), the requirement that all the terms scale equally with respect to time leads to the following equation for λ:

$$4\alpha\lambda^2 - \lambda = 0, \tag{1.125}$$

with solution $\lambda = \frac{1}{4\alpha}$. Replacing this value of λ in Equation (1.124) gives:

$$F_{\xi\xi\xi\xi} + F F_{\xi\xi} + F_\xi^2 - \frac{1}{3}\xi F_{\xi\xi} - \frac{1}{2}F_\xi = 0. \tag{1.126}$$

Equation (1.126) can be immediately integrated to:

$$F_{\xi\xi\xi} + F F_\xi - \frac{1}{3}\left(\xi F_\xi - F\right) - \frac{1}{2}F + A = 0, \tag{1.127}$$

where A is an arbitrary constant that can be set to zero with the condition that $F, F_\xi, \dots \to 0$ as $|\xi| \to \infty$. Multiplying Equation (1.127) by the integrating factor F and integrating leads to:

$$F F_{\xi\xi} - \frac{1}{2}F_\xi^2 + \frac{1}{3}F^3 - \frac{1}{6}\xi F^2 + B = 0, \tag{1.128}$$

where B is an arbitrary constant that can be set to zero by the same decreasing conditions for F and its derivatives that were imposed for the first integration constant, A. With the substitution $F = V^2$, Equation (1.128) becomes:

$$V_{\xi\xi} - \frac{1}{12}\xi V + \frac{1}{6}V^3 = 0, \tag{1.129}$$

which is the Painlevé II Equation. Since Equation (1.129) does not depend on α, it follows that solutions in terms of the Painlevé II transcendent exist for both the KPI and KPII Equations.

1.3.8 Well posedness

For a review of well-posedness and regularity results for the KP and generalised KP Equations, see [83]. Here we limit ourselves to sketching the relevant well-posedness results for the KPI and KPII Equations on the two-dimensional torus $\mathbb{T}^2$, providing references to the appropriate literature for the details.

The global well-posedness of the KPII Equation was proved in [22] for initial data in $H^s(\mathbb{T}^2)$ for any $s \geq 0$ (which includes, in particular, $L^2(\mathbb{T}^2)$). Bourgain's proof consists in setting up the following Picard fixed point iteration:

$$\left(\partial_t + \partial_{xxx} + \alpha\partial_x^{-1}\partial_{yy}\right) u^{(1)} = 0 \qquad u^{(1)}(x, y, 0) = u_0(x, y) \tag{1.130}$$

$$\left(\partial_t + \partial_{xxx} + \alpha\partial_x^{-1}\partial_{yy}\right) u^{(k+1)} = -\partial_x \left(u^{(k)}\right)^2 \qquad u^{(k+1)}(x, y, 0) = u_0(x, y), \tag{1.131}$$

which is based on a linear-nonlinear splitting, and takes advantage of the fact that the linear problems in (1.131) can be solved explicitly by Fourier series. Introducing an appropriate norm, it is possible to to prove the existence of a local solution to the iterative scheme (1.131) by means of Banach–Caccioppoli's contraction theorem. The fixed point method guarantees only local well posedness (i.e. for small times), nevertheless a global result can be obtained by repeatedly applying the contraction principle, which ensures the existence of a sequence of local solutions, defined on a relative sequence of overlapping time intervals. The conservation of the L^2 norm ensures the existence of a lower bound on the length of the successive intervals, thus the sequence of local solutions can be extended to arbitrary times. In the same paper [22], Bourgain extended the global well-posedness proof for the KPII Equation to the whole $\mathbb{R}^2$. Bourgain's proof cannot be adapted to the KPI Equation. Instead, in [72] global well posedness in $\mathbb{T}^2$ is shown under the hypothesis that the initial data lies in the space:

$$Z^2(\mathbb{T}^2) := \{u \,:\, \|u\|_{Z^2} < \infty \text{ and } \widehat{u}(0,l) = 0 \; \forall\, l \in \mathbb{Z}\setminus\{0\}\}, \tag{1.132}$$

with norm:

$$\|u\|_{Z^2} := \|\widehat{u}(j,l)\left(1 + |j|^2 + |l/j|^2\right)\|_{\ell^2(\mathbb{Z}^2)}. \tag{1.133}$$

In addition, the authors of [72] also show that such solution is in $C(\mathbb{R}, Z^2(\mathbb{T}^2)) \cap C^1(\mathbb{R}, H^{-1}(\mathbb{T}^2))$. Global well posedness of the KPI Equation in $\mathbb{R}^2$ is proved in [102] for initial data in the Sobolev space with norm:

$$\|u\|_Z = \|u\|_{L^2} + \|u_{xxx}\|_{L^2} + \|u_y\|_{L^2} + \|u_{xy}\|_{L^2} + \|\partial_x^{-1} u_y\|_{L^2} + \|\partial_x^{-2} u_{yy}\|_{L^2}, \tag{1.134}$$

and in [72] for initial data in the Sobolev space defined by the energy norm:

$$\|u\|^2_{\mathrm{E}^1} = \|u\|^2_{L^2} + \|u_x\|^2_{L^2} + \|\partial_x^{-1} u_y\|^2_{L^2}. \tag{1.135}$$

Chapter 2

Modulation Equations for the KP Equation

Whitham modulation theory is a powerful framework for the analytical description of wave propagation in dispersive PDEs.

This topic, first introduced by Whitham in [117] for the KdV Equation, has found wide application in the description of wave modulation in the NLS Equation [50, 105], Camassa–Holm Equation [2, 1]. For an introduction to Whitham modulation theory, the most valuable source is still the book by Whitham [116], and some more recent review works are [76, 59, 46].

2.1 Whitham Modulation Theory

Whitham's Averaging Method is an asymptotic technique suitable for the asymptotic description of modulated traveling waves. Here the term modulation refers to the fact that some relevant wave parameters (amplitude, speed, direction of propagation) of a certain carrier wave change slowly in space or in time. There is no rigorous definition of how slow a modulated wave must be, and sometimes Whitham's method gives good results even when the modulation takes place in just a couple of wavelengths.

Whitham's modulation theory unfolds in two steps:

1. the search for an exact traveling wave solution to the original equation;
2. the asymptotic description of wave parameters' modulation by finding their evolution equations;

followed, if possible, by the solution of the modulation equations.

The second step can be greatly simplified if the original equation can be expressed in Lagrangian form. More precisely, if the original equation is the Euler–Lagrange equation for some Lagrangian, then the modulation equations too will be the Euler–Lagrange equations for some other Lagrangian functional. The Lagrangian for the modulated parameters is related to the original Lagrangian by a clever averaging procedure.

The search for a traveling wave solution is in principle quite straightforward. For a PDE in one space dimension, the typical form of a traveling wave is:

$$\eta(x,t) = af(kx - \omega t), \tag{2.1}$$

for some unknown periodic function f. This ansatz is then plugged in the original equation, and apart from technical difficulties an explicit (or sometimes implicit, as in the Camassa–Holm Equation [1]) form for f can be obtained.

Whitham's approach stems from the observation that the functional form (2.1) can be a good approximation to other classes of solutions to the original equation, not expressible in the form (2.1), provided that the parameters a, k, ω change slowly with space and time. This latter requirement can be formalized by introducing a “fast” variable[1]:

$$\theta = \frac{1}{\epsilon}\left(kx - \omega t\right) \tag{2.2}$$

and considering the following formal ansatz:

$$\eta(\theta, x, t) = a(x,t) f(\theta), \qquad \theta = \frac{1}{\epsilon}\left(k(x,t)x - \omega(x,t)t\right). \tag{2.3}$$

Computing the Lagrangian for the ansatz (2.3) leads to an expression of the form:

$$L(\theta, x, t) = F(\eta, \eta_a a_x, \eta_a a_t, \eta_k k_x, \eta_k k_y, \dots, \frac{1}{\epsilon}\eta_\theta, \dots) \tag{2.4}$$

where F denotes some functional dependence, which in general depends both on the fast and the slow variables. It is worth noting that, as a consequence of the functional form of the ansatz for η, the derivatives with respect to the fast variables are of order $O(1/\epsilon)$, while the derivatives with respect to the slow variables are of order $O(1)$. Indeed:

$$\frac{\mathrm{d}\eta}{\mathrm{d}x} = \eta_x + \frac{1}{\epsilon}\eta_\theta \theta_x. \tag{2.5}$$

As a result, in the limit $\epsilon \to 0$, the derivatives with respect to the slow variables can be neglected when computing the Lagrangian.

Since the parameters a, k, ω depend only on the slow variables, it would be interesting to derive a Lagrangian which depends only on the slow variables, so that the Euler–Lagrange equations for such Lagrangian would provide the modulation equations for the parameters. This was achieved by Whitham by introducing an averaged Lagrangian:

$$\overline{L} = \frac{1}{2\pi}\int_0^{2\pi} L \,\mathrm{d}\theta, \tag{2.6}$$

where the integral extends over a period that is supposed to be 2π in Equation (2.6).

The least action principle is then imposed on the averaged Lagrangian:

$$\delta \int \overline{L}\,\mathrm{d}t = 0. \tag{2.7}$$

Variations with respect to the amplitude are computed as:

$$\delta a: \qquad \overline{L}_a := \frac{\partial \overline{L}}{\partial a} = 0, \tag{2.8}$$

where the subscript notation is used to denote partial differentiation. Equation (2.8) results in an algebraic constraint for the parameters called *dispersion relation*. Variations with respect to the phase give:

$$\delta\theta: \qquad \frac{\partial}{\partial t}\overline{L}_\omega - \frac{\partial}{\partial x}\overline{L}_k = 0, \tag{2.9}$$

which provides the evolutionary PDE for the parameters. The modulation equations (2.8), (2.9) form a system of conservation laws, regardless of the nature of the initial equation for u.

2.2 Traveling wave solutions for the KP Equation

In this section, the first step in the theory of modulations sketched in Section 2.1, namely the search for traveling wave solutions, is pursued specifically for the KP Equation. However, instead of seeking directly for a traveling wave solution to the original form of the KP Equation,

---

[1] with the expression “fast variable”, we stress the fact that since $\epsilon \ll 1$, even if the original (or “slow”) variables x, t change slowly, θ changes rapidly.

for the Lagrangian averaging procedure it is convenient to seek for traveling wave solutions of the Potential Equation (1.43).

The most general form for a traveling wave potential is:

$$\varphi = \psi + \phi(\theta), \tag{2.10}$$

with ψ and θ in the form:

$$\psi = \beta_1 x + \beta_2 y - \gamma t \qquad \theta = \frac{1}{\epsilon}\left(kx + ly - \omega t\right), \tag{2.11}$$

and ϕ is a periodic function of period 2π. The term ψ is required to allow for modulations of the average value of u, and in the following it will be clear that the original wave parameters a, k, l, ω depend on ψ for their modulation. The traveling wave potential (2.11) gives origin to a traveling wave solution $u = \eta(\theta)$ by:

$$\eta(\theta) = \epsilon\varphi_x = \beta_1 + k\phi_\theta. \tag{2.12}$$

From Equation (2.12) it follows that

$$\phi_\theta = \frac{1}{k}\left(\eta - \beta_1\right), \tag{2.13}$$

which in turn gives the relations:

$$\epsilon\varphi_y = \beta_2 + \frac{l}{k}\left(\eta - \beta_1\right) \qquad \epsilon\varphi_t = -\gamma - \frac{\omega}{k}\left(\eta - \beta_1\right) \tag{2.14}$$

$$\epsilon^2\varphi_{xx} = k^2\phi_{\theta\theta} = k\eta_\theta \qquad \epsilon^2\varphi_{xt} = -\omega\eta_\theta \tag{2.15}$$

$$\epsilon^2\varphi_{yy} = \frac{l^2}{k}\eta_\theta \qquad \epsilon^4\varphi_{xxxx} = k^3\eta_{\theta\theta\theta}, \tag{2.16}$$

that will prove useful in the following. Imposing that φ solves the KP Equation for the potential:

$$\varphi_{xt} + \epsilon\varphi_x\varphi_{xx} + \epsilon^2\varphi_{xxxx} + \alpha\varphi_{yy} = 0, \tag{2.17}$$

yields:

$$-\omega\eta_\theta + k\eta\eta_\theta + k^3\eta_{\theta\theta\theta} + \alpha\frac{l^2}{k}\eta_\theta = 0. \tag{2.18}$$

This equation can be integrated immediately to:

$$k^2\eta_{\theta\theta} = -\frac{1}{2}\eta^2 + \left(\frac{\omega}{k} - \alpha\frac{l^2}{k^2}\right)\eta + \frac{A}{6}, \tag{2.19}$$

with A an integration constant. Multiplying Equation (2.19) by the integrating factor η_θ and integrating gives:

$$3k^2\eta_\theta^2 = -\eta^3 + V\eta^2 + A\eta + B, \tag{2.20}$$

with

$$V = 3\left(\frac{\omega}{k} - \alpha\frac{l^2}{k^2}\right), \tag{2.21}$$

and B an arbitrary integration constant. It is interesting to note that the Equation (2.20) for the traveling wave does not depend on ϵ. Clearly, Equation (2.20) has a real solution only if η is on a positive branch of the cubic on the right hand side. Supposing that the cubic has three roots e_1, e_2, e_3, ordered as $e_1 > e_2 > e_3$ (see Figure 2.1), then a traveling wave solution exists for $e_2 \leq \eta \leq e_1$. In this case, the right hand side of Equation (2.20) can be written as:

$$-\eta^3 + V\eta^2 + A\eta + B = -(\eta - e_1)(\eta - e_2)(\eta - e_3), \tag{2.22}$$

so that:

$$\begin{aligned} e_1 + e_2 + e_3 &= V \\ e_1 e_2 + e_2 e_3 + e_3 e_1 &= -A \\ e_1 e_2 e_3 &= B, \end{aligned} \tag{2.23}$$

which makes clear the fact that the roots e_1, e_2, e_3 contain the same information as V, A, B.

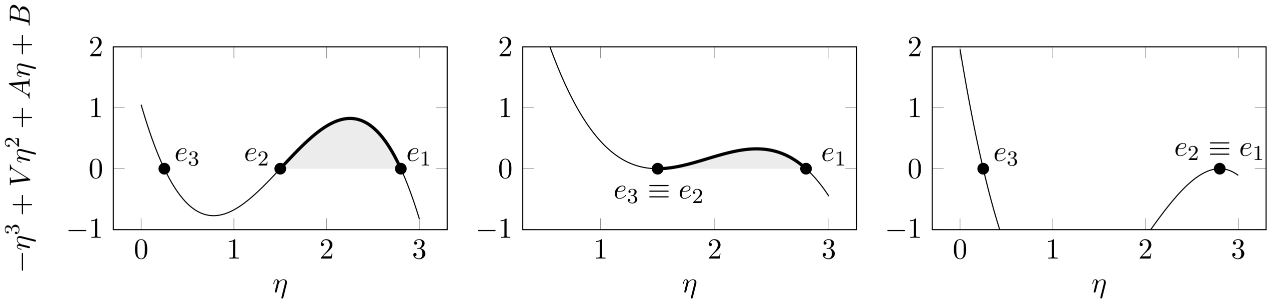


FIGURE 2.1: Cubics at the right hand side of Equation (2.20) allowing for traveling wave solutions. Only three cases are possible: $e_1 > e_2 > e_3$ (left), $e_1 > e_2 \equiv e_3$ (center), $e_1 \equiv e_2 > e_3$ (right).

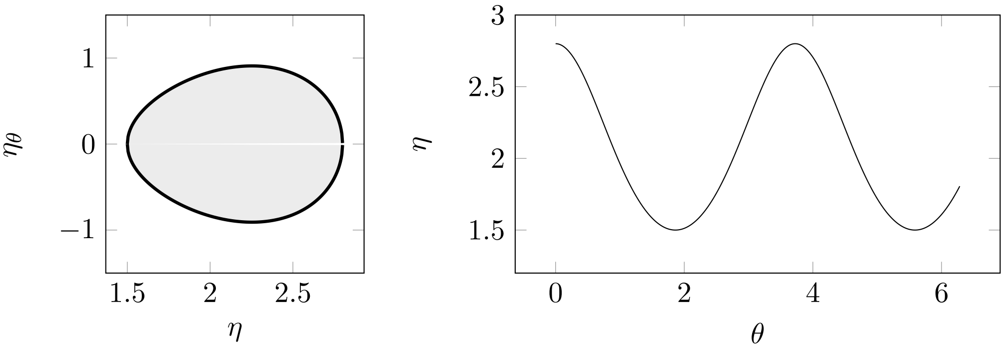


FIGURE 2.2: Parametric plot of η and η_θ for the case $e_3 < e_2 < e_1$ (left) and the corresponding traveling wave solution $\eta(\theta)$ (right). This corresponds to the picture on the left of Figure 2.1. The shaded area is proportional to the variable W defined in Equation (2.43).

Provided that $e_2 \le \eta \le e_1$, Equation (2.20) can be rewritten formally as:

$$\sqrt{3}k \frac{\mathrm{d}\eta}{\sqrt{-(\eta - e_1)(\eta - e_2)(\eta - e_3)}} = \mathrm{d}\theta. \tag{2.24}$$

Equation (2.24) can be integrated over a period to:

$$2\sqrt{3}k \int_{e_2}^{e_1} \frac{\mathrm{d}\eta}{\sqrt{-(\eta - e_1)(\eta - e_2)(\eta - e_3)}} = \oint \mathrm{d}\theta = 2\pi. \tag{2.25}$$

The integral on the left hand side of Equation (2.25) can be expressed as:

$$\int_{e_2}^{e_1} \frac{\mathrm{d}\eta}{\sqrt{-(\eta - e_1)(\eta - e_2)(\eta - e_3)}} = 2\frac{\mathrm{K}(m)}{\sqrt{e_1 - e_2}} \qquad m = \frac{e_1 - e_2}{e_1 - e_3}, \tag{2.26}$$

where $\mathrm{K}(m)$ is the complete elliptic integral of the first kind, and m is the parameter of the elliptic integral. Combining Equation (2.26) with Equation (2.25) gives an expression for the wave number k in terms of the roots e_1, e_2, e_3:

$$k = \frac{\pi}{2\sqrt{3}} \frac{\sqrt{e_1 - e_3}}{\mathrm{K}(m)}. \tag{2.27}$$

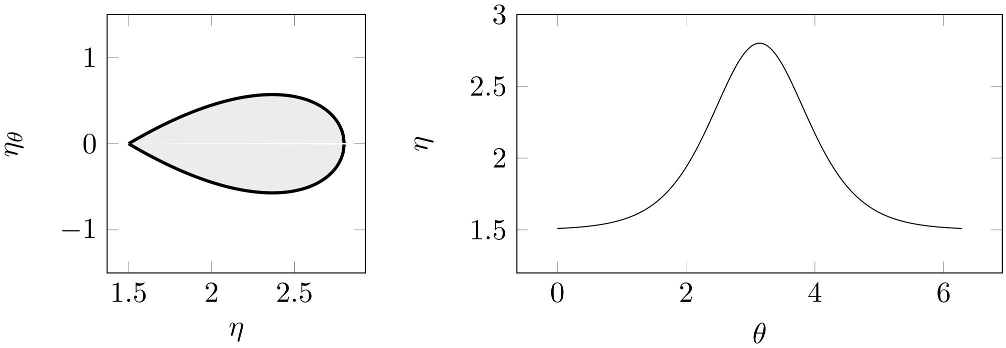


FIGURE 2.3: Parametric plot of η and η_θ for the solitonic case $e_3 \equiv e_2$ (left) and the corresponding traveling wave solution $\eta(\theta)$ (right). This corresponds to the central picture of Figure 2.1. The shaded area is proportional to the variable W defined in Equation (2.43).

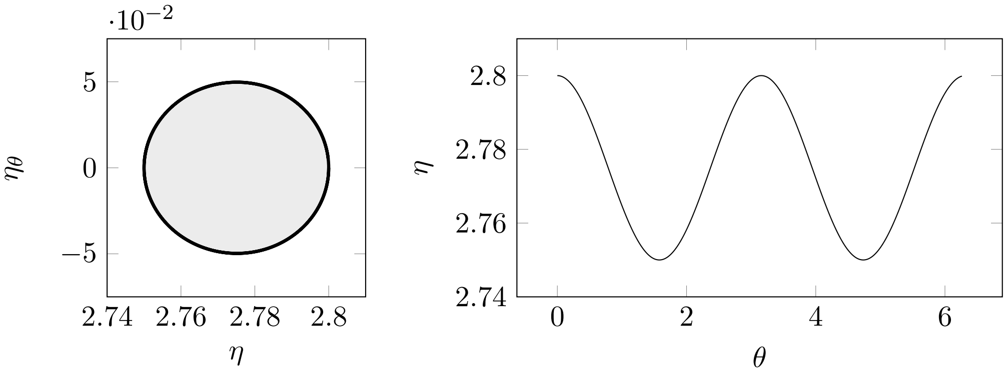


FIGURE 2.4: Parametric plot of η and η_θ for the small amplitude case $e_3 \neq e_2 \to e_1$ (left) and the corresponding periodic wave solution $\eta(\theta)$ (right). In this limit, η and η_θ parametrize a circle. The plot is made by taking $e_1 = 2.8$, $e_2 = 2.75$, $e_3 = 0.25$. This case corresponds to the picture on the right of Figure 2.1. The shaded area is proportional to the variable W defined in Equation (2.43).

The details for the integration in Equation (2.26) can be found e.g. in [11, p. 597].

An implicit expression for the the traveling wave η is found after integrating Equation (2.25) over a generic interval $e_2 \leq \eta \leq e_1$ as follows:

$$\sqrt{3}k \int_{e_2}^{\eta} \frac{\mathrm{d}\eta}{\sqrt{-(\eta - e_1)(\eta - e_2)(\eta - e_3)}} = 2\pi. \tag{2.28}$$

The integral on the left hand side can be expressed as:

$$\int_{e_2}^{\eta} \frac{\mathrm{d}\eta}{\sqrt{-(\eta - e_1)(\eta - e_2)(\eta - e_3)}} = \frac{2}{\sqrt{e_1 - e_2}} \left[\mathrm{F}\left(\arcsin\left(\sqrt{\frac{e_1 - e_2}{e_1 - \eta}} \right) |m \right) - \mathrm{K}(m) \right], \tag{2.29}$$

where F is the incomplete elliptic integral of the first kind, defined as:

$$\mathrm{F}(\phi|m) = \int_0^{\phi} \frac{\mathrm{d}\theta}{\sqrt{1 - m\sin^2\theta}}. \tag{2.30}$$

It may be useful to recall the following relation between complete and incomplete elliptic integrals of the first kind:

$$\mathrm{K}(m) = \mathrm{F}\left(\frac{\pi}{2} |m \right). \tag{2.31}$$

An explicit expression for η can be derived from Equation (2.29) via inverse functions:

$$\eta(\theta) = e_2 + (e_1 - e_2)\,\mathrm{cn}^2\left(\frac{\theta}{\pi} - \mathrm{K}(m)|m \right) \tag{2.32}$$

$$\eta(x, y, t) = e_2 + (e_1 - e_2)\,\mathrm{cn}^2\left(\frac{\sqrt{e_1 - e_3}}{2\sqrt{3}\epsilon} \left(x + \frac{l}{k}y - \frac{\omega}{k}t \right) - \mathrm{K}(m)|m \right), \tag{2.33}$$

where cn denotes the Jacobi elliptic function:

$$\mathrm{cn}(u|m) = \cos\phi \qquad \text{where } u = \int_0^{\phi} \frac{\mathrm{d}\theta}{\sqrt{1 - m\sin^2\theta}}. \tag{2.34}$$

The theory of elliptic transcendents is classical, for details on the derivation see e.g. [91] or [100].

Equation (2.33) is the most general form of a traveling wave solution for the KP Equation. The functional form is the same for the KPI and the KPII Equations, but amplitude, speed and propagation direction of the wave differ for the two equations, since the roots e_1, e_2, e_3 as well as the wave parameters k, l, ω depend on V, which in turn depends on α.

Some special forms of the traveling wave are worth mentioning.

Constant solution If $e_1 \to e_2$, clearly $\eta = e_2$ and the solution is constant for all times.

Small amplitude wave If $m \to 0$, it is possible to apply the following expansion for the elliptic integrals:

$$\begin{aligned} \mathrm{K}(m) &= \frac{\pi}{2}\left(1 + \frac{m}{4} + \frac{9}{64}m^2 \right) + O(m^3) \\ \mathrm{cn}(x|m) &= \cos x + \left(\frac{x}{4}\sin x - \frac{1}{8}\sin x \sin 2x \right) m + O(m^2), \end{aligned} \tag{2.35}$$

which lead to the periodic solution:

$$\eta(\theta) = e_2 + (e_1 - e_2)\sin^2\left(\frac{\sqrt{e_1 - e_3}}{2\sqrt{3}\epsilon} \left(x + \frac{l}{k}y - \frac{\omega}{k}t \right) \right). \tag{2.36}$$

A representation of this solution can be seen in Figure 2.4.

Soliton If $e_3 \to e_2$, such that $m \to 1$, the following asymptotic holds:

$$\lim_{m\to 1} \operatorname{cn}(x|m) = \operatorname{sech} x, \tag{2.37}$$

so that the solution is the line soliton:

$$\eta(x,y,t) = e_2 + (e_1 - e_2)\operatorname{sech}^2\left(\frac{\sqrt{e_1 - e_2}}{2\sqrt{3}\epsilon}\left(x + \frac{l}{k}y - \frac{\omega}{k}t\right)\right). \tag{2.38}$$

A plot of the soliton solution is available in Figure 2.3.

2.2.1 Lagrangian averaging for the KP traveling waves

Following Whitham's seminal work for the KdV Equation [116], it is possible to derive a system of PDEs which rules the modulation of the traveling waves solutions. A remarkable property of the Lagrangian averaging method is that the explicit knowledge of a traveling wave solution is not necessary. This fact was achieved by Whitham through a clever choice of the modulation variables.

The original form of the traveling wave potential is:

$$\varphi(x,y,t) = \beta_1 x + \beta_2 y - \gamma t + \phi(\theta) \qquad\qquad \theta = \frac{1}{\epsilon}\left(kx + ly - \omega t\right), \tag{2.39}$$

and the Lagrangian density for the KP Equation, expressed in terms of the potential is:

$$L = \epsilon^2 \varphi_t \varphi_x + \frac{\epsilon^3}{3}\varphi_x^3 - \epsilon^4 \varphi_{xx}^2 + \epsilon^2 \alpha \varphi_y^2. \tag{2.40}$$

By applying the relations (2.14), (2.15), (2.16), the Lagrangian can be rewritten in terms of η and η_θ as:

$$\begin{aligned} L &= -\left(\gamma + \frac{\omega}{k}(\eta - \beta_1)\right)\eta + \frac{1}{3}\eta^3 - k^2\eta_\theta^2 + \alpha\left(\beta_2 + \frac{l}{k}(\eta - \beta_1)\right)^2 \\ &= -k^2\eta_\theta^2 + \frac{1}{3}\eta^3 + \left(\alpha\frac{l^2}{k^2} - \frac{\omega}{k}\right)\eta^2 + \left(2\alpha\left(\beta_2 - \frac{l}{k}\beta_1\right)\frac{l}{k} + \beta_1\frac{\omega}{k} - \gamma\right)\eta + \alpha\left(\beta_2 - \frac{l}{k}\beta_1\right)^2. \end{aligned} \tag{2.41}$$

With the help of Equation (2.20), it is possible to simplify the Lagrangian (2.41) to:

$$L = -2k^2\eta_\theta^2 + \left(\frac{B}{3} - \gamma + \beta_1\frac{\omega}{k} + 2\alpha\frac{l}{k}\left(\beta_2 - \frac{l}{k}\beta_1\right)\right)\eta + \alpha\left(\beta_2 - \frac{l}{k}\beta_1\right)^2 + \frac{A}{3}. \tag{2.42}$$

Note that this simplification does not require an explicit knowledge of the traveling wave solution, although clearly some information is implicitly present in Equation (2.20).

It is convenient to introduce the following quantity:

$$\begin{aligned} W(A,B,V) &= \frac{1}{2\pi}\int_0^{2\pi} k\eta_\theta^2 \,\mathrm{d}\theta \\ &= \frac{\sqrt{3}}{2\pi}\oint \frac{\eta_\theta^2}{\sqrt{-\eta^3 + V\eta^2 + A\eta + B}}\,\mathrm{d}\eta \\ &= \frac{k}{\sqrt{3}\pi}\int_{e_2}^{e_1}\sqrt{-\eta^3 + V\eta^2 + A\eta + B}\,\mathrm{d}\eta. \end{aligned} \tag{2.43}$$

Note that in the last line of Equation (2.43), the dependence of W on A, B, V is hidden also in the integral limits e_2, e_1. An interpretation on the meaning of W can be inferred from the third equation of (2.43). Indeed, by interpreting η as a generalized coordinate q, and η_θ as a generalized momentum p, then W has the form:

$$W \simeq \oint p\,\mathrm{d}q, \tag{2.44}$$

which is the classical definition of an action in Hamiltonian Mechanics. For a visual interpretation of W, see the caption in Figures 2.2, 2.3 and 2.4. For computing the average Lagrangian, only the following integrals are needed:

$$\begin{aligned}\langle\eta\rangle &:= \frac{1}{2\pi}\int_0^{2\pi}\eta\,\mathrm{d}\theta = \beta_1\\ \langle\eta_\theta^2\rangle &:= \frac{1}{2\pi}\int_0^{2\pi}\eta_\theta^2\,\mathrm{d}\theta = \frac{W}{k},\end{aligned} \tag{2.45}$$

which lead to the following averaged Lagrangian:

$$\overline{L} = \frac{1}{2\pi}\int_0^{2\pi} L\,\mathrm{d}\theta = -2kW + \frac{A}{3} + \frac{1}{3}B\beta_1 - \gamma\beta_2 + \frac{1}{3}V\beta_1^2 + \alpha\beta_2^2. \tag{2.46}$$

Since the averaging is performed over the fast variable θ, the resulting Lagrangian $\overline{L}$ depends only on the slow variables x, y, t. The variational principle for the averaged Lagrangian is:

$$\delta\iint \overline{L}(\beta_1,\beta_2,\gamma,k,l,\omega,A,B)\,\mathrm{d}t\,\mathrm{d}x\,\mathrm{d}y, \tag{2.47}$$

where the dependence of the Lagrangian on V is missing since V is a function of k, l, ω, as in Equation (2.21). Note that the dependence of $\overline{L}$ on β_1, β_2, γ is due to the slow variable ψ, while the dependence of $\overline{L}$ on k, l, ω is due to the fast variable θ.

Taking variations with respect to the two parameters A and B gives the algebraic constraints:

$$\delta A:\ \overline{L}_A = 0 \qquad\qquad \delta B:\ \overline{L}_B = 0, \tag{2.48}$$

$$\delta A:\ kW_A = \frac{1}{6} \qquad\qquad \delta B:\ kW_B = \frac{\beta_1}{6}. \tag{2.49}$$

The two equations (2.49) can be combined to give the interesting expression for β_1:

$$\beta_1 = \frac{W_B}{W_A}. \tag{2.50}$$

It is interesting to note that such relations, together with the definition of V, give the dispersion relation:

$$\omega = \frac{1}{3}kV + \alpha\frac{l^2}{k}. \tag{2.51}$$

The dispersion relation is an important equation, since it predicts the speed of the wave as function of its amplitude[2] and propagation direction. The dependence of the wave speed on the amplitude is a typical nonlinear effect, while the dependence on the direction of propagation follows from the hypothesis of slow dependence of the solution on the y coordinate made by Kadomtsev and Petviashvili.

Variations with respect to V are not relevant to the dynamics since these are equivalent to variations with respect to θ. However, it is interesting to report the relation:

$$\delta V:\ \overline{L}_V = 0 \qquad\qquad \frac{\beta_1^2}{6} = kW_V. \tag{2.52}$$

Variations with respect to the angles ψ and θ give the equations of motion:

$$\delta\psi:\ \frac{\partial}{\partial t}\overline{L}_\gamma - \frac{\partial}{\partial x}\overline{L}_{\beta_1} - \frac{\partial}{\partial y}\overline{L}_{\beta_2} = 0 \tag{2.53}$$

$$\delta\theta:\ \frac{\partial}{\partial t}\overline{L}_\omega - \frac{\partial}{\partial x}\overline{L}_k - \frac{\partial}{\partial y}\overline{L}_l = 0. \tag{2.54}$$

---

[2] The dependence on the amplitude of the wave is due to V. Indeed, the amplitude is given by $a = e_1 - e_2$, and V depends on e_1 and e_2 via Equation (2.23).

The variational equations are complemented with the compatibility relations:

$$\theta_{xt} = \theta_{tx} \qquad \psi_{xt} = \psi_{tx} \tag{2.55}$$

$$\theta_{yt} = \theta_{ty} \qquad \psi_{yt} = \psi_{ty} \tag{2.56}$$

$$\theta_{xy} = \theta_{yx} \qquad \psi_{xy} = \psi_{yx}, \tag{2.57}$$

which are equivalent to the following wave equations:

$$k_t + \omega_x = 0 \qquad \beta_{1,t} + \gamma_x = 0 \tag{2.58}$$

$$l_t + \omega_y = 0 \qquad \beta_{2,t} + \gamma_y = 0 \tag{2.59}$$

$$l_x - k_y = 0 \qquad \beta_{2,x} - \beta_{1,y} = 0. \tag{2.60}$$

It is worth noting that Equation (2.60) is not completely independent from the first two. Indeed, by deriving Equation (2.58) by y and Equation (2.59) by x, it is found that:

$$\omega_{xy} = -k_{ty} \qquad \gamma_{xy} = -\beta_{1,ty} \tag{2.61}$$

$$\omega_{yx} = -l_{tx} \qquad \gamma_{yx} = -\beta_{2,tx}, \tag{2.62}$$

which are equivalent to:

$$k_{ty} = l_{tx} \qquad \beta_{1,ty} = \beta_{2,tx}. \tag{2.63}$$

Deriving Equation (2.60) with respect to t, it is clear that the resulting equations are the same as those of Equation (2.63).

Summarising, there are six evolutionary equations for the parameters (two from Equations (2.53) (2.54) and four out of the six Equations (2.58), (2.59), (2.60)), and two constraints (2.49), for the 8 modulation parameters $k, l, \omega, \beta_1, \beta_2, \gamma, A, B$. Consequently, the modulation system is determined.

The eight equations (2.49), (2.53), (2.54), (2.55), (2.56), (2.57) and (2.58), (2.59), (2.60), although correct, are not very much prone to analysis, hence it is preferrable to rearrange these in a form more amenable to formal manipulations. Writing down explicitly the variation with respect to ψ leads to:

$$\begin{aligned} (\beta_1)_t + \left(\frac{1}{3}B - \gamma + \frac{2}{3}V\beta_1\right)_x + (2\alpha\beta_2)_y = 0 \\ (2\beta_1)_t + \frac{1}{3}B_x + (4kVW_B)_x + 2\alpha\beta_{2,y} = 0 \\ (6kW_B)_t + \frac{1}{3}kW_AB_x + \left(\frac{1}{3}kVW_B\right)_x + \alpha\beta_{2,y} = 0, \end{aligned} \tag{2.64}$$

which, after some rearranging becomes:

$$W_{Bt} + \left(\frac{1}{3}kV\right)W_{Bx} + \frac{1}{6}B_xW_A + \left(\frac{k_t}{k} + 3\left(\frac{\omega}{k} - \alpha\frac{l^2}{k^2}\right)\frac{k_x}{k} + \left(\frac{\omega}{k} - \alpha\frac{l^2}{k^2}\right)_x\right)W_B + \alpha\beta_{2,y} = 0. \tag{2.65}$$

Introducing the new auxiliary variables:

$$p := \beta_2 - q\beta_1 \qquad q := \frac{l}{k} \tag{2.66}$$

and the "convective" derivative:

$$D := \frac{\partial}{\partial y} - q\frac{\partial}{\partial x} \tag{2.67}$$

simplifies Equation (2.65) to:

$$W_{Bt} + \left(\frac{V}{3} + \alpha q^2\right)W_{Bx} + \frac{1}{6}W_AB_x + \alpha\left(W_VDq + 2qDW_B + W_ADp\right) = 0. \tag{2.68}$$

It is convenient to express the dispersion relation in the new variables as:

$$\omega = \frac{1}{6}\left(\frac{1}{3}\frac{V}{W_A} + \alpha\frac{q^2}{W_A}\right). \tag{2.69}$$

The variation with respect to θ gives:

$$\left(6W_V + \frac{\beta_1^2}{k}\right)_t + \left(\frac{\omega - 2\alpha l q}{k^2}\left(6kW_V - \beta_1^2\right) - 2W\right)_x + 2\alpha\left(q\left(2W_V - \frac{\beta_1^2}{k}\right)\right)_y = 0, \tag{2.70}$$

which, after expanding the derivatives, and with the use of the consistency relations (2.58)-(2.60), yields the final form:

$$W_{Vt} + \left(\frac{V}{3} + \alpha q^2\right)W_{Vx} - \frac{1}{3}W_A A_x + 2\alpha\left(W_V Dq + qDW_V + W_B Dp\right) = 0. \tag{2.71}$$

The two Equations (2.65) and (2.71) are analogous to the two variational equations with respect to the angles ψ and θ respectively.

The next step consists in rearranging the variational equations with respect to the integration constants, Equations (2.49). From the first one, $kW_A = \frac{1}{6}$, it follows:

$$(kW_A)_t = 0 \qquad\qquad k_t W_A + kW_{At} = 0. \tag{2.72}$$

Recalling that

$$k_t = -\omega_x = \left(k\frac{V}{3} + \alpha\frac{l^2}{k}\right)_x, \tag{2.73}$$

it is found that:

$$\begin{aligned} &W_{At} + \frac{\omega_x}{k}W_A = 0 \\ &W_{At} + \frac{1}{k}\left(k\frac{V}{3} + \alpha\frac{l^2}{k}\right)_x = 0. \end{aligned} \tag{2.74}$$

After some algebraic manipulations, the following form appears:

$$W_{At} + \left(\frac{V}{3} + \alpha q^2\right)W_{Ax} - \frac{1}{3}W_A V_x + 2\alpha q D W_A = 0, \tag{2.75}$$

which is particularly interesting since for $\alpha = 0$ coincides with the Whitham's Equations for the KdV case.

Another equation can be obtained by time derivation of the auxiliary variable q:

$$\begin{aligned} q_t &= \frac{l_t}{k} - \frac{l}{k}\frac{k_t}{k} \\ &= -\frac{\omega_y}{k} + q\frac{\omega_x}{k} \\ &= -\frac{1}{k}D\omega \\ &= -D\left(\frac{V}{3} + \alpha q^2\right) - \left(\frac{V}{3} + \alpha q^2\right)\frac{k_y - qk_x}{k} \\ &= -D\left(\frac{V}{3} + \alpha q^2\right) - \left(\frac{V}{3} + \alpha q^2\right)q_x, \end{aligned} \tag{2.76}$$

which is equivalent to:

$$q_t + \left(\frac{V}{3} + \alpha q^2\right)q_x + \frac{1}{3}DV + 2\alpha q Dq = 0. \tag{2.77}$$

This last equation is peculiar of the two-dimensional case, since in the one-dimensional KdV Equation there is no room for modulating the wavefront direction.

A similar computation for p provides the last equation:

$$\begin{aligned} p_t &= (\beta_2 - q\beta_1)_t \\ &= q\gamma_x - \beta_1 q_t, \end{aligned} \tag{2.78}$$

that can be written as:

$$\begin{aligned} p_t &= -\beta_1 q_t - D\gamma = 0 \\ &= -\beta_1 q_t - D\left(\gamma - \frac{\omega}{k}\beta_1 + \left(\frac{V}{3} + \alpha q^2\right)\beta_1\right) \\ &= -\beta_1 q_t - D\left(\gamma - \frac{\omega}{k}\beta_1\right) - D\left(\frac{V}{3} + \alpha q^2\right)\beta_1 - \left(\frac{V}{3} + \alpha q^2\right) D\beta_1. \end{aligned} \tag{2.79}$$

Making use of the fact that $\beta_{2x} = \beta_{1y}$ leads to:

$$\begin{aligned} p_t &= -\beta_1 q_t - D\left(\gamma - \frac{\omega}{k}\beta_1\right) - D\left(\frac{V}{3} + \alpha q^2\right)\beta_1 - \left(\frac{V}{3} + \alpha q^2\right)(\beta_{2,x} - q\beta_{1,x}) \\ &= -\beta_1 q_t - D\left(\gamma - \frac{\omega}{k}\beta_1\right) - D\left(\frac{V}{3} + \alpha q^2\right)\beta_1 - \left(\frac{V}{3} + \alpha q^2\right)(p_x + \beta_1 q_x) \\ &= -D\left(\gamma - \frac{\omega}{k}\beta_1\right) - \left(\frac{V}{3} + \alpha q^2\right)p_x - \beta_1\left(q_t + \left(\frac{V}{3} + \alpha q^2\right)q_x + D\left(\frac{V}{3} + \alpha q^2\right)\right), \end{aligned} \tag{2.80}$$

which, by Equation (2.77), finally yields:

$$p_t + \left(\frac{V}{3} + \alpha q^2\right)p_x + D\left(\gamma - \frac{\omega}{k}\beta_1\right) = 0. \tag{2.81}$$

The final result of this analysis is the following modulation system for the KP Equation:

$$\begin{aligned} &W_{At} + \left(\frac{V}{3} + \alpha q^2\right)W_{Ax} - \frac{1}{3}W_A V_x + 2\alpha q D W_A = 0 \\ &W_{Bt} + \left(\frac{V}{3} + \alpha q^2\right)W_{Bx} + \frac{1}{6}W_A B_x + \alpha\left(W_B Dq + 2qDW_B + W_A Dp\right) = 0 \\ &W_{Vt} + \left(\frac{V}{3} + \alpha q^2\right)W_{Vx} - \frac{1}{3}W_A A_x + 2\alpha\left(W_V Dq + qDW_V + W_B Dp\right) = 0 \\ &q_t + \left(\frac{V}{3} + \alpha q^2\right)q_x + \frac{1}{3}DV + 2\alpha q Dq = 0 \\ &p_t + \left(\frac{V}{3} + \alpha q^2\right)p_x + D\left(\gamma - \frac{\omega}{k}\beta_1\right) = 0. \end{aligned} \tag{2.82}$$

The original, physical modulation variables are: $\beta_1, \beta_2, \gamma, k, l, \omega, A, B$. In the modulation system (2.82), instead, the modulation variables are W_A, W_B, W_V, p, q. These variables have a more symmetric form than the original modulation parameters. However, the equations (2.82) are extremely impractical for actual computations, and in the study of the physical evolution of a traveling wave it is convenient to start with a specific ansatz instead of the general form (2.33). Two examples of this kind of study will be discussed in Sections 2.3 and 2.4. Despite of this, the modulation Equations in the form of (2.82) are a valuable tool for analysis. As an example, from the ground breaking work of [49], it can be inferred that if the system (2.82) could be diagonalised, this would be equivalent to proving integrability of the KP problem.

The close relationship between the KdV and KP Equation can be seen also from the fact that, for $\alpha = 0, l = 0, q = 0, \beta_2 = 0, \partial_y = 0$, the modulation Equations (2.82) for the KP

Equation become:

$$\begin{aligned} W_{At} + \frac{V}{3}W_{Ax} - \frac{1}{3}W_A V_x = 0 \\ W_{Bt} + \frac{V}{3}W_{Bx} + \frac{1}{6}W_A B_x = 0 \\ W_{Vt} + \frac{V}{3}W_{Vx} - \frac{1}{3}W_A A_x = 0, \end{aligned} \tag{2.83}$$

which are exactly the modulation system for the KdV Equation as found originally by Whitham [116] (apart from a correction in the coefficients due to a different choice of scaling in the original PDE).

2.3 Modulation Equations for the KPI line soliton

This section is devoted to the derivation of the Whitham modulation equations for a class of exact solutions known as line solitons. The modulations for the line solitons could in principle be obtained by algebraic manipulation of the general framework of Equation (2.82). However, it is convenient to apply the Lagrangian averaging machinery directly to the line soliton traveling wave. A similar idea will be pursued in Section (2.4) for the lump solution of the KPI Equation.

The line soliton for the KPI Equation is the soliton solution for the associated KdV Equation:

$$u_t + uu_x + \epsilon^2 u_{xxx} = 0, \tag{2.84}$$

whose expression is:

$$u(x,y,t) = a\,\mathrm{sech}^2\left(\sqrt{\frac{a}{12}}(x - \frac{a}{3}t)\right) \tag{2.85}$$

for any amplitude $a > 0$. As such, the line soliton does not depend on the y coordinate.

A possible modulation for the line soliton (2.85) in the spirit of the discussion of Section 2.1 is:

$$u(\theta,x,y,t) = a\,\mathrm{sech}^2\theta \qquad \theta = \sqrt{\frac{a}{12}}\left(\frac{x}{\epsilon} + \frac{l}{k}\frac{y}{\epsilon} - \frac{\omega}{k}\frac{t}{\epsilon}\right). \tag{2.86}$$

If the parameters are allowed to change slowly with x, y, t, then the structure of the modulated line soliton becomes clear:

$$u(\theta,x,y,t) = a(x,y,t)\,\mathrm{sech}^2\left(\sqrt{\frac{a(x,y,t)}{12}}\left(\frac{x}{\epsilon} + \frac{l(x,y,t)}{k(x,y,t)}\frac{y}{\epsilon} - \frac{\omega(x,y,t)}{k(x,y,t)}\frac{t}{\epsilon}\right)\right). \tag{2.87}$$

The expression (2.87) allows for slow modulations of the soliton's wavefront shape, heigth, and speed of propagation.

The traveling wave (2.86) is associated to the following potential function:

$$\varphi = \sqrt{12a}\tanh\left(\sqrt{\frac{a}{12}}\left(\frac{x}{\epsilon} + \frac{l}{k}\frac{y}{\epsilon} - \frac{\omega}{k}\frac{t}{\epsilon}\right)\right), \tag{2.88}$$

in the sense that

$$u = \epsilon\varphi_x. \tag{2.89}$$

For clarity, we recall here the expression for the Lagrangian density:

$$L = \epsilon^2\varphi_t\varphi_x + \frac{\epsilon^3}{3}\varphi_x^3 - \epsilon^4\varphi_{xx}^2 + \alpha\epsilon^2\varphi_y^2. \tag{2.90}$$

Since φ is known, it is possible to compute the following averages:

$$\langle\varphi_t\varphi_x\rangle := \int_{-\infty}^{\infty}\varphi_t\varphi_x\,\mathrm{d}x = -\frac{8}{\sqrt{3}}\frac{\omega}{k}\frac{a^{3/2}}{\epsilon} \qquad \langle\varphi_x^3\rangle = \frac{32}{5\sqrt{3}}\frac{a^{5/2}}{\epsilon^2} \tag{2.91}$$

$$\langle\varphi_{xx}^2\rangle = \frac{8}{15\sqrt{3}}\frac{a^{5/2}}{\epsilon^3} \qquad \langle\varphi_y^2\rangle = \frac{8}{\sqrt{3}}\frac{l^2}{k^2}\frac{a^{3/2}}{\epsilon}. \tag{2.92}$$

Note that the integrals in Equations (2.92) and (2.93) are performed only on x, and not on $\mathbb{R}^2$ (and indeed, if the integrals were performed on the whole plane, these would diverge as the line soliton (2.86) is a nondecreasing function in the y direction). The integrals (2.92) are in turn useful in the computation of the averaged lagrangian:

$$\overline{L} = k\int_{-\infty}^{\infty} L\,\mathrm{d}x = -\omega a^{3/2} + \frac{1}{5}ka^{5/2} + \alpha\frac{l^2}{k}a^{3/2}, \tag{2.93}$$

where the equality in Equation (2.93) is understood up to an irrelevant multiplicative constant.

The variation with respect to the amplitude a gives the dispersion relation:

$$\frac{\partial\overline{L}}{\partial a} = 0 \qquad \omega = \frac{1}{3}ka + \alpha\frac{l^2}{k}. \tag{2.94}$$

The variation with respect to the angle θ leads to the following conservation law:

$$(a^{3/2})_t + \left(\frac{1}{5}a^{5/2} - \alpha q^2 a^{3/2}\right)_x + \left(2\alpha q a^{3/2}\right)_y = 0, \tag{2.95}$$

that can be written in advective form as:

$$a_t + \left(\frac{1}{3}a - \alpha q^2\right)a_x - \frac{4}{3}\alpha a q q_x + 2\alpha q a_y + \frac{4}{3}\alpha a q_y = 0, \tag{2.96}$$

with $q = l/k$. An evolution equation for q can be derived using the compatibility conditions (2.58), (2.59), (2.60) and the dispersion relation (2.94):

$$\begin{aligned} q_t &= \frac{l_t}{k} - \frac{l}{k}\frac{k_t}{k} \\ q_t - \frac{l}{k^2}\omega_x + \frac{1}{k}\omega_y &= 0, \end{aligned} \tag{2.97}$$

that after some algebraic computations simplifies to:

$$q_t - \frac{1}{3}qa_x + \left(\frac{1}{3}a - \alpha q^2\right)q_x + \frac{1}{3}a_y + 2\alpha q q_y = 0. \tag{2.98}$$

The two modulation equations for a and q can be written jointly as:

$$\begin{pmatrix} a \\ q \end{pmatrix}_t + \begin{bmatrix} \frac{1}{3}a - \alpha q^2 & -\frac{4}{3}\alpha a q \\ -\frac{1}{3}q & \frac{1}{3}a - \alpha q^2 \end{bmatrix}\begin{pmatrix} a \\ q \end{pmatrix}_x + \begin{bmatrix} 2\alpha q & \frac{4}{3}\alpha a \\ \frac{1}{3} & 2\alpha q \end{bmatrix}\begin{pmatrix} a \\ q \end{pmatrix}_y = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{2.99}$$

It will be convenient to denote with A_1 and A_2 the two matrices of equation (2.99), so that if $\boldsymbol{v} = (a, q)$, then Equation (2.99) can be expressed more compactly as:

$$\boldsymbol{v}_t + A_1\boldsymbol{v}_x + A_2\boldsymbol{v}_y = \boldsymbol{0}. \tag{2.100}$$

After the system (2.99) is solved for a and q, the remaining modulation parameters ω, k, l can be obtained, e.g. by the following procedure. Starting with the consistency condition:

$$k_t + \omega_x = 0, \tag{2.101}$$

and making use of the dispersion relation (2.94) gives the evolutionary equation for k:

$$k_t + \left(\frac{1}{3}a - \alpha q^2\right) k_x + 2\alpha q k_y + \frac{1}{3}\alpha a_x k = 0, \tag{2.102}$$

with a and q known. Solving Equation (2.102) allows in turn to get l and ω from the simple expressions:

$$l = qk \tag{2.103}$$

and (2.94), that we rewrite for convenience:

$$\omega = \frac{1}{3}ka + \alpha\frac{l^2}{k}. \tag{2.104}$$

Alternatively, it would be possible to start with the consistency relation dual to the (2.101):

$$l_t + \omega_y = 0 \tag{2.105}$$

and to repeat the same computations as above to obtain an evolutionary equation for l:

$$l_t + \left(\frac{1}{3}a - \alpha q^2\right) l_x + 2\alpha q l_y + \frac{1}{3}\frac{a_y}{q} l = 0. \tag{2.106}$$

Writing the two uncoupled equations (2.102) and (2.106) as a system:

$$\begin{pmatrix} k \\ l \end{pmatrix}_t + \begin{bmatrix} \frac{1}{3}a - \alpha q^2 & 0 \\ 0 & \frac{1}{3}a - \alpha q^2 \end{bmatrix} \begin{pmatrix} k \\ l \end{pmatrix}_x + \begin{bmatrix} 2\alpha q & 0 \\ 0 & 2\alpha q \end{bmatrix} \begin{pmatrix} k \\ l \end{pmatrix}_y + \begin{pmatrix} \frac{1}{3}a_x k \\ \frac{1}{3}\frac{a_y}{q} l \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{2.107}$$

shows that k and l are advected with the same positive velocity in the x direction and with negative velocity in the y direction.

We close this section by showing how the modulation equations (2.99) can be derived directly from the general Whitham system (2.82). For the line soliton solution, we have that the two roots of smaller magnitude, e_2 and e_3 are both equal to zero, and the third root e_1 is generally a nonzero positive number:

$$e_1 > 0, \qquad e_2 = e_3 = 0. \tag{2.108}$$

In this case, from Equation (2.23), we have for the potential and the integration constants:

$$A = B = 0, \qquad V = e_1. \tag{2.109}$$

Furthermore, from Equation (2.38), the amplitude of the line soliton is given by:

$$a = e_1 - e_2 = e_1 = V. \tag{2.110}$$

With these simplifications, the action W can be computed explicitly as:

$$W = \frac{k}{\sqrt{3}\pi}\int_0^{e_1} \eta\sqrt{(e_1 - \eta)}\,\mathrm{d}\eta = \frac{4}{15}\frac{k}{\sqrt{3}\pi}e_1^{5/2}. \tag{2.111}$$

Note in particular that in Equation (2.111), e_1 can be freely interchanged with V or with a, in virtue of Equation (2.110).

Since W does not depend explicitly on A or B, the first two equations of the (2.82) contain no information. Similarly, the fifth equation of the (2.82) is an identity. From the third equation of (2.82), we have:

$$\frac{k}{\sqrt{3}\pi}V^{1/2}\left(V_t + \left(\frac{V}{3} + \alpha q^2\right)V_x + \frac{4}{3}\alpha V(q_y - qq_x) + 2\alpha q\,(V_y - qV_x)\right) = 0, \tag{2.112}$$

which, after replacing $V = a$ and simplifying gives:

$$a_t + \left(\frac{a}{3} - \alpha q^2\right) a_x + 2\alpha q a_y + \frac{4}{3}\alpha a\,(q_y - q q_x) = 0, \tag{2.113}$$

which concides with the modulation equation for the amplitude found by a different route in Equation (2.99). Similarly, the fourth equation of (2.82), after replacing $V = a$ and repeating the computation gives:

$$q_t + \left(\frac{a}{3} + \alpha q^2\right) q_x + \frac{1}{3}(a_y - q a_x) + 2\alpha q\,(q_y - q q_x) = 0, \tag{2.114}$$

which concides with the modulation equation for q in Equation (2.99).

2.4 Modulation Equations for the KPI lump solution

This section is devoted to the derivation of Whitham's modulation equations for the lump solution of the KPI Equation. Lumps are a family of rational solution (see the discussion in Section 1.3.7), which consist in a very localized peak. The simplest lump solution is defined by:

$$u(x,y,t) = 24\frac{-(x-3b^2t)^2 + 3b^2y^2 + \frac{1}{b^2}}{\left[(x-3b^2t)^2 + 3b^2y^2 + \frac{1}{b^2}\right]^2}, \tag{2.115}$$

which describes a single peak moving in the x direction with velocity proportional to its amplitude (and both are proportional to the parameter b^2). The tau function for the lump described above is given by:

$$\tau(x,y,t) = \left((x-3b^2t)^2 + 3b^2y^2 + \frac{1}{b^2}\right), \tag{2.116}$$

so that indeed

$$u(x,y,t) = 12\partial_{xx}\log\tau(x,y,t). \tag{2.117}$$

It is convenient to introduce the modulation parameters directly in the τ function. This amounts to the assumption that the lump's speed, amplitude, and propagation direction are allowed to change weakly in space and time, independently of each other. Reintroducing also the dependence on ϵ, the modulated lump is expressed by:

$$\tau(x,y,t) = \left(\frac{x}{\epsilon} - 3\frac{\omega}{k}\frac{t}{\epsilon}\right)^2 + 3\left(\frac{l}{k}\frac{y}{\epsilon}\right)^2 + \frac{1}{a}, \tag{2.118}$$

where the modulation parameters are related to the original parameter b by:

$$\frac{\omega}{k} = b^2 \qquad\qquad \frac{l}{k} = b \qquad\qquad a = b^2. \tag{2.119}$$

Note in particular that the modulated amplitude a is a positive function of x, y and t.

The velocity potential in this case is given by:

$$\varphi(x,y,t) = 12\partial_x\log\tau(x,y,t), \tag{2.120}$$

which gives immediately:

$$\varphi_t = -3\frac{\omega}{k}\varphi_x. \tag{2.121}$$

As a result, the Lagrangian density for the lump solution becomes:

$$L = -3\epsilon^2\frac{\omega}{k}\varphi_x^2 + \frac{\epsilon^3}{3}\varphi_x^3 - \epsilon^4\varphi_{xx}^2 + \alpha\epsilon^2\varphi_y^2. \tag{2.122}$$

In this case, the terms of the averaged Lagrangian are:

$$\langle\varphi_x^2\rangle := \int_{\mathbb{R}^2} \varphi_x^2 \,\mathrm{d}x\,\mathrm{d}y = 96\sqrt{3}\pi\frac{k}{l}a \qquad \langle\varphi_x^3\rangle = 1152\sqrt{3}\pi\frac{k}{l}\frac{a^2}{\epsilon} \tag{2.123}$$

$$\langle\varphi_{xx}^2\rangle = 192\sqrt{3}\pi\frac{k}{l}\frac{a^2}{\epsilon^2} \qquad \langle\varphi_y^2\rangle = 96\sqrt{3}\pi\frac{l}{k}a. \tag{2.124}$$

The averaged Lagrangian density for the modulated lump solution is then:

$$\overline{L} = l\int_{\mathbb{R}^2} L \,\mathrm{d}x\,\mathrm{d}y = -3\omega a + 2ka^2 + \alpha\frac{l^2}{k}a. \tag{2.125}$$

Note that the lagrangian averaging in this case differs from the averaging introduced in Equation (2.93) for the line soliton. Indeed, since the lump is a highly localized solution, it belongs to the class of square integrable functions on $\mathbb{R}^2$, $L^2(\mathbb{R}^2)$.

Requiring that the action is stationary with respect to amplitude variations leads to the dispersion relation:

$$\delta a: \qquad \omega = \frac{4}{3}ka + \frac{1}{3}\alpha\frac{l^2}{k}. \tag{2.126}$$

The dispersion relation (2.126) for the lump solution is remarkably similar to the dispersion relation (2.94) for the line soliton. To check the consistency of the dispersion relation so obtained, it is sufficient to replace the definitions (2.119) in Equation (2.126), which is indeed an identity if $\alpha = -1$.

Stationarity of the action with respect to angular variations gives the conservation law:

$$\delta\theta: \qquad (3a)_t + \left(2a^2 - \alpha q^2 a\right)_x + (2\alpha q a)_y = 0. \tag{2.127}$$

Equation (2.127) can be expressed in the equivalent form:

$$a_t + \left(\frac{4}{3}a - \frac{1}{3}\alpha q^2\right)a_x - \frac{2}{3}\alpha a q q_x + \frac{2}{3}\alpha q a_y + \frac{2}{3}\alpha a q_y = 0. \tag{2.128}$$

A second modulation equation can be obtained by combining the definition of q with the integrability conditions:

$$k_t + \omega_x = 0 \qquad l_t + \omega_y = 0 \qquad k_y - l_x = 0, \tag{2.129}$$

as follows. Starting with the identity:

$$q_t = \frac{l_t}{k} - \frac{l}{k}\frac{k_t}{k} = -\frac{1}{k}\omega_x + \frac{l}{k^2}\omega_y \tag{2.130}$$

and replacing in Equation (2.130) the definition of ω coming from the dispersion relation (2.126),

$$q_t - \frac{l}{k^2}\left(\frac{4}{3}ka + \frac{1}{3}\alpha\frac{l^2}{k}\right)_x + \frac{1}{k}\left(\frac{4}{3}ka + \frac{1}{3}\alpha\frac{l^2}{k}\right)_y, \tag{2.131}$$

which, after some algebraic manipulations, yields the modulation equation for q:

$$q_t - \frac{4}{3}q a_x + \left(\frac{4}{3}a - \frac{1}{3}\alpha q^2\right)q_x + \frac{4}{3}a_y + \frac{2}{3}\alpha q q_y = 0. \tag{2.132}$$

Summarising, the modulation equations for the lump, written in advective form are:

$$\begin{pmatrix} a \\ q \end{pmatrix}_t + \begin{bmatrix} \frac{4}{3}a - \frac{1}{3}\alpha q^2 & -\frac{2}{3}\alpha a q \\ -\frac{4}{3}q & \frac{4}{3}a - \frac{1}{3}\alpha q^2 \end{bmatrix} \begin{pmatrix} a \\ q \end{pmatrix}_x + \begin{bmatrix} \frac{2}{3}\alpha q & \frac{2}{3}\alpha a \\ \frac{4}{3} & \frac{2}{3}\alpha q \end{bmatrix} \begin{pmatrix} a \\ q \end{pmatrix}_y = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{2.133}$$

If a and q are known, e.g. by solving the system (2.133), the remaining modulation parameters ω, k, l can be obtained by imposing the relations:

$$q = \frac{l}{k} \qquad \omega = \frac{4}{3}ak + \frac{1}{3}\alpha\frac{l^2}{k} \qquad k_x = l_y. \tag{2.134}$$

An more practical way for reconstructing the wave parameters ω, k, l is to start with the relation:

$$q_x = \frac{l_x}{k} - \frac{l}{k}\frac{k_x}{k}, \tag{2.135}$$

that after replacement of the rightmost equation in (2.134) becomes:

$$q_x = \frac{1}{k}\left(l_x - q l_y\right). \tag{2.136}$$

Finally, multiplying the (2.136) by l and rearranging gives a modulation equation for l:

$$l_x - q l_y = l\partial_x \log q. \tag{2.137}$$

After solving Equation (2.137), ω and k can be found with the algebraic relations:

$$k = \frac{l}{q} \qquad\qquad \omega = \frac{4}{3}a\frac{l}{q} + \frac{1}{3}\alpha q l. \tag{2.138}$$

A symmetric alternative to Equation (2.137) can be derived for k, by writing down explicitly the derivative of q_y and repeating the same passages as done in Equations (2.135) and (2.136), leading to the modulation equation for k:

$$k_x - q k_y = k q_y. \tag{2.139}$$

Alternatively, it is possible to proceed as in Equations (2.102)-(2.106) to obtain the two equations for k and l:

$$k_t + \left(\frac{4}{3}a - \frac{1}{3}\alpha q^2\right) k_x + \frac{2}{3}\alpha q k_y + \frac{4}{3}\alpha a_x k = 0 \tag{2.140}$$

$$l_t + \left(\frac{4}{3}a - \frac{1}{3}\alpha q^2\right) l_x + \frac{2}{3}\alpha q l_y + \frac{4}{3}\alpha\frac{a_y}{q} l = 0 \tag{2.141}$$

$$\tag{2.142}$$

which are the analogous for the lump of Equations (2.107).

2.5 Stability of line solitons and lumps

The stability of solutions to the KP Equation is an active research topic, which started with the seminal work by [120] on the stability of line solitons using IST. In this work, it is shown that in the KPI Equation line solitons are unstable since these tend to radiate away their mass and energy in a packet of small amplitude waves. Numerical evidence of the instability of line solitons in the KPI Equation, leading to wave breaking and the development of lumps was provided in [83]. A similar kind of instability for line solitons, leading to the development of lumps, was derived analytically in [68]. In [101] the asymptotic stability of lumps is shown, and it is proved that in some configurations, lumps can lose mass by low-amplitude radiation, stabilizing after a while as a lump with smaller amplitude.

It is interesting to study the stability of line soliton and lumps by means of Whitham Modulation Theory. The stability properties of line solitons and lumps can be studied by checking whether the modulation systems (2.99) and (2.133) respectively are hyperbolic or elliptic. To this end, consider the linear combination:

$$A_1 + \xi A_2, \tag{2.143}$$

where A_1 is the matrix that multiplies the x-derivatives in Equation (2.133), A_2 is the matrix that multiplies the y-derivatives in Equation (2.133), and ξ is a real number. System (2.133) is hyperbolic if the eigenvalues of $A_1 + \xi A_2$ are real for all $\xi \in \mathbb{R}$, otherwise if the eigenvalues are complex for some ξ, then the system is elliptic.

Both for the line soliton and the lump, the eigenvalues $\lambda_{1,2}$ of $A_1 + \xi A_2$ can be computed explicitly as:

$$\lambda_{1,2} = \frac{1}{3}a - \alpha q^2 + 2\alpha q\xi \pm \frac{2}{3}\sqrt{\alpha a(\xi - q)^2} \tag{2.144}$$

for the line soliton, and

$$\lambda_{1,2} = \frac{4}{3}a - \frac{1}{3}\alpha q^2 + \frac{2}{3}\alpha q\xi \pm \frac{2\sqrt{2}}{3}\sqrt{\alpha a(\xi - q)^2}, \tag{2.145}$$

for the lump. Since $a > 0$, for both cases the modulation equations are elliptic for the KPI Equation ($\alpha = -1$). It was shown in [41] that elliptic systems of conservation laws develop a singular point, and the singularity unfolds as an elliptic umbilic catastrophe. This behaviour is similar with the semiclassical limit of the focusing nonlinear Schrödinger Equation in one space dimension, and this analogy will be explored further in Chapter 5.

2.6 Discussion

The main result of this Chapter consists in the modulation equations for the line soliton and for the lump, respectively Equation (2.99) and (2.133). A similar, but not equivalent set of modulation equations was derived in [6] and in [10], using multiple scale analysis. The system (2.82) differs from the cited works because the equation for p is an evolution equation, while in [6, 10] is a constraint on the space derivatives of p. We regard that in principle it is not possible to assume that the average of p is constant in time in a periodic setting.

The modulation equations are not suitable for practical computations due to the difficulty of providing initial or boundary values if the modulation system is respectively hyperbolic or elliptic. Nevertheless, the modulation equations constitute a valuable tool for stability analysis.

According to the discussion of Section 2.5, it should be expected that the modulation of the line soliton is not stable for the KPI Equation. A similar result was shown numerically in [83], where it can clearly be seen that the evolution under the KPI flow of a line soliton with a very small perturbation leads very soon to a wave breaking with the formation of very high localized peaks (interpreted as lump solutions).

A second apparent instability result of Section 2.5 regards lump solutions. In this case, the results obtained from modulation theory are open to interpretation, since it is known that pure lump solutions are stable in the KPI Equation. A possible explanation for this observed behaviour is that lump modulation cannot happen for a single lump, but is inherently related to some secondary effect, such as dispersive radiation or interaction with other lumps (as is the case in the numerical computations, since the algebraic decay of lumps at infinity is likely responsible for some lump-lump interaction).

It would be interesting to study the integrability properties of the modulation system for the KP Equation. Since the original equation is integrable, it is expected that the modulation system too should be integrable. Given the hydrodynamic form of the modulation system:

$$\boldsymbol{z}_t + \mathrm{A}_1 \boldsymbol{z}_x + \mathrm{A}_2 \boldsymbol{z}_y = \boldsymbol{0}, \tag{2.146}$$

a convenient definition of integrability is that the system above can be written in diagonal form.

A possible way to check for integrability of a system of hydrodynamic type in the form (2.146) is due to Ferapontov and Khusnutdinova, that in the paper [48] have given a remarkably simple integrability criterion. This criterion consists in assembling the matrix:

$$(\rho \mathrm{I} + \mathrm{A}_1)^{-1} (\sigma \mathrm{I} + \mathrm{A}_2)\,; \tag{2.147}$$

then the integrability of system (2.146) is ensured if the matrix (2.147) can be diagonalised for any real ρ, σ.

In this way, integrability of the modulation systems for the line soliton and for the lump can be checked immediately. Indeed, considering the modulation system for the variables (a, q, k), the matrices A_1 and A_2 take the form:

$$\mathrm{A}_1 = \begin{bmatrix} \frac{1}{3}a - \frac{1}{3}\alpha q^2 & -\frac{4}{3}\alpha q & 0 \\ -\frac{1}{3}q & \frac{1}{3}a - \frac{1}{3}\alpha q^2 & 0 \\ \frac{1}{3}\alpha k & 0 & \frac{1}{3}a - \frac{1}{3}\alpha q^2 \end{bmatrix} \qquad \mathrm{A}_2 = \begin{bmatrix} 2\alpha q & \frac{4}{3}\alpha a & 0 \\ \frac{1}{3} & 2\alpha q & 0 \\ 0 & 0 & 2\alpha q \end{bmatrix} \tag{2.148}$$

for the line soliton, and:

$$\mathrm{A}_1 = \begin{bmatrix} \frac{4}{3}a - \frac{1}{3}\alpha q^2 & -\frac{2}{3}\alpha q & 0 \\ -\frac{4}{3}q & \frac{4}{3}a - \frac{1}{3}\alpha q^2 & 0 \\ \frac{4}{3}\alpha k & 0 & \frac{4}{3}a - \frac{1}{3}\alpha q^2 \end{bmatrix} \qquad \mathrm{A}_2 = \begin{bmatrix} \frac{2}{3}\alpha q & \frac{2}{3}\alpha a & 0 \\ \frac{4}{3} & \frac{2}{3}\alpha q & 0 \\ 0 & 0 & \frac{2}{3}\alpha q \end{bmatrix} \tag{2.149}$$

for the lump. The eigenvalues of the matrix (2.147) with the two special choices (2.148) and (2.149) for A_1 and A_2 can be computed easily with any computer algebra system, showing that generically the matrix (2.147) can be diagonalised.

For the full Whitham system (2.82), the procedure outlines here cannot be applied since the algebraic computations involved would require too much time for a 5×5 or a 6×6 system. To address this difficulty, in the work [48], the authors propose an alternative approach that consists in computing the “Haantjes tensor” associated to the matrix (2.147), and may be applied to larger systems such as the one of Equation (2.82). This possibility is kept open for a future work on the integrability and Hamiltonian properties of Whitham’s system for the KP Equation.

Chapter 3

Numerical Method

Despite the spectacular theory that allows to generate exact solutions for the KP Equation in terms of tau functions or by means of elliptic functions on Riemann surfaces, the direct problem is much harder. Specifically, the problem of finding the evolution of a given initial datum under the KP flow is in general not as simple as giving arbitrary exact solutions, mainly due to the heavy computational machinery required by the Inverse Scattering Transform.

As a result, numerical methods are still an important tool for the study of the KP Equation. There are however several difficulties that make the numerical study of such equations a particularly delicate task, the most formidable being the highly oscillatory behaviour of KP's traveling wave solutions for small dispersion. For any discretization method, a fast changing solution requires the use of high resolution numerics. This in turn implies an important effort on the algorithmic side, asking for the full exploitation of the possibilities brought forward by modern (vectorised, parallel) cluster architectures. Highly localised solutions, such as lumps, contribute as well to this computational burden.

For such reasons, it is worthwile to discuss the details of the numerical methods and of the practical implementation on modern clusters. This Chapter describes precisely our approach to address these difficulties with a suitable numerical method.

3.1 Fourier pseudospectral method

Almost all numerical works on the KP Equations, use a Fourier pseudospectral method. The main reasons for the popularity of Fourier pseudospectral methods are the difficulty of discretising the antiderivative operator, the high order linear dispersive operator, and the difficulty of imposing boundary conditions on non-periodic domains (but see [28] for a workaround in the imposition of initial data with non-periodic derivatives in Fourier methods). One of the first numerical studies of the KP Equations with Fourier methods was [21], and more recently a series of works by Klein [80, 81] have shown how to obtain accurate numerical solution for a variety of regimes (solitons, lumps, blowup [84, 83], dispersive shock waves [82], oscillatory regimes [85]). A hybrid method combining a Fourier pseudospectral method for the linear part with a discontinuous Galerkin scheme for the nonlinear part and operator splitting in time is proposed in [45]. A rare alternative to Fourier methods is presented in [29], where the authors develop a compact Finite Difference Method and validate it against a Fourier pseudospectral method. For a different flavour of numerical analysis, focused on the class of elliptic theta function solutions based on Chebyshev Spectral Methods was pursued in [52]. The numerical method chosen in this work is the *Fourier Pseudospectral Method* (also called *Fourier Collocation Method*).

There are some advantages of Fourier methods (or of Spectral Methods in general) that cannot be underestimated, namely:

- the knowledge of the expansion coefficients allows to extract relevant information on the quality of the numerical solution and to reckon some relevant features of the solution which otherwise would be out of reach;
- several differential or integral operators are hard to implement in a numerical code based on the space representation, but have straightforward or at least reachable implementation in a spectral code;

- zero numerical dissipation and dispersion.

Furthermore, the presence of a structured problem allows to achieve tremendous throughput, communication balance, and a very high sustained performance (in terms of the peak percentage) in modern cluster architectures. The details for an efficient implementation of a Fourier spectral method are described in Chapter 4. Clearly, there are also drawbacks in the choice of a Fourier pseudospectral method, some of which are:

- it is possible to consider only domains with periodic boundary conditions;
- the resolution is fixed throughout the domain, thus regions in which the solution changes slowly are likely to be overresolved, leading to suboptimal memory usage and generating very large output files.

The remaining part of this Section describes the Fourier pseudospectral method used for the numerical approximation of the KP Equation. Some important references on Fourier and other Spectral Methods are [27, 23, 56, 113].

3.1.1 Fourier Interpolation

In the following, all the equations are set in $\mathbb{T}^2$, the two-dimensional torus, as Fourier expansions allow only for periodic boundary conditions. It is understood that x, y form a cartesian coordinate system on $\mathbb{T}^2$, with x, y ranging in the interval $[0, 2\pi]$. The approximation method adopted in this work is based on the Fourier series representation of the solution to the KP Equation:

$$u(x,y,t) = \Re \sum_{\xi_1=-\infty}^{\infty} \sum_{\xi_2=-\infty}^{\infty} \widehat{u}_{\xi_1\xi_2}(t) \mathrm{e}^{\mathrm{i}\xi_1 x} \mathrm{e}^{\mathrm{i}\xi_2 y}, \tag{3.1}$$

with the symbol $\Re$ denoting the real part of a complex number. The coefficients $\widehat{u}_{\xi_1\xi_2}(t)$ are related to the original function u as:

$$\widehat{u}_{\xi_1\xi_2}(t) = \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^{2\pi} u(x,y,t) \mathrm{e}^{-\mathrm{i}\xi_1 x} \mathrm{e}^{-\mathrm{i}\xi_2 y} \, \mathrm{d}x \, \mathrm{d}y. \tag{3.2}$$

If the series in the representation (3.1) is truncated after N terms in each variable, the resulting function is called u^N:

$$u^N(x,y,t) = \Re \sum_{\xi_1=-N/2}^{N/2-1} \sum_{\xi_2=-N/2}^{N/2-1} \widehat{u}_{\xi_1\xi_2}(t) \mathrm{e}^{\mathrm{i}\xi_1 x} \mathrm{e}^{\mathrm{i}\xi_2 y}. \tag{3.3}$$

It is well known [27] that as $N \to \infty$, for a continuous function u of bounded variation, the expression (3.3) will converge uniformly to u. In the following, often the symbol $\Re$ will be understood. The series representation (3.1) can be a useful tool for constructing a convergence estimate: by applying the Parseval identity to the norm of the difference between u and u^N leads to the following estimate for the L^2 norm of the truncation error:

$$\|u - u^N\|_{L^2}^2 = 2\pi \sum_{|\xi_1| \geq N/2} \sum_{|\xi_2| \geq N/2} |\widehat{u}_{\xi_1\xi_2}|^2. \tag{3.4}$$

In a similar fashion, noting that that basis functions $\mathrm{e}^{\mathrm{i}\xi x}$ are unimodular, we obtain an estimate for the L^∞ norm of the error:

$$\|u - u^N\|_{L^\infty} \leq \sum_{|\xi_1| \geq N/2} \sum_{|\xi_2| \geq N/2} |\widehat{u}_{\xi_1\xi_2}|. \tag{3.5}$$

For a periodic, smooth function the error decays very rapidly with N. More precisely, the magnitude of the Fourier coefficients of the error approach zero faster [56] than any polynomial in N. This behaviour is often called *spectral convergence*, or *exponential convergence*.

In practical computations, the Fourier transform (3.2) is often replaced by its discrete counterpart. The discrete Fourier transform is based on an equispaced grid, whose N points x_i are equally spaced on the domain:

$$x_j = \frac{2\pi j}{N}, \qquad y_l = \frac{2\pi l}{N}, \qquad j, l = 0, \dots, N-1 \tag{3.6}$$

so that the computational points on $\mathbb{T}^2$ are:

$$x_{jl} := x_j y_l = \frac{(2\pi)^2 jl}{N^2} \qquad j, l = 0, \dots, N-1. \tag{3.7}$$

Then, the discrete Fourier transform is defined as:

$$\widehat{u}_{jl}(t) = \frac{1}{N} \sum_{j=0}^{N-1} \sum_{l=0}^{N-1} u(x_j, y_l, t) \mathrm{e}^{-\mathrm{i}jx_j} \mathrm{e}^{-\mathrm{i}ly_l}. \tag{3.8}$$

The inverse discrete Fourier transform allows to recover the value of u at the grid points (3.6) by means of the sum:

$$u(x_j, y_l, t) = \sum_{m=-N/2}^{N/2-1} \sum_{n=-N/2}^{N/2-1} \widehat{u}_{mn} \mathrm{e}^{\mathrm{i}jx_m} \mathrm{e}^{\mathrm{i}ly_n}. \tag{3.9}$$

If the pointwise values (3.9) are known, the value of the approximant u^N can be defined at any point in $\mathbb{T}^2$ by interpolation with trigonometric polynomials:

$$u^N(x, y, t) = \sum_{j,l=-N/2}^{N/2-1} \widehat{u}_{jl} \mathrm{e}^{\mathrm{i}jx} \mathrm{e}^{\mathrm{i}ly}, \tag{3.10}$$

or equivalently by interpolation with cardinal sine functions:

$$u^N(x, y, t) = \sum_{j,l=-N/2}^{N/2-1} \widehat{u}_{jl} \operatorname{sinc}(x - x_j) \operatorname{sinc}(y - y_l). \tag{3.11}$$

The Discrete Fourier Transform (3.8) can be interpreted as an approximation of the exact Fourier Transform (3.2). An important consequence of this approximation is that the coefficients of the trigonometric interpolant (3.10) are in general different from the coefficients of the Fourier projection (3.3). More precisely, denoting by $\mathcal{I}\widehat{u}_{jl}$ the Fourier coefficients of the interpolant (3.10) and by $\mathcal{P}\widehat{u}_{\xi_1\xi_2}$ the Fourier coefficients of the projection, it can be shown [27, Ch. 2] that:

$$\mathcal{I}\widehat{u}_{jl} = \mathcal{P}\widehat{u}_{jl} + \sum_{\substack{m,n=-\infty \\ m,n\neq 0}}^{\infty} \mathcal{P}\widehat{u}_{k+Nm,j+Nn} \qquad \forall\, j, l = 0, \dots, N-1. \tag{3.12}$$

The summation on the right hand side is called the *aliasing error*. The presence of the aliasing error implies that the error of the interpolant (3.10) is greater than the error for the projection (3.3). This is a typical effect of pseudospectral methods, as Galerkin spectral methods do not suffer from aliasing errors. It has been shown both in theory and in the computational practice that the aliasing error decays as fast as the truncation error if the number of modes is sufficiently large. Unfortunately, practical computations are often quite far from the large N asymptotic, so aliasing errors cannot be ignored. There exist several approaches to reduce aliasing in practical computations, such as the 3/2-rule, or the phase shift method [27, Ch. 2]. The approached followed here is different and guided more by practical considerations available in the literature than by theoretical analysis, and consists in the removal of aliasing by means of a Krasny filtering [87]. This filtering consists in setting to zero all the Fourier coefficients with magnitude smaller than a threshold, here taken equal to 10^{-10}, after every operator application. As an example, the application of $\mathbf{L}$ to the Fourier representation $\widehat{u}$ of some function u takes the form:

$$\mathbf{L}u := \sigma_{jl}\widehat{\mathbf{L}}\widehat{u}_{jl}, \tag{3.13}$$

with

$$\sigma_{jl} = \begin{cases} 1 & \text{if } |\widehat{u}_{jl}| > 10^{-10} \\ 0 & \text{otherwise.} \end{cases} \tag{3.14}$$

In [80] it is observed that the combination of Kransy filtering with sufficiently high resolution in space allows to obtain very accurate results in highly oscillatory regions even with Fourier pseudospectral expansions. An influential work on the use of Krasny-inspired filtering for the removal of aliasing in the Euler hydrodynamic equations is [66], and this kind of de-aliasing filters are popular in the hydrodynamics community.

An important consequence of the representation (3.10) is that linear differential operators act as multiplications on the Fourier components:

$$\begin{aligned} \partial_x u^N &= \partial_x \sum_{j,l=0}^{N-1} \widehat{u}_{jl} \mathrm{e}^{\mathrm{i}jx} \mathrm{e}^{\mathrm{i}ly} \\ &= \sum_{j,l=0}^{N-1} \mathrm{i}j \widehat{u}_{jl} \mathrm{e}^{\mathrm{i}jx} \mathrm{e}^{\mathrm{i}ly}. \end{aligned} \tag{3.15}$$

An analogous property holds for the antiderivative, when expressed in frequency space:

$$\begin{aligned} \partial_x^{-1} u^N &= \int_{\mathbb{R}} \left(\sum_{j,l=0}^{N-1} \widehat{u}_{jl} \mathrm{e}^{\mathrm{i}jx} \mathrm{e}^{\mathrm{i}ky} \right) \mathrm{d}x \\ &= \sum_{j,l=0}^{N-1} \frac{1}{\mathrm{i}j} \widehat{u}_{jl} \mathrm{e}^{\mathrm{i}jx} \mathrm{e}^{\mathrm{i}ky}. \end{aligned} \tag{3.16}$$

The last important property that is used in the analysis of the method (but not in the practical computations) is the following representation for space multiplications:

$$\begin{aligned} u^N v^N &= \left(\sum_{j,l=0}^{N-1} \widehat{u}_{jl} \mathrm{e}^{\mathrm{i}jx} \mathrm{e}^{\mathrm{i}ly} \right) \left(\sum_{m,n=0}^{N-1} \widehat{v}_{mn} \mathrm{e}^{\mathrm{i}mx} \mathrm{e}^{\mathrm{i}ny} \right) \\ &= \sum_{j,l=0}^{N-1} \sum_{p+r=j} \sum_{q+s=l} \widehat{u}_{pq} \widehat{v}_{rs} \mathrm{e}^{\mathrm{i}jx} \mathrm{e}^{\mathrm{i}ly}. \end{aligned} \tag{3.17}$$

Namely, products in physical space become convolutions in Fourier variables. Usually convolution is denoted by a star $\star$, so that the abbreviated notation for Equation (3.17) is:

$$u^N v^N = \widehat{u} \star \widehat{v} \qquad\qquad \widehat{uv} = u^N \star v^N. \tag{3.18}$$

Summing up, the important relations (3.15), (3.16), (3.17) can be given in terms of the Fourier coefficients as:

space	frequency	
$\partial_x u^N$	$\mathrm{i}j\widehat{u}_{jl}$	(3.19)
$\partial_y u^N$	$\mathrm{i}l\widehat{u}_{jl}$	(3.20)
$\partial_x^{-1} u^N$	$\frac{1}{\mathrm{i}j}\widehat{u}_{jl}$	(3.21)
$u^N v^N$	$\widehat{u} \star \widehat{v}.$	(3.22)

With this machinery in place, it is possible to introduce the approximation scheme. For generality, the original PDE is recast in the form:

$$u_t = \mathbf{L}u + \mathbf{N}(u), \tag{3.23}$$

where $\mathbf{L}$ denotes the linear part of the equation and $\mathbf{N}$ the nonlinear part. In the KP Equation, these operators are defined as:

$$\mathbf{L} = -\epsilon^2 \partial_{xxx} - \alpha \partial_x^{-1} \partial_{yy} \qquad\qquad \mathbf{N}(u) = u u_x = \frac{1}{2} \partial_x \left(u^2\right). \tag{3.24}$$

Note that in order to recast the KP Equation in the form (3.23), the antiderivative with respect to x had to be introduced in the definition of $\mathbf{L}$.

3.1.2 Fourier Projection and Collocation

Up to this point, only Fourier interpolation and projection have been discussed, meaning that the expansion coefficients for the Fourier series are supposed to be known, and several important properties following from this representation have been discussed. The purpose of a numerical approximation scheme is however that of finding an approximate value for the solution which is not known a priori. Fourier Spectral Methods find an approximation of the solution by prescribing a rule that allows to find an approximation to the Fourier expansion coefficients. The numerical solution can then be evaluated at any point following the interpolation techniques presented in the previous section.

Spectral Methods are generally classified in two main families: Galerkin methods and collocation methods. Galerkin methods require that a weak form of the original Equation (3.23) holds on some appropriate finite-dimensional functional space. This requires the introduction of an inner product $(\cdot,\cdot)$, which here we suppose to be the standard L^2 inner product on the two-dimensional torus. The Galerkin projection defining the Fourier Spectral Method then reads:

$$(\mathrm{e}^{\mathrm{i}mx}\mathrm{e}^{\mathrm{i}ny}, u_t^N) = (\mathrm{e}^{\mathrm{i}mx}\mathrm{e}^{\mathrm{i}ny}, \mathbf{L}u^N + \mathbf{N}(u^N)) \qquad \forall\, m, n = 0, \dots, N-1. \tag{3.25}$$

Equation (3.25) represents a system of N^2 equations for the N^2 expansion coefficients $\widehat{u}_{jl}^N$ of the numerical approximation u^N, that can be determined by solving this nonlinear system.

An alternative to Fourier Spectral Methods comes from Fourier pseudospectral methods, which require that the PDE (3.23) is fulfilled at the collocation points introduced in (3.7). Formally, this is equivalent to the Galerkin form (3.25) where the test functions $\mathrm{e}^{\mathrm{i}mx}\mathrm{e}^{\mathrm{i}ny}$ are replaced by Dirac deltas centered in the collocation points (3.7):

$$(\delta(x - x_m)\delta(y - y_n), u_t^N) = (\delta(x - x_m)\delta(y - y_n), \mathbf{L}u^N + \mathbf{N}(u^N)) \qquad \forall\, m, n = 0, \dots, N-1. \tag{3.26}$$

The unknown expansion coefficients $\widehat{u}_{jl}(t)$ are thus given by the system of ODEs:

$$\frac{\mathrm{d}}{\mathrm{d}t}\widehat{u}_{jl} = \widehat{\mathbf{L}}\widehat{u}_{jl} + \widehat{\mathbf{N}(u)} \qquad \text{for } j, l = 0, \dots, N-1. \tag{3.27}$$

Equation (3.27) is a system of N^2 ODEs and N^2 unknowns. The symbol $\widehat{\mathbf{L}}$ denotes the Fourier representation of the linear operator $\mathbf{L}$, and for the KP Equation is given by:

$$\widehat{\mathbf{L}} = \mathrm{i}\epsilon^2 j^3 - \alpha \mathrm{i}\frac{l^2}{j}. \tag{3.28}$$

The antiderivative in the definition of $\mathbf{L}$ makes its Fourier transform singular. For practical computations, the singularity is regularized as:

$$\widehat{\mathbf{L}} = \begin{cases} \mathrm{i}\epsilon^2 j^3 - \alpha\frac{\mathrm{i}l^2}{j} & \text{if } j \neq 0 \\ \mathrm{i}\epsilon^2 (j + \mathtt{eps})^3 - \alpha\frac{\mathrm{i}l^2}{j+\mathtt{eps}} & \text{otherwise,} \end{cases} \tag{3.29}$$

where `eps` $\simeq 10^{-16}$ is the machine epsilon, or the number of digits representable in double precision arithmetic.

The nonlinear operator is expressed as the Fourier transform of the result that follows by computing the nonlinearity in the original variables. This matches closely the strategy used for actual computations, since the steps needed to evaluate the nonlinear operator in this way are:

$$\widehat{\mathbf{N}(u)} = \mathcal{F}\left[\frac{1}{2}\mathrm{i}j\,\left(\mathcal{F}^{-1}\left[\widehat{u}\right]\right)^2\right], \tag{3.30}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the direct and inverse Fourier transforms. The computation in (3.30) requires just $O(N^2 \log N)$ operations due to Fast Fourier Transform algorithms [55]. In contrast, computing the nonlinearity directly in the transformed variables would mean carrying out the following explicit computation:

$$\widehat{\mathbf{N}}(\widehat{u}) = \widehat{u} \star (\mathrm{i}j\widehat{u}), \tag{3.31}$$

requiring $O(N^4)$ operations.

3.2 Time integration

The space discretization algorithm changes dramatically the nature of the approximation problem, in the sense that the initial PDE is replaced by a system of ODEs (3.27) for the Fourier coefficients. The goal of this Section is the description of an appropriate, efficient (not necessarily the best) numerical method for advancing in time the Fourier coefficients according to some approximation of Equation (3.27).

There are two very important observations regarding the system of ODEs (3.27):

- the linear part is purely imaginary;
- the linear operator grows quite rapidly with the wavenumber ($O(j^3)$).

These two facts combined make the problem "stiff", meaning that standard explicit time advancement methods are unlikely to work well on such problem, due to excessive restrictions on the timestep size. Implicit methods however are impractical for large scale computations, since these would require some form of nonlinear iteration, which in turn would need the solution of a large, dense linear system, thus jeopardising one of the greatest advantages of the Fourier representation: the fact that linear operators are diagonal (and as such, can be inverted in $O(N^2)$ operations, versus the $O(N^4)$ operations required by a well-preconditioned iterative method[1]).

In addition to this difficulty, it is important to make sure that the time integrator does not compromise the accuracy of the Spectral Method, and that the numerical errors do not grow too fast, otherwise long-time computations would be completely unreliable.

Some important observations regarding the time integrator that can be helpful in guiding the choice or the development of a numerical scheme are:

1. the integrator should behave well for a purely dispersive, stiff problem (i.e., it should not require a time step size very much smaller than the smallest physical time scale);
2. a high resolution in time is required, and should be mantained for long times. This implies that the time integrator should have at least order 3;
3. fully implicit methods should be avoided, since the cost of nonlinear iterations on the convective term would be prohibitive for the required high space resolution.

It is clear that there is some tension between the requirements 1) and 3). Implicit methods are known to be much more stable than explicit methods with stiff problems, but at the same

---

[1] assuming that a good preconditioner is available, which however does not seem to be the case for this problem, at least to the best of our knowledge. The worst case complexity, assuming a Krylov method without restart and exact arithmetic would be $O(N^6)$, but with restart a Krylov method may well stagnate undefinitely.

time they break rule 3). Strong stability preserving methods, which are very popular for hyperbolic equations, are very efficient and preserve accurately some important properties of the nonlinear part of the equation, but are not suitable for stiff problems, thus breaking rule 1). A class of methods studied in [15, 44] precisely to address this tension between a stiff linear part and an explicit, yet accurate advancement of the nonlinear part are implicit-explicit (IMEX) methods. IMEX Methods originally relied on multistep architectures, and where extended to the multistage case starting with the work [25] on Implicit-Explicit Runge–Kutta Methods.

In this work, multistage (Runge–Kutta) methods are preferred over multistep methods (such as the Adams methods) due to their simplicity, as the former do not need a starting procedure.

Several state-of-the-art numerical methods suitable for the KP Equations are studied in [81] and [109], and the results of these references point to Driscoll's Composite Runge Kutta method [37], which proved to be the most efficient among fourth order schemes in terms of error vs time step size. Klein suggests to avoid Exponential Time Differencing [64] for purely dispersive equations, because in this case the ϕ functions which appear in such methods are known to perturb the spectrum of the nonlinear part of the equation.

In the remaining part of this section, two new composite Runge Kutta method will be derived with the aim of providing similar performance to Driscoll's method, while being suitable as a predictor/corrector for the future implementation of a Deferred Correction Method [31, 32].

3.2.1 Standard Implicit-Explicit Runge–Kutta Methods

The general theory of Runge–Kutta methods is a classic subject, and is assumed here. For a detailed introduction to the theory of Runge–Kutta Methods, see e.g. [61]. In the following, it is assumed that the time interval $[0, T]$ is divided in M equally spaced subintervals $[t^i, t^{i+1}]$, such that

$$h := t^{n+1} - t^n = \frac{T}{M}. \tag{3.32}$$

The approximate solution to the system of ODEs (3.23) at time t^n is denoted by $u^n := u^N(t^n)$.

Implicit-Explicit Runge–Kutta (here denoted by IMEX-RK) Methods adopt a diagonally implicitRunge–Kutta Method for the linear, stiff part of the equation and an explicit method for the nonlinear part.

The standard form of an s-stage IMEX-RK integrator is:

$$u^{n+1} = u^n + h\sum_{i=0}^{s} b_i\,\mathbf{L}k^i + h\sum_{i=0}^{s} \widehat{b}_i\,\mathbf{N}(t^n + \widehat{c}_i h, k^i) \tag{3.33}$$

$$k^i = u^n + h\sum_{j=0}^{i} a_{ij}\,\mathbf{L}\left(u^n + hb_j k^j\right) + h\sum_{l=0}^{i-1} \widehat{a}_{il}\,\mathbf{N}(t^n + \widehat{c}_i h, u^n + h\widehat{b}_l k^l). \tag{3.34}$$

A scheme is completely defined by the constants $a_{ij}, b_i, c_i, \widehat{a}_{ij}, \widehat{b}_i, \widehat{c}_i$ for $i, j = 1, \ldots, s$, which in this case are real numbers. The coefficients a_{ij}, b_i, c_j are used for advancing the linear part of the equation, while the coefficients $\widehat{a}_{ij}, \widehat{b}_i, \widehat{c}_i$ are used for the remaining terms. A concise way to present an IMEX Runge–Kutta method is by means of its Butcher tableau, that consists in arranging the triples (a_{ij}, b_i, c_i) and $(\widehat{a}_{ij}, \widehat{b}_i, \widehat{c}_i)$ in two side-by-side tables as follows:

$$\begin{array}{c|ccccc}
0 & & & & & \\
c_2 & a_{21} & a_{22} & 0 & 0 & \\
c_3 & a_{31} & a_{32} & a_{33} & 0 & \\
\vdots & \vdots & & & \ddots & \\
c_s & a_{s1} & a_{s2} & a_{s3} & a_{s4} & \ldots \\
\hline
 & b_1 & b_2 & b_3 & b_4 & \ldots
\end{array}
\qquad
\begin{array}{c|cccc}
0 & & & & \\
\widehat{c}_2 & \widehat{a}_{21} & 0 & & \\
\widehat{c}_3 & \widehat{a}_{31} & \widehat{a}_{32} & 0 & \\
\vdots & \vdots & & & \ddots \\
\widehat{c}_s & \widehat{a}_{s1} & \widehat{a}_{s2} & \widehat{a}_{s3} & \ldots \\
\hline
 & \widehat{b}_1 & \widehat{b}_2 & \widehat{b}_3 & \ldots
\end{array} \tag{3.35}$$

This paragraph focuses on third-order, L-stable methods. It will become clear in the following that third-order IMEX-RK methods are a good starting point to construct higher-order methods, as these are a reasonable compromise between accuracy and time step restrictions.

Before discussing the importance of L-stability, it is convenient to recall a few basic notions concerning the linear stability of a numerical method. Linear stability is studied by setting $\mathbf{N} = \mathbf{0}$ and writing down explicitly a single step of the numerical method (3.34). The resulting expression is of the form:

$$u^1 = R(z)u^0, \tag{3.36}$$

with $z = h\,\mathbf{L}$, and $R(z)$ is a rational function of z, that characterises the linear properties of the scheme. For Runge–Kutta schemes, it can be shown [60] that the stability function has the expression

$$R(z) = \frac{\det\left(I - zA + ze_1 b^T\right)}{\det\left(I - zA\right)}, \tag{3.37}$$

where I is the $s \times s$ identity matrix, A is the matrix with entries a_{ij}, b is the vector formed by the b_i, and e_1 is a vector of length s (equal to the length of b), with all entries equal to one. The expression (3.37) can be used in practical stability computations, provided that the number of stages s is not too large.

Definition 1 (Stability region)**.** *The* stability region *of a numerical method is the set $\mathcal{S} \subseteq \mathbb{C}$ where:*

$$\mathcal{S} = \{z \in \mathbb{C} : |R(z)| < 1\}. \tag{3.38}$$

If the stability region contains the left half plane,

$$\{z \in \mathbb{C} : \Re z \leq 0\} \subseteq \mathcal{S}, \tag{3.39}$$

then the method is called A-stable.

The rationale behind the definition of the stability region lies in the repeated application of Equation (3.36),

$$u^n = (R(z))^n\, u^0, \tag{3.40}$$

which means that, under the reasonable hypothesis that $R(z)$ does not depend on t, the numerical solution is bounded for all times if and only if $|R(h\,\mathbf{L})| < 1$, or equivalently if the spectrum of $h\,\mathbf{L}$ belongs to $\mathcal{S}$.

As for L-stability, the definition adopted here is the one introduced in [61, Ch. IV.3], namely:

Definition 2 (L-stability)**.** *A method is called* L-stable *if it is A-stable and in addition:*

$$\lim_{z \to -\infty} R(z) = 0. \tag{3.41}$$

L-stability is a desirable property for a numerical scheme because the fact that the stability function vanishes at ∞ guarantees that those parts of the spectrum of $\mathbf{L}$ that are further away from the origin, which are responsible for the strictest restrictions on the time step, will be well within the stability region of the scheme and that any spurious, high-frequency oscillations caused by round-off errors will be damped sufficiently fast as $|R(z)| \to 0$.

It can be proved (see e.g. [60, Ch. II.2]) that a third order Runge–Kutta method must obey to the following conditions:

$$\sum_i b_i = 1 \qquad \sum_i \widehat{b}_i = 1 \tag{3.42}$$

$$\sum_i b_i c_i = \frac{1}{2} \qquad \sum_i \widehat{b}_i \widehat{c}_i = \frac{1}{2} \tag{3.43}$$

$$\sum_{i,j} b_i a_{ij} c_j = \frac{1}{6} \qquad \sum_{i,j} \widehat{b}_i \widehat{a}_{ij} \widehat{c}_j = \frac{1}{6}, \tag{3.44}$$

provided that also:

$$c_i = \sum_{j=0}^{i} a_{ij} \qquad\qquad \widehat{c}_i = \sum_{j=0}^{i-1} \widehat{a}_{ij} \tag{3.45}$$

hold for all $i = 1, \dots, s$. It is understood that the in all the equations (3.42), (3.43), (3.44) and (3.45) the sums run from 1 to s, and that $a_{ij} = 0$ for $j > i$ and $\widehat{a}_{ij} = 0$ for $j \geq i$.

Even if the conditions (3.42) to (3.45) are fulfilled separately by the implicit and explicit parts of the method, this is not sufficient to guarantee that overall the IMEX method is third order accurate. In [25] it is proven that the following coupling conditions are required for a third order IMEX-RK method:

$$\sum_i \widehat{b}_i c_i = \frac{1}{2} \qquad\qquad \sum_i \widehat{b}_i c_i^2 = \frac{1}{3} \tag{3.46}$$

$$\sum_{i,j} \widehat{b}_i \widehat{a}_{ij} c_j = \frac{1}{6} \qquad\qquad \sum_{i,j} b_i \widehat{a}_{ij} c_j = \frac{1}{6} \tag{3.47}$$

$$\sum_{i,j} \widehat{b}_i a_{ij} c_j = \frac{1}{6}. \tag{3.48}$$

In this work, the following method, introduced in [37] will be regarded as the reference IMEX-RK method of order 3:

$$\begin{array}{c|cccc}
0 & 0 & & & \\
\frac{1}{2} & \frac{1}{6} & \frac{1}{3} & & \\
\frac{1}{2} & \frac{1}{2} & -1 & 1 & \\
1 & 0 & 0 & \frac{2}{3} & \frac{1}{3} \\
\hline
 & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6}
\end{array}
\qquad\qquad
\begin{array}{c|cccc}
0 & 0 & & & \\
\frac{1}{2} & \frac{1}{2} & 0 & & \\
\frac{1}{2} & 0 & \frac{1}{2} & 0 & \\
1 & 0 & 0 & 1 & 0 \\
\hline
 & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6}
\end{array} \tag{3.49}$$

In addition of being third order, this method is L-stable. Among the family of IMEX-RK method, the (3.49) is also particularly simple to implement and study, since $b_i = \widehat{b}_i$ and $c_i = \widehat{c}_i$ for all $i = 1, \dots, s$. Written down explicitly, the scheme (3.49) reads:

$$k^1 = u^n \tag{3.50}$$

$$k^2 = \left(1 - \frac{1}{3} h\, \mathbf{L}\right)^{-1} \left(u^n + \frac{1}{6} h\, \mathbf{L} k^1 + \frac{1}{2} h\, \mathbf{N}(k^1)\right) \tag{3.51}$$

$$k^3 = (1 - h\, \mathbf{L})^{-1} \left(u^n + \frac{1}{2} h\, \mathbf{L} k^1 - h\, \mathbf{L} k^2 + \frac{1}{2} h\, \mathbf{N}(k^2)\right) \tag{3.52}$$

$$k^4 = \left(1 - \frac{1}{3} h\, \mathbf{L}\right)^{-1} \left(u^n + h\, \mathbf{L} k^3 + h\, \mathbf{N}(k^3)\right) \tag{3.53}$$

$$u^{n+1} = u^n + \frac{1}{6} h \left[\left(\mathbf{L} k^1 + \mathbf{N}(k^1)\right) + 2\left(\mathbf{L} k^2 + \mathbf{N}(k^2)\right) + 2\left(\mathbf{L} k^3 + \mathbf{N}(k^3)\right) + \left(\mathbf{L} k^4 + \mathbf{N}(k^4)\right)\right]. \tag{3.54}$$

Another property of Driscoll's scheme that is clear from the form (3.54) is the small number of functions evaluations and summations. This is a consequence of the many zeros in the lower triangular part of the Butcher Tableau (3.49).

For time interpolation purposes, the choice of c_i is not satisfactory. Instead, it would be preferable to consider the following midpoints for the stages of the scheme:

$$c_1 = 0 \qquad c_2 = \frac{1}{3} \qquad c_3 = \frac{2}{3} \qquad c_4 = 1. \tag{3.55}$$

The remaining of this section is concerned with the derivation and the subsequent study of a third-order, L-stable IMEX-RK method that is competitive with (3.49), and realises the choice (3.55).

As an initial simplification, in this work only schemes with $b_i = \widehat{b}_i$ and $c_i = \widehat{c}_i$ for all $i = 1, \dots, s$ are considered. Since the $\widehat{c}_i$ coefficients are determined by Equation (3.55), considering the order conditions involving only the b_i and the c_i gives a linear system for the b_i coefficients:

$$\begin{aligned} b_1 + b_2 + b_3 + b_4 &= 1 \\ \frac{1}{3}b_2 + \frac{2}{3}b_3 + b_4 &= \frac{1}{2} \\ \frac{1}{9}b_2 + \frac{4}{9}b_3 + b_4 &= \frac{1}{3}. \end{aligned} \tag{3.56}$$

Imposing for simplicity $b_2 = b_3$ determines the b_i coefficients:

$$b_1 = \frac{1}{8} \qquad b_2 = \frac{3}{8} \qquad b_3 = \frac{3}{8} \qquad b_4 = \frac{1}{8}. \tag{3.57}$$

Additionally, this choice fulfills the fourth order condition:

$$\sum_{i=1}^{4} b_i c_i^3 = \frac{1}{4}. \tag{3.58}$$

As for the coefficients a_{ij}, there are 9 unknowns for the implicit part and 6 for the explicit part, with two constraints from the order conditions (3.44), other two constraints from the hypotheses (3.45) and two from the (3.48). As a result, there are $15 - 8 = 7$ free variables, two of which in this work are arbitrarily set to zero:

$$a_{41} = 0 \qquad a_{42} = 0. \tag{3.59}$$

This choice, together with the constraints discussed above leads to the class of methods characterised by the following Butcher Tableau:

$$\begin{array}{c|cccc|c|cccc} 0 & & & & & 0 & & & & \\ \frac{1}{3} & \frac{1}{3} - \gamma & \gamma & & & \frac{1}{3} & \frac{1}{3} & 0 & & \\ \frac{2}{3} & \gamma + \delta + \frac{1}{3}\zeta & \frac{2}{3} - \gamma - 2\delta - \frac{1}{3}\zeta & \delta & & \frac{2}{3} & \frac{2}{3} - \widehat{a}_{32} & \widehat{a}_{32} & 0 & \\ 1 & 0 & 0 & 1 - \zeta & \zeta & 1 & \widehat{a}_{43} + 3\widehat{a}_{32} - 3 & 4 - 3\widehat{a}_{32} - 2\widehat{a}_{43} & \widehat{a}_{43} & 0 \\ \hline & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} & & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \end{array} \tag{3.60}$$

where for convenience three variables were renamed as: $a_{22} = \gamma$, $a_{33} = \delta$, $a_{44} = \zeta$. The free variables that have to be specified to define the third order method are $\gamma, \delta, \zeta, \widehat{a}_{32}, \widehat{a}_{43}$. Requiring the scheme (3.60) to be L-stable leads to:

$$\lim_{t \to -\infty} R(t) = 0 \tag{3.61}$$

which in this case is achieved by imposing:

$$\delta = \frac{36\gamma\zeta - 9\gamma + 4\zeta^2 - 9\zeta + 2}{3(24\gamma\zeta - 3\gamma - 11\zeta + 2)} \qquad 4\zeta^6 - 104\zeta^5 - 100\zeta^4 + 76\zeta^3 + 5\zeta^2 - 8\zeta + 1 = 0, \tag{3.62}$$

where the second equation in (3.62) is imposed to simplify the subsequent expressions. The root chosen for the equation on the right of (3.62) is:

$$\zeta = \frac{\sqrt{3} - 1}{2}, \tag{3.63}$$

which leads to the scheme:

$$
\begin{array}{c|cccc}
0 & & & & \\
\frac{1}{3} & -\frac{1+\sqrt{3}}{6} & \frac{3+\sqrt{3}}{6} & & \\
\frac{2}{3} & \frac{2+\sqrt{3}}{3} & -\frac{1+\sqrt{3}}{3} & \frac{1}{3} & \\
1 & 0 & 0 & \frac{3-\sqrt{3}}{2} & \frac{\sqrt{3}-1}{2} \\
\hline
 & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8}
\end{array}
\qquad
\begin{array}{c|cccc}
0 & & & & \\
\frac{1}{3} & \frac{1}{3} & 0 & & \\
\frac{2}{3} & 0 & \frac{2}{3} & 0 & \\
1 & 0 & 0 & 1 & 0 \\
\hline
 & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8}
\end{array}
\tag{3.64}
$$

The linear stability function of the scheme (3.64) is:

$$R_m(z) = \frac{(5\sqrt{3}-3)z^2 + 12(\sqrt{3}-1)z - 18}{(z-3)(\sqrt{3}z^2 - 4\sqrt{3}z + 6)}, \tag{3.65}$$

for comparison, the linear stability function of Driscoll's scheme (3.49) is:

$$R_d(z) = \frac{7z^2 + 12z - 18}{2(z-3)^2(z-1)}. \tag{3.66}$$

The stability regions of the two methods are compared in Figure 3.1. The stability regions are remarkably similar for the two methods, and both are concave close to the imaginary axis, a feature that helps ensuring that a purely imaginary spectrum lies inside the stability region even after small perturbations of its eigenvalues. Another way to evaluate the relative performance of the two methods is by highlighting that part of the complex plane where:

$$\frac{|R(z)|}{|R_m(z)|} > 1, \tag{3.67}$$

meaning that in the highlighted region the original method by Driscoll is more stable that the modified method. For the connection between this indicator and the theory of order stars see [62].

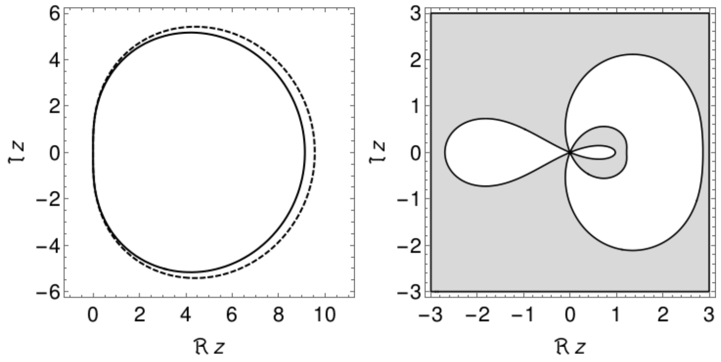


FIGURE 3.1: Left: stability regions for the two Implicit-Explicit Runge–Kutta methods with Butcher Tableaus (3.49) (dashed line) and (3.64) (continuous line). The stability regions are on the exterior part of the curves. Right: region where $|R(z)| \geq |R_m(z)|$, as defined in Equation (3.67)

Finally, it is interesting to inspect the truncation error of the original and of the modified methods. This is usually done by noting that, for a scheme of order p it holds:

$$\mathrm{e}^z - R(z) = Cz^{p+1} + O(z^{p+2}), \tag{3.68}$$

and consequently, a Taylor expansion of the left hand side of Equation (3.68) confirms that the numerical methods have order 3, and allows to compare the quality of the scheme by inspecting the coefficient of the first nonzero term in the expansion. Clearly, it is desirable for C to be as small as possible. For the original method by Driscoll, the expansion is:

$$\mathrm{e}^z - R(z) = \frac{1}{24}z^4 + \frac{89}{1080}z^5 + O(z^6), \tag{3.69}$$

and for the alternative method (3.64) is:

$$e^z - R_m(z) = \frac{2\sqrt{3}-1}{72} z^4 + \frac{49+10\sqrt{3}}{1080} z^5 + O(z^6). \tag{3.70}$$

Consequently, the modified method has indeed order 3 and can rely on an error constant roughly 20% smaller than the original method.

Non L-stable IMEX Runge–Kutta methods

It is interesting to study which possibilities arise when the L-stability requirement (3.41) is replaced by a weaker asymptotic stability constraint, namely:

$$\lim_{z \to -\infty} |R(z)| < 1. \tag{3.71}$$

We repeat the computations described in Section 3.2.1, with the purpose of maximising by trial and error the stability region of the method. This is an attractive feature both in view of nonlinear stability and for ensuring a safety region agains perturbations of the linear operator as in Equation (3.29). This has lead us to the following scheme:

$$\begin{array}{c|cccc}
0 & 0 & & & \\
\frac{1}{3} & -\frac{7}{15} & \frac{4}{5} & & \\
\frac{2}{3} & \frac{1643}{1110} & -\frac{491}{370} & \frac{19}{37} & \\
1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\
\hline
 & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8}
\end{array}
\qquad
\begin{array}{c|cccc}
0 & 0 & & & \\
\frac{1}{3} & \frac{1}{3} & 0 & & \\
\frac{2}{3} & -\frac{13}{66} & \frac{19}{22} & 0 & \\
1 & \frac{313}{418} & -\frac{379}{418} & \frac{22}{19} & 0 \\
\hline
 & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8}
\end{array} \tag{3.72}$$

The stability function of the method (3.72) shows that the method is indeed third order:

$$e^z - R(z) = \frac{301}{4440} z^4 + O(z^5), \tag{3.73}$$

even though the error constant is larger than the schemes (3.49) and (3.64). The difference that the requirement (3.71) induces on the scheme is well visible on a Nyquist diagram, that consists in a plot of the image under the map $R(z)$ of the imaginary axis. In Figure 3.2, the Nyquist plots for the three IMEX-RK methods (3.49), (3.64) and (3.72) are shown. From this plot, it is clear that the L-stability requirement makes the image of $\{\Re z = 0\}$ to pass through the origin, with a winding number of one. Despite the loss of L-stability, the new method (3.72) has actually a larger linear stability domain than both methods (3.49) and (3.64), and is less oscillatory than (3.49) on a large portion of the left half plane, as can be seen from Figure 3.3.

3.2.2 Extrapolation Methods

There exist several strategies to accelerate the convergence rate of an approximating sequence. For example, Aitken Extrapolation and the Shanks Transformation are popular tools in several areas of Approximation Theory, as are usually involved with root finding or fixed point iterations. A valuable source of information concerning extrapolation methods in the context of ODEs is [60], and for recent developments regarding the use of extrapolation methods in the context of parallel-in-time algorithms is [79].

Let us denote with $u_h(t)$ the approximate solution, evaluated at time t, of a differential problem with a numerical method of order p and time step h. Then, the global error expansion has the form:

$$u(t) - u_h(t) = c_1 h^p + c_2 h^{p+1} + O(h^{p+2}), \tag{3.74}$$

where $u(t)$ is the exact solution at time t. The same expression for the same numerical method with a halved step size, $h/2$, is:

$$u(t) - u_{\frac{h}{2}}(t) = c_1 \left(\frac{h}{2}\right)^p + c_2 \left(\frac{h}{2}\right)^{p+1} + O(h^{p+2}). \tag{3.75}$$

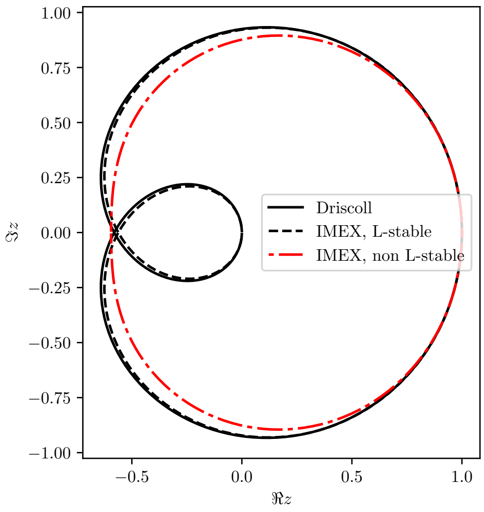


FIGURE 3.2: Nyquist plot for the linear stability functions of the three Implicit-Explicit Runge–Kutta schemes (3.49), (3.64) and (3.72). The two L-stable schemes (3.49) and (3.64) are clearly identified by the winding number equal to one. All the three curves are oriented in the counterclockwise direction if the imaginary axis is transversed from $-\mathrm{i}\infty$ to $\mathrm{i}\infty$.

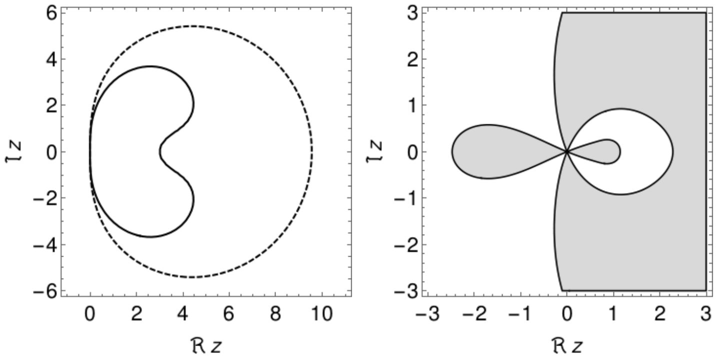


FIGURE 3.3: Left: stability regions for the two Implicit-Explicit Runge–Kutta methods with Butcher Tableaus (3.49) (dashed line) and (3.72) (continuous line). The stability regions are on the exterior part of the curves. Right: the shaded region is defined by the inequality $|R(z)| \geq |R_m(z)|$, as in Equation (3.67).

Supposing that both approximations u_h and $u_{\frac{h}{2}}$ are actually computed, it is possible to eliminate the first term on the right hand side from the expansions (3.74) and (3.75) as follows:

$$(1-2^p)\,u(t)-u_h(t)+2^p u_{\frac{h}{2}}(t)=\left(1-\frac{1}{2}\right)c_2h^{p+1}+O(h^{p+2}),\tag{3.76}$$

which can be rearranged to give:

$$u(t)=\frac{2^p}{2^p-1}u_{\frac{h}{2}}(t)-\frac{1}{2^p-1}u_h(t)+O(h^{p+1}).\tag{3.77}$$

The expression (3.77) is one of the most basic examples of extrapolation techniques to accelerate the convergence rate of a sequence of approximations. In this work, the extrapolation (3.77) with $p=4$ is used in combination with the third order scheme (3.64), advanced with two time steps h and $h/2$, to extrapolate a fourth order, L-stable scheme.

3.2.3 Composite Implicit-Explicit Runge–Kutta Methods

In his influential paper [37], Driscoll made a clever observation which allowed him to preserve the good stability properties of third order IMEX-RK methods while improving (3.49) almost to fourth order.

The idea of Composite methods, introduced first in [51] in the context of multistep methods, consists in dividing the Fourier modes in (at least) two regions: "fast" modes and "slow" modes, depending on the magnitude of the linear operator associated to each mode. High-frequency modes are those responsible for the stiffness of a time advancement problem, so these should be treated with an implicit method in order to avoid an excessively restrictive stability condition on the time step. Conversely, low-frequency modes do not require such a strong stability constraint. Moreover, if the solution is sufficiently smooth, its Fourier coefficients will decay very fast, so only a small fraction of the total energy resides in the high-frequency modes.

Such considerations lead to the idea of a "composite" method, which takes advantage of explicit, high order methods for advancing the highly energetic but nonstiff slow modes, combined with an L-stable, diagonally implicit, but lower order method for advancing the stiff but low energy fast modes. In principle, it is possible to construct Composite methods by choosing (at least) two methods for the fast and slow frequencies, which in turn could take advantage of different methods for the linear and nonlinear parts, as done in IMEX methods.

Driscoll proposed a Composite Runge–Kutta method by combining the L-stable, third order method (3.49) for the fast modes with the explicit, fourth order method on the right hand side of the Tableau (3.49) for the slow modes. As for the distinction of fast and slow modes, Driscoll proposed simply to consider as slow all those modes for which the product of time step and linear operator fits within the stability region of the explicit method:

$$h|\,\mathbf{L}_j|\leq 2\sqrt{2},\tag{3.78}$$

where $2\sqrt{2}$ is the stability ordinate for any explicit Runge–Kutta method of order 4.

For the IMEX-RK method (3.64), it is possible to check that the nonlinear part is advanced only with a method of order 3. Indeed, the stability function for the nonlinear method is:

$$R_n(z)=1+z+\frac{z^2}{2}+\frac{z^3}{6}+\frac{z^4}{36},\tag{3.79}$$

from which it follows that:

$$\mathrm{e}^z-R_n(z)=\frac{z^4}{72}+O(z^5),\tag{3.80}$$

and since the coefficient of the term z^4 is nonzero, the method is only third order. For the nonlinear part of scheme (3.72), the stability function is:

$$R(z)=1+1+\frac{z^2}{2}+\frac{z^3}{6}+\frac{z^4}{24},\tag{3.81}$$

with expansion:

$$e^z - R(z) = \frac{z^5}{120} + O(z^6), \tag{3.82}$$

confirming that the method has indeed order four. For these reasons, the schemes (3.64) and (3.72) cannot be employed as they are for the construction of a Composite method, but need to be supplemented by an explicit method of order at least 4, for the slow unknowns.

An explicit Runge–Kutta method suitable for extending the IMEX scheme (3.64) to a Composite method needs, in addition of being high order, to comply with the choice made for the coefficients b_i and c_i. This is required both to reduce the code complexity and to have a good starting point for the construction of more sophisticated schemes. The general form for the "slow" method is:

$$\begin{array}{c|cccc}
0 & & & & \\
\frac{1}{3} & \frac{1}{3} & 0 & & \\
\frac{2}{3} & a_{31} & a_{32} & 0 & \\
1 & a_{41} & a_{42} & a_{43} & 0 \\
\hline
 & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8}
\end{array} \tag{3.83}$$

To simplify the notation, two of the coefficients will be renamed as $a_{32} = \gamma$ and $a_{43} = \delta$. The conditions up to order 3 give:

$$\begin{array}{c|cccc}
0 & & & & \\
\frac{1}{3} & \frac{1}{3} & 0 & & \\
\frac{2}{3} & \frac{2}{3} - \gamma & \gamma & 0 & \\
1 & 3\gamma + \delta - 3 & 4 - 3\gamma - 2\delta & \delta & 0 \\
\hline
 & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8}
\end{array} \tag{3.84}$$

The stability function for the scheme of Tableau (3.84) is:

$$R_s(z) = 1 + z + \frac{z^2}{2} + \frac{z^3}{6} + \frac{1}{24}\gamma\delta z^4, \tag{3.85}$$

and from the series expansion:

$$e^z - R_s(z) = \frac{1}{24}\left(1 - \gamma\delta\right) z^4 + \frac{1}{120} z^5 + O(z^6) \tag{3.86}$$

it is clear that the method has order 4 if and only if:

$$\gamma\delta = 1. \tag{3.87}$$

There is one free variable, so a possible choice to simplify the expressions is $\gamma = 1$. The resulting scheme coincides with a fourth order method that was derived by Kutta himself (see [24] for the history of Runge–Kutta methods), sometimes called "the 3/8-rule":

$$\begin{array}{c|cccc}
0 & & & & \\
\frac{1}{3} & \frac{1}{3} & 0 & & \\
\frac{2}{3} & -\frac{1}{3} & 1 & 0 & \\
1 & 1 & -1 & 1 & 0 \\
\hline
 & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8}
\end{array} \tag{3.88}$$

Written down explicitly, the Composite method reads:

$$u^{n+1} = u^n + \frac{1}{8}h\left[\mathbf{F}(k^1) + 3\,\mathbf{F}(k^2) + 3\,\mathbf{F}(k^3) + \mathbf{F}(k^4)\right], \tag{3.89}$$

with

$$k^1_{\mathscr{S}} = u^n \tag{3.90}$$

$$k^1_{\mathscr{F}} = u^n \tag{3.91}$$

$$k^2_{\mathscr{S}} = u^n + \frac{1}{3} h\, \mathbf{F}(k^1) \tag{3.92}$$

$$k^2_{\mathscr{F}} = \left(1 - \frac{3+\sqrt{3}}{6} h\, \mathbf{L}\right)^{-1} \left(u^n + h\left(-\frac{1+\sqrt{6}}{6}\, \mathbf{L} k^1_{\mathscr{F}} + \frac{1}{3}\, \mathbf{N}(k^1)\right)\right) \tag{3.93}$$

$$k^3_{\mathscr{S}} = u^n + \frac{2}{3} h \left(\mathbf{F}(k^2) - \frac{1}{3}\, \mathbf{F}(k^1)\right) \tag{3.94}$$

$$k^3_{\mathscr{F}} = \left(1 - \frac{1}{3} h\, \mathbf{L}\right)^{-1} \left(u^n + h\left(\frac{2+\sqrt{3}}{3}\, \mathbf{L} k^1_{\mathscr{F}} - \frac{1+\sqrt{3}}{3}\, \mathbf{L} k^2_{\mathscr{F}} + \frac{2}{3}\, \mathbf{N}(k^2)\right)\right) \tag{3.95}$$

$$k^4_{\mathscr{S}} = u^n + h\left(\mathbf{F}(k^3) - \mathbf{F}(k^2) + \mathbf{F}(k^1)\right) \tag{3.96}$$

$$k^4_{\mathscr{F}} = \left(1 - \frac{\sqrt{3}-1}{2} h\, \mathbf{L}\right)^{-1} \left(u^n + h\left(\frac{3-\sqrt{3}}{2}\, \mathbf{L} k^3_{\mathscr{F}} + \mathbf{N}(k^3)\right)\right) \tag{3.97}$$

$$\tag{3.98}$$

and

$$\begin{cases} k^i := k^i_{\mathscr{F}} & \text{if } h|\mathbf{L}| < 2\sqrt{2} \\ k^i := k^i_{\mathscr{S}} & \text{if } h|\mathbf{L}| \geq 2\sqrt{2} \end{cases} \qquad \mathbf{F} := \mathbf{L} + \mathbf{N}. \tag{3.99}$$

A similar strategy is used to set up Composite methods based on the scheme (3.72) for the high frequencies, combined with (3.88) for the low frequencies.

3.3 Numerical accuracy validation

The quality of the numerical solution is validated in two ways. First, the space resolution is checked by inspecting the magnitude of Fourier coefficients. Secondly, the time resolution is checked by inspecting the numerical conservation of the L^2 norm, which is an integral of the motion.

Definition 3 (Space criterion)**.** *If any Fourier coefficient with magnitude greater than the machine epsilon (*$\simeq 10^{-16}$ *in double precision floating-point arithmetic) lies outside the current resolution (after filtering, if applied), the computation is rejected.*

This criterion may seem to be not very practical, since the coefficients outside resolution are unknown. A possible workaround is to check that for all $l = 0, \ldots, N-1$, the curves $|\widehat{u}^N_{jl}|$, seen as functions of j, decrease down to machine epsilon within the available resolution (i.e. for $j \leq N-1$). In order to comply with the second requirement, we check the conservation of the L^2 norm. In Spectral Methods, the L^2 norm is particularly simple to compute, thanks to Parseval's theorem:

$$\|u\|^2_{L^2} = \sum_{j,k=-N/2}^{N/2-1} |u_{jk}|^2, \tag{3.100}$$

and the numerical results are deemed acceptable according to the following criterion.

Definition 4 (Time criterion)**.** *Provided that the space criterion of Definition 3 holds, the numerical solution is acceptable as long as the* L^2 *norm of the numerical solution is conserved up to a given tolerance* `tol`*. In practice, this latter requirement is computed as:*

$$\left| 1 - \frac{\left(\sum_{j=-N/2}^{N/2+1} \sum_{k=-N/2}^{N/2+1} |u_{jk}|^2\right)^{\frac{1}{2}}}{\|u(x,y,t=0)\|_{L^2}} \right| \leq \texttt{tol}. \tag{3.101}$$

TABLE 3.1: Main features of the time integration schemes considered for the comparison of Section 3.4.

short name	order	details
DIRK3	3	L-stable Diagonally-implicit RK used in [37]
IERK3L	3	L-stable Implicit-Explicit RK method of Table (3.64)
IERK3W	3	non L-stable Implicit-Explicit RK method of Table (3.72)
CDIRK3	$3 \div 4$	Driscoll's Composite RK method [37]
CIERK3L	$3 \div 4$	Composite method based on the IERK3L, with slow-fast splitting (3.105)
CIERK3W	$3 \div 4$	Composite method based on the IERK3W, with slow-fast splitting (3.105)
EIERK3W	4	Extrapolation method based on IERK3W

For the present work the tolerance adopted is `tol` $= 10^{-5}$.

3.4 Comparison

A natural choice to compare the performance of the numerical methods described in the previous sections and summarized in Table 3.1, is to evolve numerically an exact, periodic solution of the KPI Equation. In principle there are many possible choices, such as line solitons, cnoidal waves, lumps, or explicit solutions given in terms of theta functions on a Riemann surface. However, each of these solutions has its own disadvantages such as trivial depedence on y (line solitons, cnoidal waves), algebraic decrease at infinity (lumps), or an involved explicit representation (theta functions). All these considerations reduce the number of “practical” exact solutions. One possibility, considered also in the works [81], [109] is the Zaitsev solution [119]:

$$u(\xi, y) = 2\alpha^2 \frac{(1-\beta\cosh(\alpha\xi))\cos\delta y}{\cosh\alpha\xi - \beta\cos\delta y}, \tag{3.102}$$

with parameters given by:

$$\xi = x - ct \qquad c = \alpha^2\frac{4-\beta^2}{1-\beta^2} \qquad \delta = \alpha^2\sqrt{\frac{3}{1-\beta^2}}. \tag{3.103}$$

In [81, 109], the Zaitsev solution is the basis for a comparison of two Exponential Time Differencing schemes (the fourth order methods by Cox and Matthews [34] and Krogstad [90] and the fifth order method by Ostermann [95]), a fourth order Integrating Factor Runge–Kutta, and Driscoll's Composite Runge–Kutta scheme (3.49).

For the numerical test, the parameters chosen are:

$$\alpha = 1 \qquad \beta = \frac{1}{2} \tag{3.104}$$

and the solution is evolved up to time $t = 2$. In [83] the authors show that Zaitsev's solution is unstable, however the numerical method is sufficiently accurate to guarantee that the unavoidable roundoff perturbations do not grow significantly up to $t = 2$. The computational domain is $[-5\pi, 5\pi] \times [0, 10\pi]$ and the number of Fourier coefficients in the two directions are respectively 2^{11} and 2^9, for a total of 1 048 576 unknowns. Zaitsev's solution changes relatively slowly in both directions, so the grid $2^{11} \times 2^9$ is sufficient to comply with the resolution requirements discussed in Section 3.3, namely that all the Fourier coefficients with magnitude greater than the smallest representable number in double precision floating-point arithmetic are represented. The space discretization and resolution is the same for all the time advancing schemes considered here. The computations are performed on 32 Intel Xeon E5-2640 cores, and the results are available in Figure 3.4. On Figure 3.4 (left), the L^2 norm of the error is shown as a function of the time step size h.

On Figure 3.4 (right), the same L^2 norm of the error is plotted as a function of the total computational time. All of the methods seem to achieve only third order accuracy, which probably reveals some form of order reduction phenomenon, that requires further investigation. However, for the Extrapolation Method the error constant is clearly smaller, and as a result the error for a fixed time step is constantly one order of magnitude smaller than the IMEX methods. Despite this, the higher number of operations required by the Extrapolation method does not yield any advantage when compared in terms of actual computational time. The three implicit-explicit methods seem to behave similarly, with the original DIRK3 that proves to be effective across a larger region of time steps sizes, that are precluded to the other methods. The composite methods, despite failing to achieve fourth order accuracy in this test, are certainly the most effective, but Driscoll's original Composite method seems to be stable only for very small time step sizes. The reason for this behaviour is unclear and deserves further investigation. The stability of composite methods can be greatly improved if the slow-fast splitting (3.78) is modified as:

$$h|\mathbf{L}| \leq \frac{2\sqrt{2}}{15}, \tag{3.105}$$

which is not justified rigorously, but by trial and error seems to be a good compromise, as testified by the greatly improved stability of the two composite methods CIERK3L and CIERK3W.

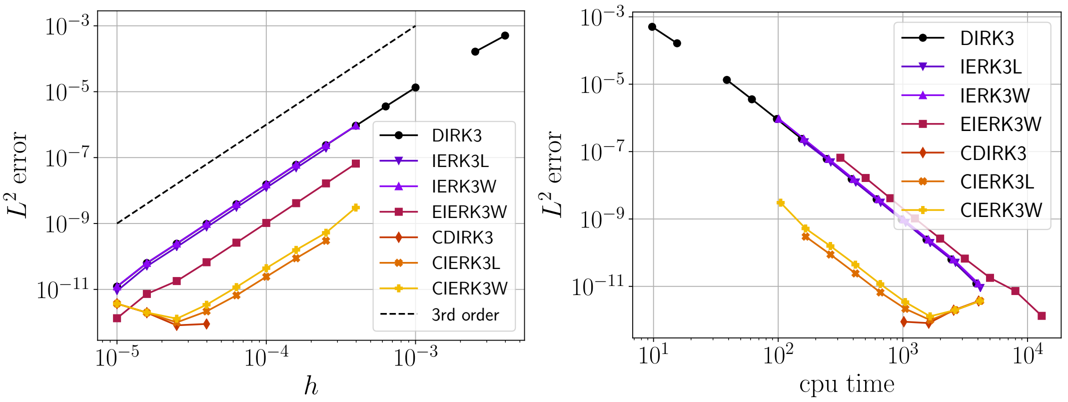


FIGURE 3.4: Error size measured in the L^2 norm as function of the time advancement step h (left) and as function of the computational time (right) required on 32 cores for a problem with approximately 1 million unknowns.

Despite the obstruction to fourth order convergence rate, it is still convenient to exploit composite Runge–Kutta methods, due to their stability region about as large as DIRK Methods, but with errors smaller by two orders of magnitude for a fixed time step size.

We conclude the chapter with a comment on the energy distribution between the two frequency regions of a Composite Runge–Kutta method. To this end, a suitable numerical test consists of a run with $\epsilon = 0.1$ and resolution $2^{13} \times 2^{13}$. The computation is performed until $t = 1$, for three time step values: $h = 10^{-4.1}$, $h = 10^{-4.7}$ and $h = 10^{-5.1}$. For each value of the time step, the energy distribution between fast and slow wavenumbers is computed and the time evolution of this ratio is shown in Figure 3.5. Initially, the solution is well described by a relatively small number of Fourier components, and practically all of the energy is stored in the slow modes. After some time, the first lump appears, redistributing some of the energy from the slow to the fast modes. This can be seen by the sharp drop in the slow energy ratio at $t \simeq 0.15$ in Figure 3.5. For longer times, the lump interacts with the dispersive waves re-entering the domain through the periodic boundaries, and this introduces some further energy exchange between slow and fast modes, corresponding to the oscillations between $t \simeq 0.6$ and $t \simeq 1$ in Figure 3.5. Despite of the lump formation and the relative energy increase in high-frequency components, more than 99% of the energy lies in the low-frequency modes. This test provides further justification to the application of Composite methods (even if with a different partition threshold) to purely dispersive PDEs, since it shows that the vast majority of the energy is evolved with a fourth order method.

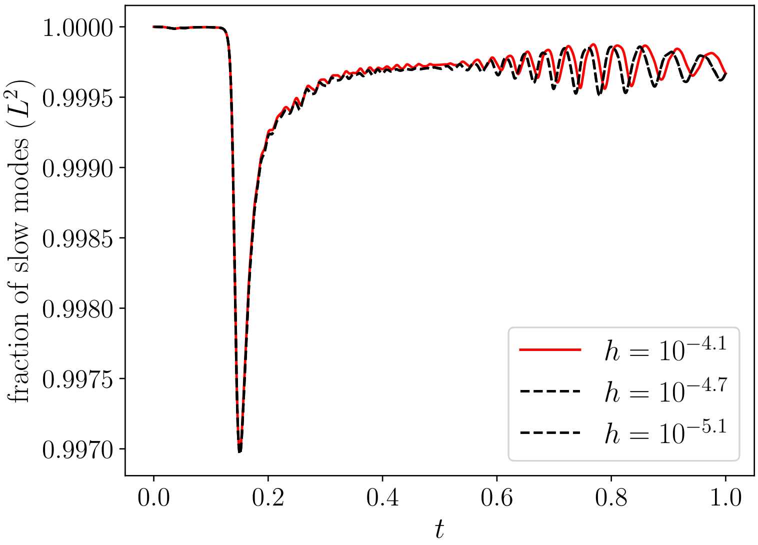


FIGURE 3.5: Fraction of the L^2 norm lying in the slow modes for the Composite Runge–Kutta Method based on the scheme (3.72) for the fast modes, and on the scheme (3.88) for the slow modes. The two cases with $h = 10^{-4.7}$ and $h = 10^{-5.1}$ are plot with the same line style since these would not be distinguishable anyways.

Chapter 4

Parallel Implementation

In this work, numerical methods are considered as a complementary tool to Whitham's theory for the study of dispersive shock wave formation and propagation in the KPI Equation. Fourier Spectral Methods provide uniform resolution throughout the spatial domain. As soon as the solution develops a region of high frequency oscillations in a small part of the domain, such oscillations can be faithfully represented at the cost of introducing a very large number of (collocation or quadrature) points even in regions where the solution is changing slowly, which usually cover most of the domain. Nevertheless, it is possible to tolerate this burden to a large extent thanks to the availability of highly efficient algorithms with computational complexity of $O(N \log N)$, where N is the number of Fourier modes, or equivalently the number of grid points, retained by the approximation method.

However, when studying the small dispersion case, ($\epsilon \ll 1$), the number of Fourier modes required for an acceptable computation[1] grows rapidly, and already for $2^{11} \times 2^{11}$ points the memory required by the solution is

$$8 \cdot 2^{22}\text{b} = 32\text{Mb}, \tag{4.1}$$

which is enough to induce a high number of cache misses in most modern CPU architectures, with a noticeable deterioration in the computational efficiency. Considering the fact that for $\epsilon = \frac{1}{10}$, at least $2^{13} \times 2^{14}$ points are required, it follows that at least

$$8 \cdot 2^{27}\text{b} = 1\text{Gb}, \tag{4.2}$$

are required for a case that can hardly be considered in the small dispersion regime.

This situation, intertwining algorithmic and hardware capacity considerations, is typical for a wide range of problems in Scientific Computing. Since the late 1990s [35], the approach of hardware manufacturers relies on providing clusters with a large number of cpu cores, and to encourage programmers to distribute their computations among hundreds or thousands of cores, so that each core contributes only to a small chunk of the computation, which becomes a concerted effort. As for the cpu cores, the memory too is distributed among many nodes, and even if memory bandwidth is not increasing dramatically, the replication of many independent cpu-memory units still makes it possible to achieve a very high throughput. Clearly this approach has some drawbacks, notably the time required to transfer data on a network connecting the nodes and the additional requirement of designing and maintaining a more complex code, where memory distribution and data sharing are much more difficult tasks than in a standard serial code.

4.1 Parallel Strategies

In a Fourier pseudospectral code, the relevant information possesses naturally a tensor product structure. Specifically, the Fourier coefficients $\widehat{u}_{jl}^N$ of the approximate solution can be thought of as the integer points in a lattice of dimension $N_1 \times N_2$, where $N_{1,2}$ are the number of Fourier basis in each direction. A similar consideration holds for the values of the approximate solution at the physical grid points, u_{jl}^N, and for the structure of the linear operator $\mathbf{L}$. A consequence of this simple data structure is that the task of distributing the information between a certain

---

[1] in view of criteria 3 and 4.

number of cores is straightforward, as opposed to other discretization schemes for PDEs (such as Finite Element Methods) which rely on a triangulation of the computational domain, that in turn requires the programmer to deal with much less strucured data sets.

In this context, the most natural choice for distributing the computation between m cpus is that of dividing the square domain in slabs and assigning a slab to each core, as shown in Figure 4.1 (left). Despite the simple implementation, this choice has the intrinsic limitation that while the amount of memory required grows linearly with the number of modes N (that as a first approximation are supposed to be equally distributed in the two directions), the maximum number of allowed cores grows only as $N^{1/2}$. Furthermore, the slab partitioning requires a global (also called *all-to-all*) inter-core communication for each 2D Discrete Fourier Transform, as shown in Figure 4.1 (right).

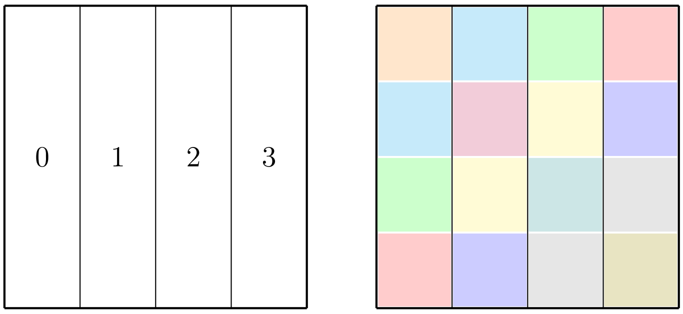


FIGURE 4.1: Left: slab data distribution for a parallel Fourier pseudospectral code. The Fourier coefficients, the values at the gridpoints and the Fourier representation of the linear operator are distributed by dividing the domain in slabs along one of the two directions. The number within each slab denotes the id (or *rank*) of each of the 4 cpus. Right: communication pattern required to compute a 2D Discrete Fourier Transform in the slab distribution. The matching colors need to be exchanged between the respective owners. It is clear that the transposition requires that each cpu communicates with every other cpu, leading to an all-to-all communication pattern.

An alternative to the slab partition is the *pencil* partition, shown in Figure 4.2 (left). In this case the data is more localised, it is possible to scale the number of cpus linearly with N, as opposed to $N^{1/2}$ of the slab configuration, and there is no need of global communication, as shown schematically in Figure 4.2 (right). Some disadvantages of the pencil configuration are the greater code complexity and the lower performance of 1D FFTs on shared data (this second aspect could be partly overcome by using hybrid MPI-OpenMP programming techniques). For a two-dimensional code, usually the burden of the higher code complexity is not rewarded by a significantly higher performance, so the slab strategy is preferred in this work. This perspective is completely different for the case of massively parallel 3D computations, which lead in recent years to the development of several FFT libraries tailored specifically to simplify the pencil-like decomposition [106, 92, 39].

There exist strategies for parallelising the time advancement part of the computation. An intuitive way to parallelise an extrapolation method would be that of assigning the coarse and the fine parts of the computation to two distinct cpu cores. In the case of a computation which is parallel both in space and time, this “stage-parallel” approach would require to set up two communication grids that are called by the FFT routines. This in turn would dramatically complicate the use of efficient hybrid MPI/openMP parallelization strategies, since a subthread would need to be part of an MPI communicator. In addition, since the fine solver would require twice the work of the coarse solver, it would be difficult to ensure a proper load balancing between cpus, with a concrete risk that as many as half of the cpus spend half of their walltime idle.

As a consequence of these difficulties with stage-parallel time integrations, over the last 20 years there have been remarkable developments on step-parallel time stepping schemes, the most popular being the Parareal algorithm [94] with its variants [74], parallel versions of the Spectral Deferred Correction method [30], and hybrid Parareal–Deferred Correction schemes [47]. Such methods however are not considered here, since it will be clear from the

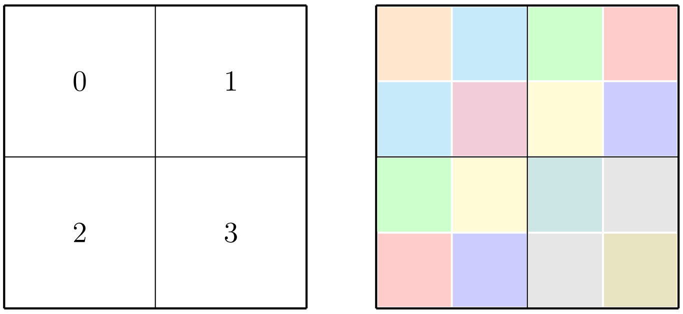


Figure 4.2: Left: pencil data distribution for a parallel Fourier pseudospectral code. The Fourier coefficients, the values at the gridpoints and the Fourier representation of the linear operator are distributed by dividing the domain in subsquares along each direction. The number within each square denotes the id (or *rank*) of each of the 4 cpus. Right: communication pattern required to compute a 2D Discrete Fourier Transform in the pencil distribution. The matching colors need to be exchanged between the respective owners. In this case the transposition requires that only the cpus 1 and 2 exchange data.

scalability results that this additional level of parallelism is not needed as long as the space parallelism is not being fully exploited.

4.2 Code structure

The code structure at large is composed of the following fundamental operations:

1. initialise the MPI communication grid, allocate and initialise memory;
2. start the time stepping cycle:
    i. advance and update the solution;
    ii. check numerical accuracy according to criteria 3 and 4;
    iii. compute some relevant quantity or output data to disk;
3. finalise MPI and deallocate memory.

In a typical computational code, it is expected that the time stepping cycle requires far more cpu time than the initialisation and finalisation tasks. The remaining operations are distinguished between operations that can be performed independently on each cpu and operations that require some data exchange between cpus. Checking the numerical accuracy requires two calls (one for checking mass conservation, and one for the inspection of the Fourier coefficients) to the `MPI_Reduce` primitive, which amounts to an all-to-one communication; advancing for a single time step the numerical integrator requires 2 `MPI_Alltoall` collective calls (required by the Discrete Fourier Transform) for each stage of the Runge–Kutta scheme. As a result, for each time step there are 8 all-to-all calls followed by two all-to-one calls. By checking the numerical accuracy e.g. once every 10 time steps, the time spent on this operation can be minimised, and all the algorithmic efforts can be directed at the optimisation of the Runge–Kutta stages. The typical time marching step for a high order Runge–Kutta scheme is shown schematically in Algorithm 1. From the Algorithm, it is clear that each time step requires 8 Fourier Transforms and 10 MPI collective communications.

It is possible to identify three code blocks that are likely to consume a rather large amount of computational power, as described below.

- Fast Fourier Transforms (and the underlying inter-processor communications). This operation is outsourced to the `FFTW` [53] library, so the only possibility for optimisation is to reduce, if possible, the number of calls to the library.
- A routine for computing the nonlinear term on the physical grid. The related loop involves the same operations for all of the Runge–Kutta stages, and are executed by a subroutine called `update_nonlinearities`.

- A routine for computing the intermediate solutions $\widehat{k}^l$ in Fourier variables. In this case, the operations required at each stage differ significantly and must be treated separately. The subroutines devoted to this computation are called `update_stage_i` in the code, with `i` ranging from 1 to the number of stages s. The solution update at the end of the time step (step 17 in Algorithm 1) is realised by a subroutine called `update_solution`.

4.3 Scalability and profiling

Current cpu architectures are sufficiently complex that a reliable a-priori estimate of the computational time required by each subroutine is practically out of reach. In particular, the complexity level of a modern cpu makes the linear dependence between operation count and computational time obsolete. It is pointless to optimise a certain part of the code based only on considerations related to the number of operations involved. During the development phase of a modern computational code it is necessary to include a profiling phase to check where the computational time is spent and, if possible, to devise new strategies for the parts of the code that the profiler has marked as more resource-intensive.

There exist several high-quality tools for profiling MPI-parallel codes, such as `hpctoolkit` [12], `scalasca` [99] and the `tau` performance system [114], which are capable of returning remarkably detailed information about the time spent by each process on every part of the code. To obtain simpler, less detailed profiling information, a good option comes fron the `MPI_Wtime` directives, that are used in the following for measuring the time required by the main computational blocks in the code.

4.3.1 Strong scalability on Ulysses

We start with a scalability analysis on SISSA's inhouse cluster, Ulysses. Each computational node on Ulysses consists of two Intel Xeon E5-2640 CPUs, for a total of 12 physical cores and 24 virtual cores per node. The scalability analysis considered here consists in a run on a grid with $2^{13}\times2^{13}$ points, which correspond roughly to 67 million unknowns, for $\epsilon = 0.1$, a timestep $h = 10^{-4}$ and for a total of 1000 timesteps. This computation is repeated several times, from 16 up to 512 cpu cores. Since 16 threads are run on every node, arguably some of the cores are running two threads, while the others are probably running a single thread. Despite of this expected imbalance, the cpu manufacturer claims that running a larger number of threads, even if executing on virtual cores, is preferrable to running a smaller number of threads on physical cores only. During each run, the time spent on each routine by every process is measured using `MPI_Wtime`. This kind of scalability test is called *strong scalability test* because the total number of degrees of freedom is fixed. Conversely, a *weak scalability test* consists in fixing the number of degrees of freedom per process, so that the total number of unknowns grows with the cpu count. The former case is the one that mimics more closely the actual practice, and is considered a stricter test for the parallel performance of a code. The results of the strong scalability test are shown in Figures 4.3 and 4.4. From the graphs of Figure 4.3, it is clear that the code scales almost perfectly up to 128 cores, but a further upscaling to 256 and 512 cores, despite still providing a remarkable saving in terms of computational time, is quite off from the ideal scaling line. This behaviour is puzzling; we speculate that it may be attributed to the faulty setup of Ulysses's network interface, as confirmed by some warning messages related to the failure of the TCP connection to a node when 256 or more cpu are used. This however requires further investigation, as this slowdown could be attributed to the network topology design, that may induce a reduction in the average number of Infiniband connections between nodes when the number of nodes is sufficiently high. In the plots of Figure 4.4, the relative importance of the main subroutines is shown as a function of the core count. This plot is of fundamental importance in guiding the optimisation process: the cycle for computing the nonlinear term in the spatial grid requires an overall 15% of the total computational time, and the Fourier transform subroutines take 80% of the time. The cycles for computing the intermediate stages and for the final summation in frequency space contribute to less than 1% of the computing time. Since the Fourier transform is outsourced

Algorithm 1 Operations involved in a typical explicit 4-stage Runge–Kutta integrator. For an implicit-explicit Runge–Kutta and a Composite Runge–Kutta method the main steps are fundamentally the same, with a different expression to take into account the linear-nonlinear and slow-fast splitting.

```
procedure Explicit Runge–Kutta step
    $k^1 = \mathcal{F}^{-1}(\widehat{u}^n)$
    for $i, j = 1, \ldots, N$ do
        $\mathbf{N}^1_{ij} = \mathbf{N}(k^1_{ij})$
        ▷ Computes the nonlinear term on the physical grid based on the solution at the previous time step
    $\widehat{\mathbf{N}}^1 = \mathcal{F}(\mathbf{N}^1)$
    for $i, j = 1, \ldots, N$ do
        $\widehat{k}^2_{ij} = \widehat{u}^n_{ij} + h a_{21} \left( \widehat{\mathbf{L}} \widehat{k}^1_{ij} + \widehat{\mathbf{N}}^1_{ij} \right)$
        ▷ Second stage
    $k^2 = \mathcal{F}^{-1}(\widehat{k}^2)$
    for $i, j = 1, \ldots, N$ do
        $\mathbf{N}^2_{ij} = \mathbf{N}(k^2_{ij})$
    $\widehat{\mathbf{N}}^2 = \mathcal{F}(\mathbf{N}^2)$
    for $i, j = 1, \ldots, N$ do
        $\widehat{k}^3_{ij} = \widehat{u}^n_{ij} + h \sum_{l=1}^{2} a_{3l} \left( \widehat{\mathbf{L}} \widehat{k}^l_{ij} + \widehat{\mathbf{N}}^l_{ij} \right)$
        ▷ Third stage
    $k^3 = \mathcal{F}^{-1}(\widehat{k}^3)$
    for $i, j = 1, \ldots, N$ do
        $\mathbf{N}^3_{ij} = \mathbf{N}(k^3_{ij})$
    $\widehat{\mathbf{N}}^2 = \mathcal{F}(\mathbf{N}^2)$
    for $i, j = 1, \ldots, N$ do
        $\widehat{k}^4_{ij} = \widehat{u}^n_{ij} + h \sum_{l=1}^{3} a_{4l} \left( \widehat{\mathbf{L}} \widehat{k}^l_{ij} + \widehat{\mathbf{N}}^l_{ij} \right)$
        ▷ Fourth stage
    $k^4 = \mathcal{F}^{-1}(\widehat{k}^4)$
    for $i, j = 1, \ldots, N$ do
        $\mathbf{N}^4_{ij} = \mathbf{N}(k^4_{ij})$
    $\widehat{\mathbf{N}}^4 = \mathcal{F}(\mathbf{N}^4)$
    for $i, j = 1, \ldots, N$ do
        $\widehat{u}^{n+1}_{ij} = \widehat{u}^n_{ij} + h \sum_{l=1}^{4} b_l \left( \widehat{\mathbf{L}} \widehat{k}^l_{ij} + \widehat{\mathbf{N}}^l_{ij} \right)$
        ▷ Solution update
```

to a highly optimised library, it follows that any optimisation effort is best spent in trying to improve the efficiency of the nonlinear subroutine. In the second plot, Figure 4.4, the puzzling asymmetry in computational time between forward and backward Fourier transforms in the runs with 256 and 512 cores is even more clearly visible.

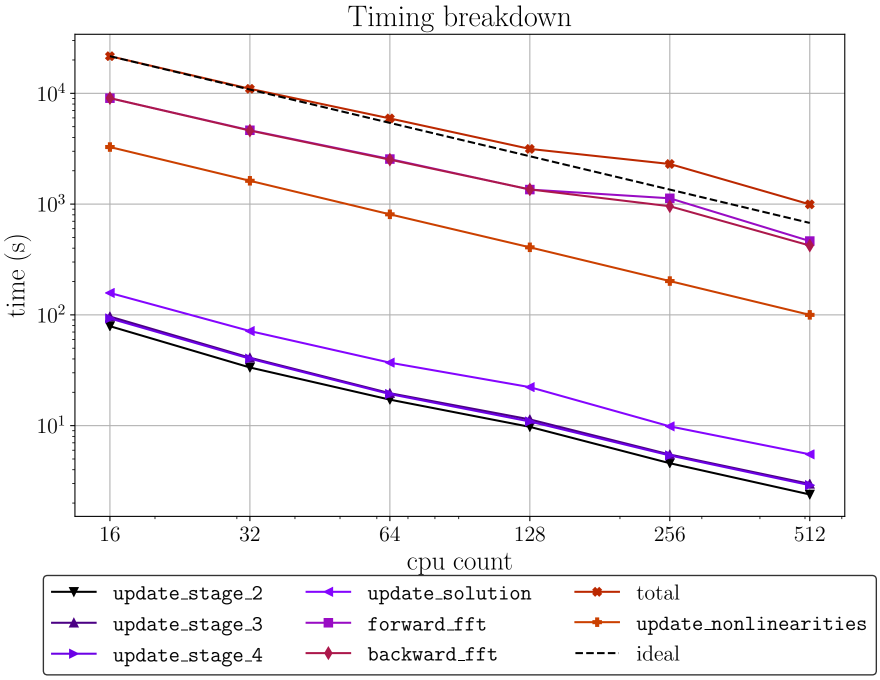


FIGURE 4.3: Results of the strong scalability test with ≃67 million unknowns. The dashed line denotes the ideal case for which the computational time is proportional to the inverse of the core count. The processors are Intel Xeon E5-2640 and the compiler is the GNU Compiler, version 6.2.0 with OpenMPI 2.0.0.

It is interesting to study more in detail the time spent by each process on a given subroutine, and comparing it with the time spent on average by the other processes. This information is shown in Figures 4.5 and 4.6, for the three performance-critical subroutines.

In case of an ideal scaling, the average values (marked with the dashed lines) would divide the plotting area in four horizontal stripes with equal area. The aberrant behaviour for the case with 256 processes is clearly visible from these plots. It is interesting to observe that for the `update_nonlinearities` subroutine, the time required is more or less equally distributed between all the processes, while for the FFT routines this is not the case. We suspect again that this behaviour is related to network installation errors. The fact that the computational times required by the `update_nonlinearities` subroutine are quite balanced between the processes in Figure 4.5 suggests that there is little that can be done on the communication part of the implementation, and that the optimisation effort must be directed to the purely computational part of the subroutine.

To this end, let us write down explicitly the two lines of C code that absorb 15% of the total computational time:

```c
for (ptrdiff_t i=0; i<N; i++)
    nn[i] = cpow(creal(u[i]), 2)/2.;
```

where `u` is an array of dimension N that stores the grid values of the numerical solution and `nn` is an array with the grid values of the nonlinearity $u^2/2$. The code is compiled with the GNU Compiler, version 6.2.0 and OpenMPI version 2.0.0. The optimisation flag used is `-O3`. Both `u` and `nn` are arrays of complex numbers, but since the solution is real, if the software is running correctly these should contain complex values with imaginary part equal to zero, or very small anyways. For this reason, it may be worthwhile to test the following alternative:

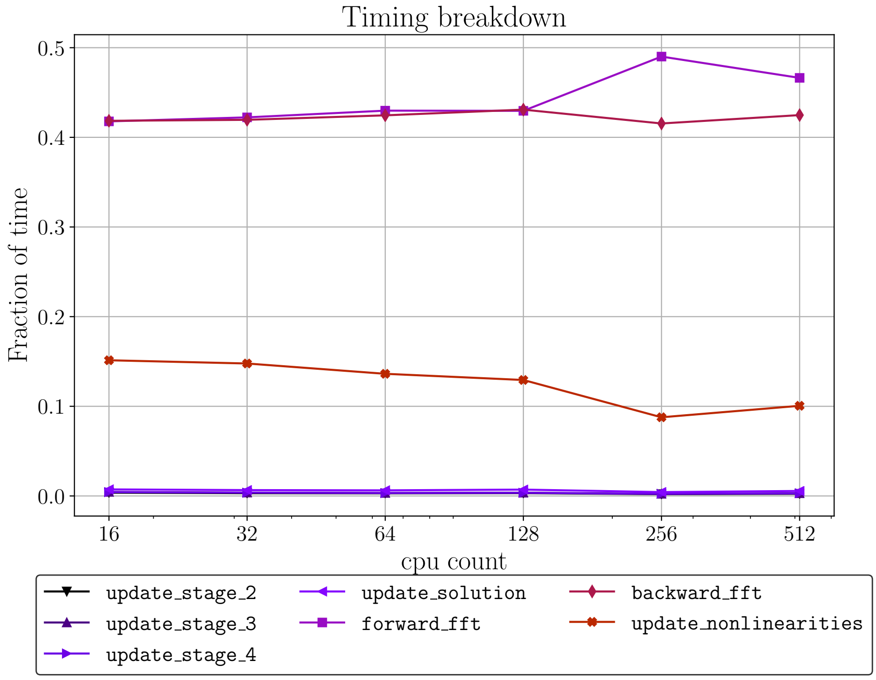


FIGURE 4.4: Results of the strong scalability test with ≃67 million unknowns. In this plot, each line shows the fraction of the total time taken up by the subroutines described in the legend. The processors are Intel Xeon E5-2640 and the compiler is the GNU Compiler, version 6.2.0 with OpenMPI 2.0.0.

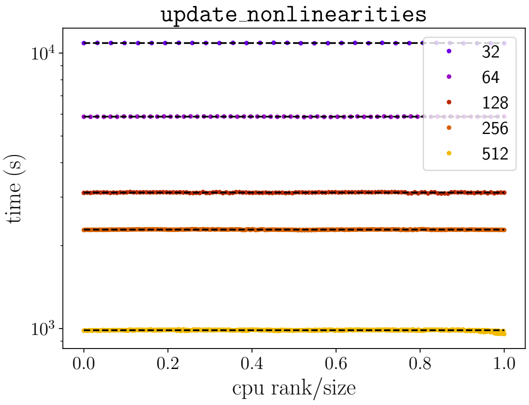


FIGURE 4.5: Time required by the `update_nonlinearities` subroutine for all the processes, for the number of cores shown in the legend. The dashed line correponds to the average value.

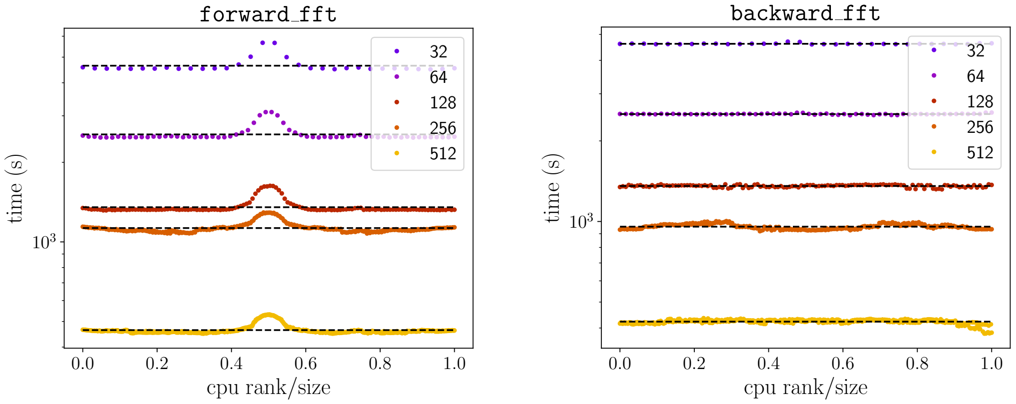


FIGURE 4.6: Time required by the `forward_fft` (left) and `backward_fft` (right) subroutines for all the processes, for the number of cores shown in the legend. The dashed line correponds to the average value.

```
for (ptrdiff_t i=0; i<N; i++)
    nn[i] = u[i]*u[i]/2.;
```

in the hope that since two function calls are avoided, it may enable the compiler to generate a better performing executable. Also, the variable `u` is qualified as `restrict`, to suggest the compiler that there is no pointer aliasing in this subroutine, and further compiler optimisations may be enabled. Replacing the first loop with the second one, we measured a reduction to almost a third of the average computational time for running each loop, with compiler optimisations disabled. However, as soon as compiler optimisations are turned on, there is a very mild difference in the execution time between the two loops.

Alternatively, following the guidelines in [71, 67], it may be worthwhile to directly use the arithmetic functions provided in the Intel Advanced Vector Extensions (AVX). The AVX instruction set takes advantage of the vectorised nature of modern Intel cpus to process 4 double precision entries for each clock cycle. Formally, in the previous loop only a single double precision number is processed at each clock cycle, and we checked by inspecting the disassembled code that there are no compiler optimisations which take advantage of the AVX instruction set. We rewrite the same loop with AVX extensions[2] as follows:

```
for (ptrdiff_t i=0; i<N; i+=4)
{
    __m256d u0 = _mm256_loadu_pd((double*)u);
    __m256d u2 = _mm256_loadu_pd((double*)(u+2));
    __m256d ur = _mm256_unpacklo_pd(u0, u2);

    __m256d tr = _mm256_mul_pd(ur, ur);
    __m256d ti = _mm256_setzero_pd();

    __m256d s0 = _mm256_shuffle_pd(tr, ti, 0b0000);
    __m256d s2 = _mm256_shuffle_pd(tr, ti, 0b1111);

    _mm256_storeu_pd((double*)nn, s0);
    _mm256_storeu_pd((double*)(nn+2), s2);
}
```

After some testing, it was found that the compiler can automatically vectorise the second of the three loops if the flags `-ftree-vectorize -mavx -march=native -ffp-contract=fast` are specified.

The scaling results with the new optimised loop are shown in Figures 4.7 and 4.8. The most striking result of the optimised code is the dramatic reduction in time spent by the `update_nonlinearities` subroutine, which after optimisation takes slightly more than 1% of the total time. As a result, computational times have decreased considerably, in many cases by more than half (see the results in Table 4.1), and more than 90% of the computational time is spent in taking Fourier transforms.

4.3.2 Strong scalability on Marconi

Our experience with the Ulysses cluster is that medium sized computations, with grids up to $2^{15} \times 2^{15}$ ($\simeq 1$ billion) points, can be run at a satisfactory performance. However, for very large scale computations, the time required to perform a long-time simulation is unacceptable. For this reason, we decided to perform our very-high resolution computations on a Tier-0 machine, namely the Marconi cluster at CINECA. The Marconi cluster consists of three partitions,

---

[2] in this case, the loops computes as output just u^2, and not $u^2/2$. The division by 2 is performed later, when the x-derivative is applied in Fourier variables. The reason for this slight difference is that AVX intrinsics do not provide an instruction to perform multiplication and division by 2 within a single clock cycle.

TABLE 4.1: Results of the strong scalability tests on the problem with 67 million unknowns run on the Ulysses cluster. The comparison is between the naive implementation and the loops hand tuned using Intel AVX intrinsics. The compiler is GNU 6.2.0 with OpenMPI 2.0.0.

cpus	time without AVX (s)	time with AVX (s)	% speedup
16	21637	12106	44
32	10990	5691	48
64	5935	2906	51
128	3149	1523	52
256	2302	985	57
512	995	443	55

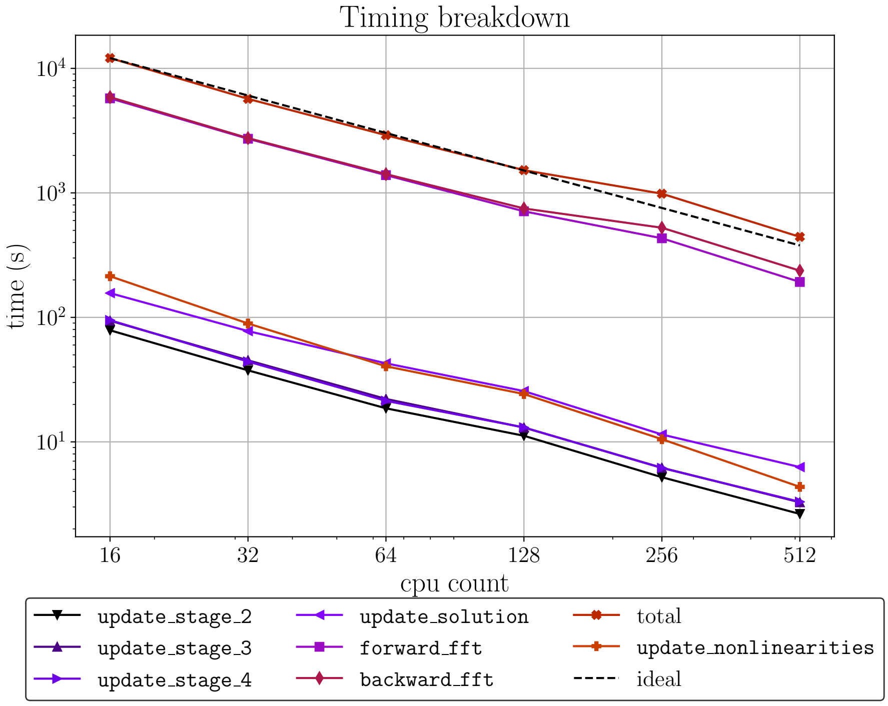


FIGURE 4.7: Results of the strong scalability test with ≃67 million unknowns for the optimised loop using AVX instructions. The dashed line denotes the ideal scaling that would happen if the computational time were scaling exactly as the inverse of the core count. The processors are Intel Xeon E5-2640 and the compiler used is the Intel Compiler, version 14.0 with OpenMPI 1.8.3.

named A1, A2, and A3. The typical node of the first partition (which will be also referred to by the commercial name of Broadwell) consists of a single Intel Xeon E5-2697 v4 processor, with 18 cores and 36 threads, and 126Gb of volatile memory. The second partition takes advantage of the Intel Xeon Phi 7250 accelerator (which in the following will be referred to with the commercial name of Knights Landing, or KNL), providing 68 cores per node with 96Gb of volatile memory and 16Gb of dedicated high-bandwidth cache memory. The third partition is equipped with two Intel Xeon 8160 per node, which offer 48 cores per node. This latter partition is not used in the present work.

Due to the several possible hardware choices, the scalability analysis on Marconi consists in two parts: a first part to test the scalability of our code under Intel Xeon CPUs and a second part to test the scalability under Intel KNL accelerators. Both partitions can take advantage of Intel AVX instructions, which are exploited as described in Section 4.3.1 In addition, the KNL cards have even more vectorisation capabilities thanks to the support of AVX512 directives. These latter directives do not require any change in the AVX code, and to enable the automatic conversion from AVX to AVX512 instructions it is sufficient to provide the Intel compiler with the flags `-O3 -axMIC-AVX512`.

The two tests on Marconi consist of a strong scaling on the Broadwell and then on the KNL partitions with $2^{16} \times 2^{16}$ Fourier modes, corresponding roughly to 4 billion unknowns. In both cases, the compiler is the Intel Compiler, version 18.0.3 with IntelMPI version 2018, update 3.

The scaling results on the Broadwell partition are shown in Figures 4.9 and 4.10. Code scalability is almost ideal from 256 to 2048 cores, but slows down significantly at 4096 cores. The reason for this less than optimal performance with 4096 cores seems to lie in the last two summation stages in the Runge–Kutta integrator. Such summation loops are indeed quite complex, since a relatively large number of sums coming from different arrays are involved. However, these summation loops did not have scalability issues in the very similar Intel Xeon E5-2670 cpus used for the test on the Ulysses cluster. Moreover, each of the last two summation loops is requiring on average 15% of the computational time on Marconi Broadwell, while on Ulysses these loops required less than 5% combined. Repeating the scalability test on the Broadwell partition with the GNU C compiler, version 6.1.0 and OpenMPI version 1.10.3, the abnormal behaviour of the last two loops disappears. In this work the Intel compiler is nevertheless preferred over the GNU compiler, due to its ability to generate significantly more efficient code, and to the errating behaviour of the scalability tests performed with the GNU compiler. This behaviour is puzzling, and deserves further investigation. The other parts of the code, including the Fourier Transforms provided by the library, are scaling very well.

Despite of the less than optimal scalability with very large core counts, still for large scale computations it is convenient to exploit as many cores as possible. This is revealed by the fact that the advancement of 50 time steps for a problem with 4 billion unknowns takes roughly 500 seconds, or 10 seconds per time step, on 4096 cores. On 256 cores, this same computation requires roughly 6000 seconds, or 120 seconds per time step. Since the expected number of time steps for a long-time simulation is of the order of $2 \cdot 10^4$, this means that a high resolution, large scale computation can be completed in 55 hours (of “human” time, equivalent to 225000 cpu hours) on 4096 cores, while the same computation would require 667 hours (equivalent to 170000 cpu hours) on 256 cores. We let the reader ponder on the relative value of human time and cpu time.

The scaling results on KNL accelerators are available in figures 4.11 and 4.12. In this case, the scaling is almost ideal from 1024 to 4096 cores, then the trend departs considerably from the inverse proportionality with respect to core count when passing from 4096 to 8192 cores. As for the Broadwell test, the last two summation loops of the Runge–Kutta integrator are requiring a much larger share (15% each) of the total time than what was observed in the Ulysses cluster. However, on the KNL partition all components of the time integrator, including the demanding loops of the last two stages, scale perfectly up to 8192 cores. In contrast, the Fourier Transforms show a less than optimal scalability in the final part of the test.

In Table 4.2, we compare the performance of regular Xeon cpus and Xeon Phi accelerators, by keeping fixed the number of nodes. A comparison based on the node count has sense from

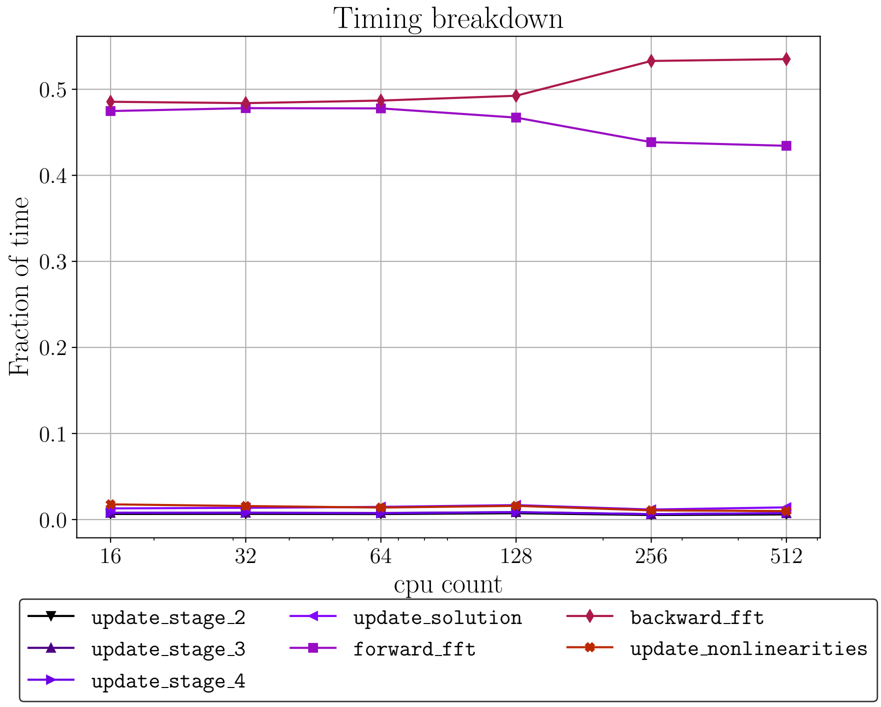


FIGURE 4.8: Results of the strong scalability test with ≃67 million unknowns for the optimised loop using AVX instructions. In this plot, each line shows the fraction of the total time taken up by the subroutines described in the legend. The processors are Intel Xeon E5-2640 and the compiler used is the Intel Compiler, version 14.0 with OpenMPI 1.8.3.

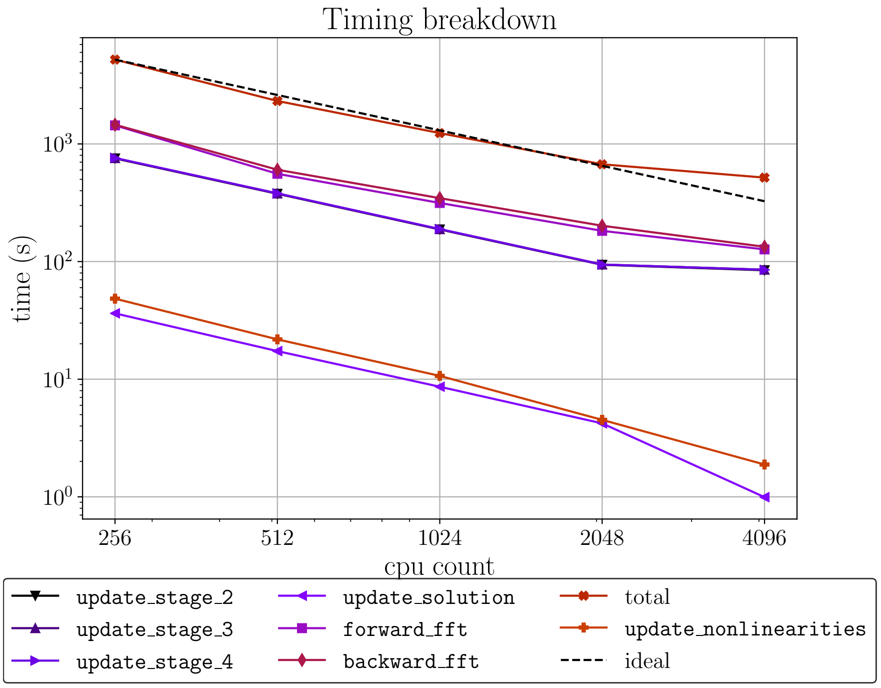


FIGURE 4.9: Results of the strong scalability test with ≃4 billion unknowns on the Marconi Broadwell partition. The dashed line denotes the ideal case where the computational time is proportional to the inverse of the core count. The processors are Intel Xeon E5-2697 v4 and the compiler used is the Intel Compiler, version 18.0.3 with the Intel MPI Library, version 2018, update 3.

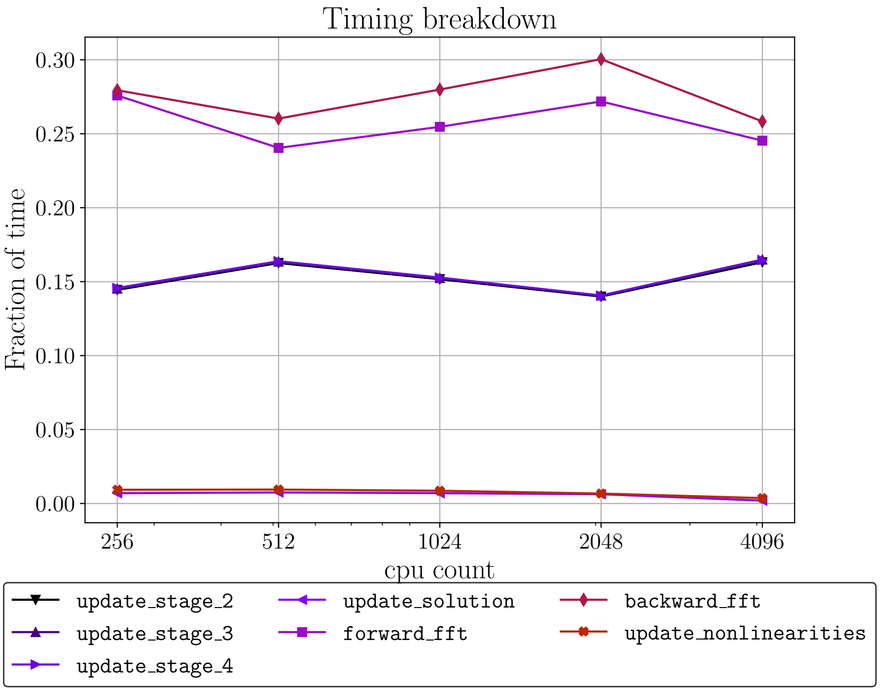


FIGURE 4.10: Results of the strong scalability test with ≃4 billion unknowns on the Marconi Broadwell partition. In this plot, each line shows the fraction of the total time taken up by the subroutines described in the legend. The computations are performed on Intel Xeon Phi 7250 accelerators and the compiler used is the Intel Compiler, version 18.0.3 with the Intel MPI Library, version 2018, update 3.

TABLE 4.2: Time required for a computation with 4 billion unknowns on the Marconi cluster. The performance of Broadwell and KNL chips is compared by equating the number of nodes (each node hosts 32 Broadwell cores and 64 KNL cores).

nodes	Broadwell cores	Broadwell time (s)	KNL cores	KNL time (s)
16	512	2318	1024	2043
32	1024	1236	2048	977
64	2048	671	4096	503
128	4096	517	8192	349

the economic viewpoint (raw cost and energy expense), and additionally a comparison based on the core count is not meaningful since accelerators have a larger number of smaller cores than a standard general purpose cpu. The results of this comparison show that in all the cases, accelerators are more efficient than standard cpus in this kind of computations, with an increase in computational efficiency ranging between 12% and 32%, the higher figures being related to higher core counts (and thus to lower "human" time).

The scalability results shown in this section confirm that modern clusters are sufficiently complex that profiling and optimisation are difficult tasks, and that the results of these activities have a strong variability even between similar machines, and in the same machine between different compilers. Indeed, despite the apparent similarity of the Ulysses and Marconi Broadwell clusters, it was found that a code performing very close to an optimum on Ulysses had issues on Marconi Broadwell, and that the scalability bottlenecks on cpus had little to do with scalability bottlenecks found on accelerators.

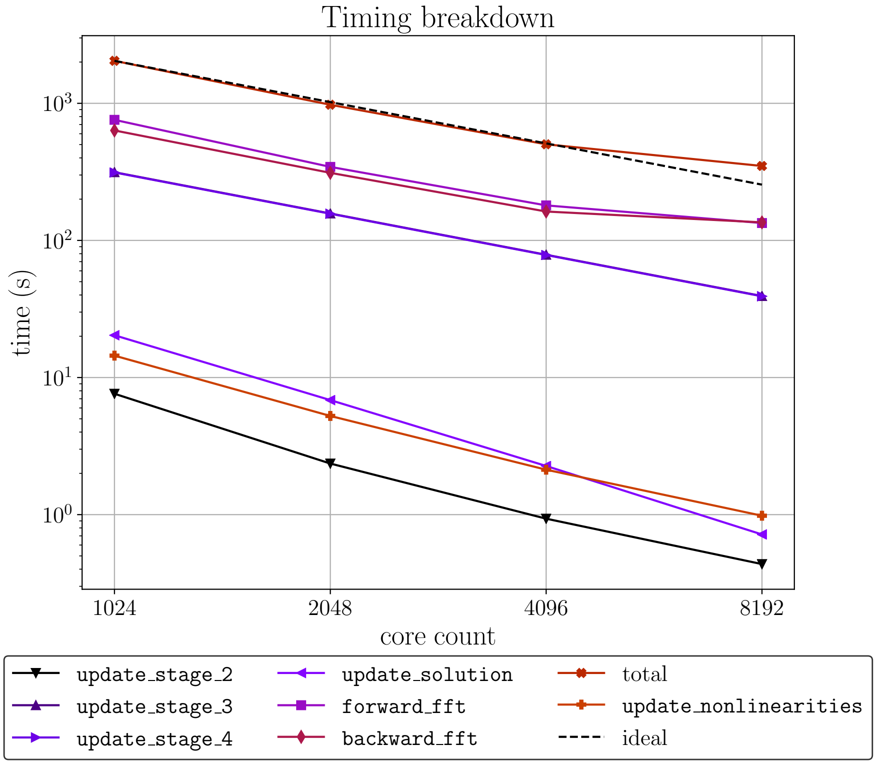


FIGURE 4.11: Results of the strong scalability test with ≃4 billion unknowns on the Marconi KNL partition. The dashed line denotes the ideal scaling that would happen if the computational time were proportional to the inverse of the core count. The computations are performed on Intel Xeon Phi 7250 accelerators and the compiler used is the Intel Compiler, version 18.0.3 with the Intel MPI Library, version 2018, update 3.

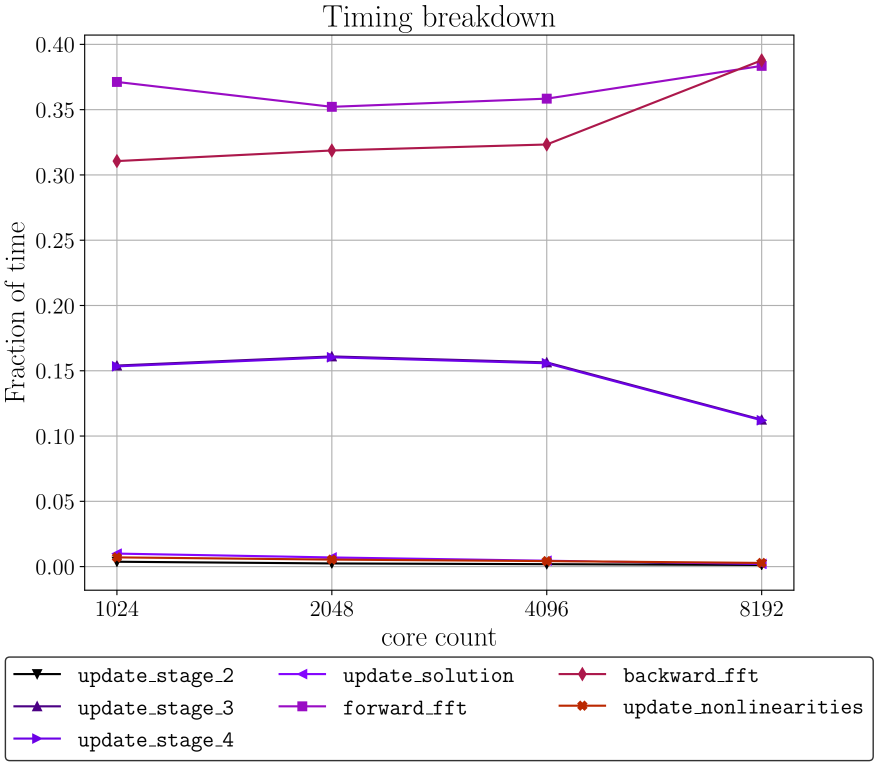


FIGURE 4.12: Results of the strong scalability test with ≃4 billion unknowns on the Marconi KNL partition. In this plot, each line shows the fraction of the total time absorbed by the subroutines described in the legend. The processors are Intel Xeon E5-2697 v4 and the compiler used is the Intel Compiler, version 18.0.3 with the Intel MPI Library, version 2018, update 3.

Chapter 5

Numerical results

We discovered the secondary wave breaking phenomenon and the lump formation in the KPI Equation by means of a numerical computation. This Chapter describes the setup and the results of a systematic, detailed numerical study of the KPI Equation with small dispersion.

In the first part of the chapter, we study the nature of the isolated high peaks seen numerically, and we show that these are practically indistinguishable from the lump exact solution. We then give a qualitative description of the lump-wave interaction that is the main mechanism by which the lumps arrange in a triangular lattice.

Secondly, we pursue a more quantitative study on the scaling properties of the wave breaking phenomenon. This is achieved first by observing numerically the scaling of some relevant quantities, with the norm of the initial data, for a fixed (small) value of ϵ. Then, we repeat the same study for a fixed initial datum, and for varying (small) values of ϵ.

We conclude the chapter by presenting an analogy with the semiclassical limit of the nonlinear Schrödinger Equation, that suggests a path to prove the scalings observed in the previous sections, and to determine other relevant quantities, such as the position of the lumps in a triangular lattice.

5.1 Lump formation mechanism

The peaks appearing in the numerical computations resemble closely the shape of lumps, so it is interesting to check how the numerical structures are close to the lump solution. To answer this question, we fit the expression for the lump solution to the numerical data. Let us recall here the analytic form for a lump symmetric with respect to the x axis:

$$u(x,y,t) = 24\frac{-(x-3b^2t)^2+3b^2y^2+\frac{1}{b^2}}{\left((x-3b^2t)^2+3b^2y^2+\frac{1}{b^2}\right)^2}. \tag{5.1}$$

In this case, b is the only parameter that specifies at once the height, width and speed of the lump. As discussed in Section 1.3.7, the position of the maximum height for the lump is:

$$x_{\max} = 3b^2t. \tag{5.2}$$

It is also quite simple to set up an algorithm to detect the position of $x_{\max}$ from the numerical data. As a result, we determine numerically the position of maximum height $x_{\max}(t)$ of the numerical solution at each time step following the appearance of the first lump, and estimate the parameter b by solving Equation (5.2) for b. The speed of the lump too is estimated numerically by a finite incremental ratio:

$$c(t) = \frac{x_{\max}(t+h)-x_{\max}(t)}{h} \tag{5.3}$$

and is compared with the exact value for the lump's speed, namely $3b^2$, b being estimated from the numerical data as explained above. The results of this comparison are shown in Figure 5.1, where the agreement between the numerical estimate for the lump's speed and the theoretical relation is excellent.

Another, more qualitative, test consists in estimating b by solving Equation (5.2) as described above, and fitting the exact expression (5.1) for the lump with the numerical data in a

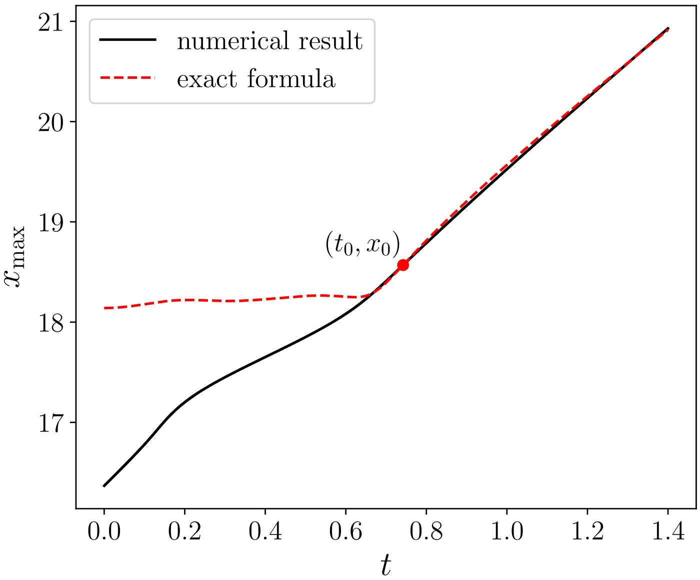


FIGURE 5.1: Horizontal coordinate ($x_{\max}$) of the maximum peak for a numerical solution of the KPI Equation with $\epsilon = 0.10$ and initial datum corresponding to Equation (5.6) with $C_0 = 6$ (case n. 6 of Table 5.2). The numerical results are shown with the black, continuous line, while the position obtained from Equation (5.2), after estimating b by fitting the numerical data is shown with the red, dashed line. The highlighted point, labelled with (t_0, x_0) corresponds to the absolute maximum of $|u|_\infty$, taken for all the times for which the numerical solution was computed.

neighbourhood of $x_{\max}$. This fit is shown in Figure 5.2 for some representative snapshots, with the plots taken along the line $y = 0$. The same comparison is shown as a three-dimensional plot in Figure 5.3, where the best fit to the numerical data is subtracted to the computational results. From this plot, the excellent fit of Equation (5.1) is clear, since in the right plot the peak is removed almost completely.

The fits with the lump solution described in this section are based on a numerical solution with initial datum:

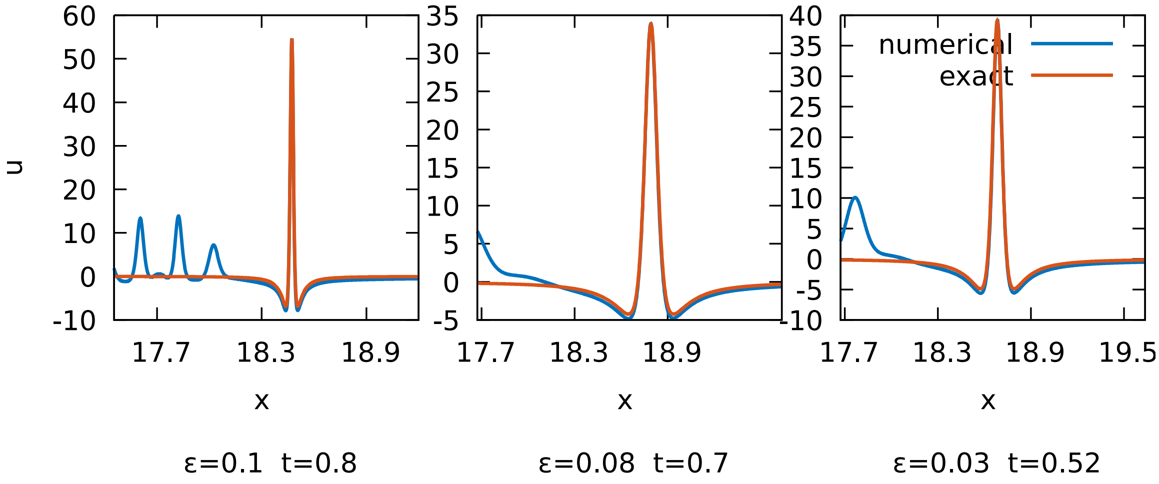
$$u_0(x, y) = -6\partial_x \operatorname{sech}^2 \sqrt{x^2 + y^2}, \tag{5.4}$$

which is a symmetric function with respect to x, namely $u_0(-x, y) = u_0(x, y)$. It may be that the existence of a lump lattice arrangement depends on this symmetric initial datum. To rule out this possibility, we repeat the numerical computations with the initial datum:

$$u_0(x, y) = -6\partial_x \exp\left(-x^2 - xy - y^2\right), \tag{5.5}$$

which does not possess any symmetry. The results for the unsymmetric datum are shown in Figure 5.4. In this case, the dispersive waves propagate in a direction which is not aligned with x, however wave focusing the the formation of a triangular lump lattice still take place. The solitons too are traveling at an angular direction, as testified by the fact that the triangular lattice is in this case no longer symmetric.

We conclude this section by describing more accurately the process by which lumps form and arrange in a triangular array. Our description is inspired by a numerical computation, whose results are depicted in Figures 5.5 and 5.6, and these figures will be often referred to in the following as a visual aid for understanding this phenomenon. The twelve insets that compose Figure 5.6 are numbered here in the western natural reading ordering. The first plot shows the last stages of the focusing process in the first wavefront, that leads to the detachment of a lump, shown in the second plot. As a consequence of the lump formation, a void region is left in the first wavefront, that corresponds to the darker area in the center of the leftmost wave in the second and third plot. In the second and third plot, it is interesting to note that a similar focusing process is taking place along the symmetry axis of the the second wavefront, in a fashion similar to the focusing of the first wavefront for earlier times. The



ε=0.1 t=0.8 ε=0.08 t=0.7 ε=0.03 t=0.52

Figure 5.2: Comparison between the numerical solution and the exact expression for the lump soliton, Equation (5.1). The parameter for the lump solution is fitted to the peak of the numerical solution by solving Equation (5.2) for b.

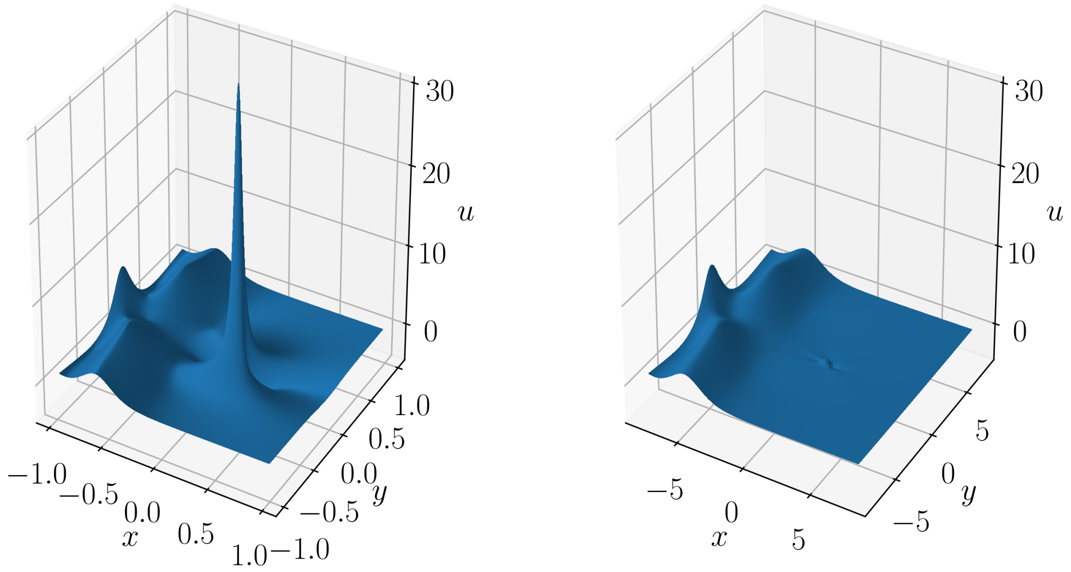


Figure 5.3: Surface plot of the numerical solution to the KPI Equation for $\epsilon = 0.10$ (left), and the same surface after subtracting the expression (5.1) for the lump, fitted to the numerical data (right).

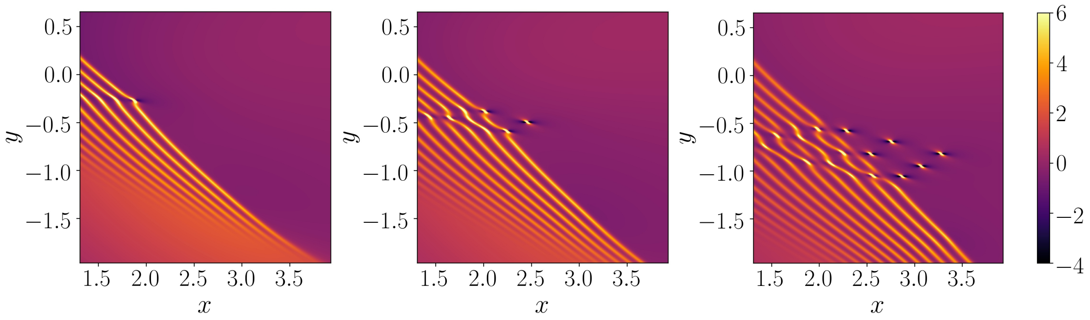


Figure 5.4: Formation of a lattice of lumps from a numerical computation with $\epsilon = 0.02$ and initial datum as of Equation (5.5). The three figures correspond to $t = 0.61$ (left), $t = 0.71$ (center), and $t = 0.86$ (right).

TABLE 5.1: Relevant parameters for the numerical computations for a fixed dispersion ($\epsilon = 0.1$), and varying initial data (C_0). In the table, "n" denotes a numeric identifier for each run, C_0 is the constant in Equation (5.6) defining the initial datum, h is the fixed time step chosen, and "grid" denotes the number of grid points (or, equivalently, Fourier modes) in each direction.

n	ϵ	C_0	h	grid
1	0.10	4	$2 \cdot 10^{-4}$	$2^{13} \times 2^{13}$
2	0.10	5	$1 \cdot 10^{-4}$	$2^{13} \times 2^{13}$
3	0.10	6	$1 \cdot 10^{-4}$	$2^{13} \times 2^{13}$
4	0.10	7	$1 \cdot 10^{-4}$	$2^{14} \times 2^{14}$
5	0.10	8	$1 \cdot 10^{-4}$	$2^{14} \times 2^{14}$

plots from the third to the sixth show that the second lump, moving across the void region ahead, interacts with the first wavefront in a way that leads to the formation of two lumps, that emerge from the first wavefront. In the same plots, it is possible to note a third wave focusing leading to the formation of a new, central lump from the third wavefront. From the fifth to the seventh plot, a similar interaction between the third lump and the second wavefront takes place, leading to the formation of a couple of symmetric solitons in the region between the first and the second wavefront. These two solitons (plots from seventh to twelfth) at later times interact with the two void regions left in the first wavefront by the previous couple of lumps, and the result of this interaction is the formation of three lumps emerging from a third wave breaking undergone by the first wavefront.

The process succintly described here continues at later times, and as a result of the lump-wave interaction, it is possible to identify three regions in the positive part of the solution, namely:

- a triangular lattice of two-dimensional peaks moving in the x-positive direction in a solitonic fashion;
- a specular region of continuous lump-wave interaction, whose extension and shape matches that of the solitonic region, and keeps growing in time;
- a region of unperturbed dispersive waves, propagating in both space directions, and whose amplitude decreases as the waves spread.

The behaviour described in this section resembles quite closely that of the KdV Equation, where it is well known [9] that a generic initial datum will evolve in solitons and radiation in the long-time asymptotic.

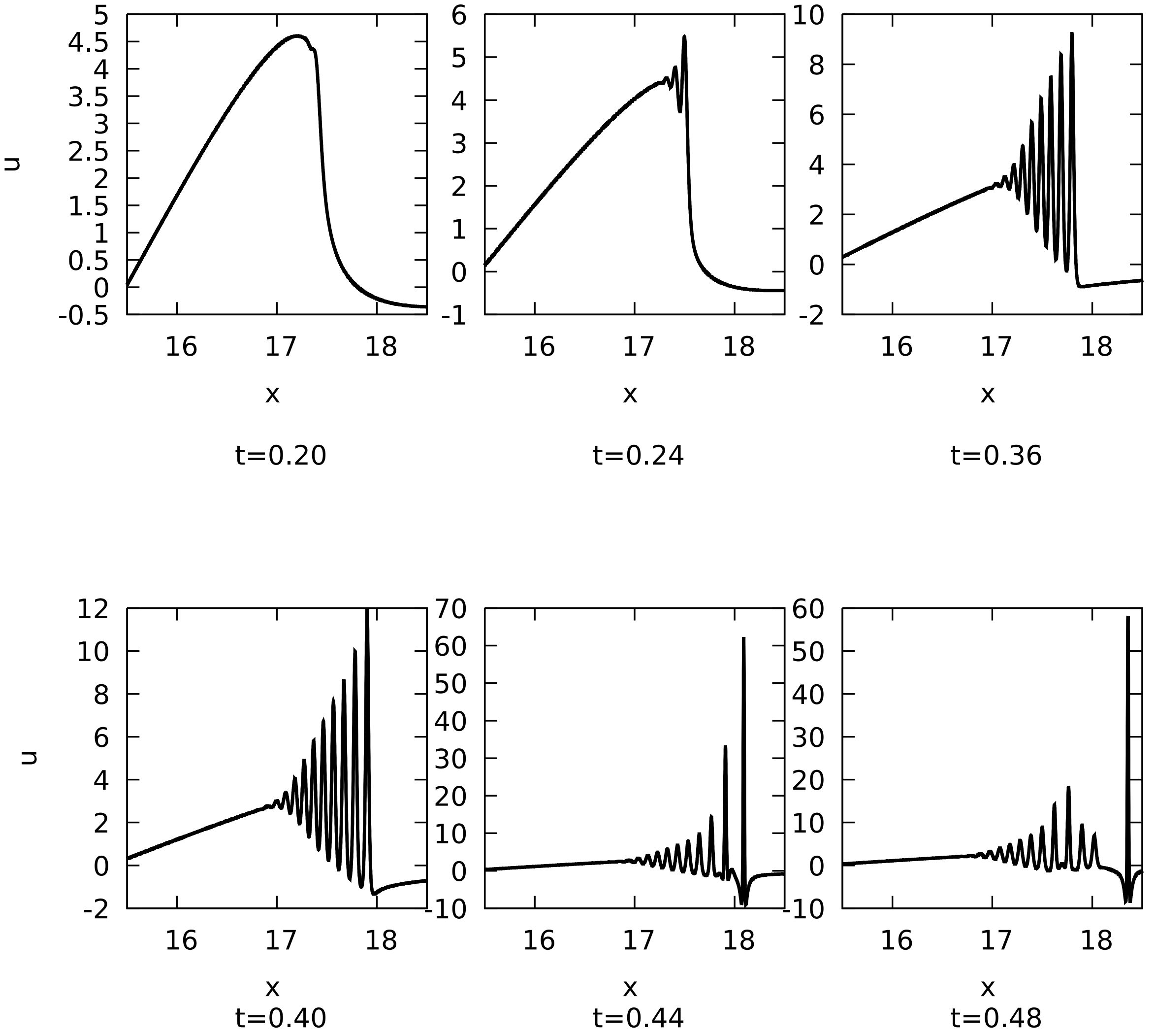
5.2 Lump scaling for fixed ϵ

We now proceed by studying how the height (and thus the speed) of the lumps changes with the energy of the initial datum, for a fixed value of ϵ. This test is performed by selecting an initial datum with expression:

$$u_0(x, y) = -C_0 \partial_x \operatorname{sech}^2 \sqrt{x^2 + y^2}, \tag{5.6}$$

with several values of C_0 as shown in Table 5.1, and $\epsilon = 0.10$. The initial datum consists of a positive and a negative bump (see Figure 1.1 (left)), disposed so that these will move away from each other during the time evolution. Furthermore, since the initial datum is a derivative, it fulfills the first of the infinitely many constraints discussed in Section 1.3.4.

Clearly, the height of the initial datum depends on C_0, and as C_0 is increased, in can be expected that the shock in the related dispersionless KP Equation becomes stronger, leading to the formation of higher amplitude waves and to higher peaks in the KP Equation after wave focusing instability. The results of this test are shown in Figure 5.7, where the lump appearance corresponds to the sharp rise in the curves on the left plot. The linear fit on



t=0.20 t=0.24 t=0.36

t=0.40 t=0.44 t=0.48

FIGURE 5.5: Slice along the line $y = 0$ of the numerical solution for $\epsilon = 0.02$, taken at a few representative times (note the different vertical scale between the plots). The shock waves are moving from left to right in the pictures. The steepening of the initial bump can be seen at $t = 0.20$, and already at $t = 0.24$ there is a dispersive regularisation for the shock, represented by the onset of a few small amplitude oscillations. At $t = 0.36$ and $t = 0.40$, the dispersive waves are propagating and the energy content of the waves is increasing, as testified by the growing wave amplitude. Some focusing effects can be seen at $t = 0.40$, as the leading wavefront departs from the line connecting the crests of the other dispersive waves. At $t = 0.44$ a lump has completely formed, as shown by the very high and localised peak, followed by another lump that is forming right ahead of the wave train. The second lump has completely formed at $t = 0.48$, and the effect of lump-wave interaction can be inferred from the fact that the wave packet that has a much less regular structure than it had for previous times.

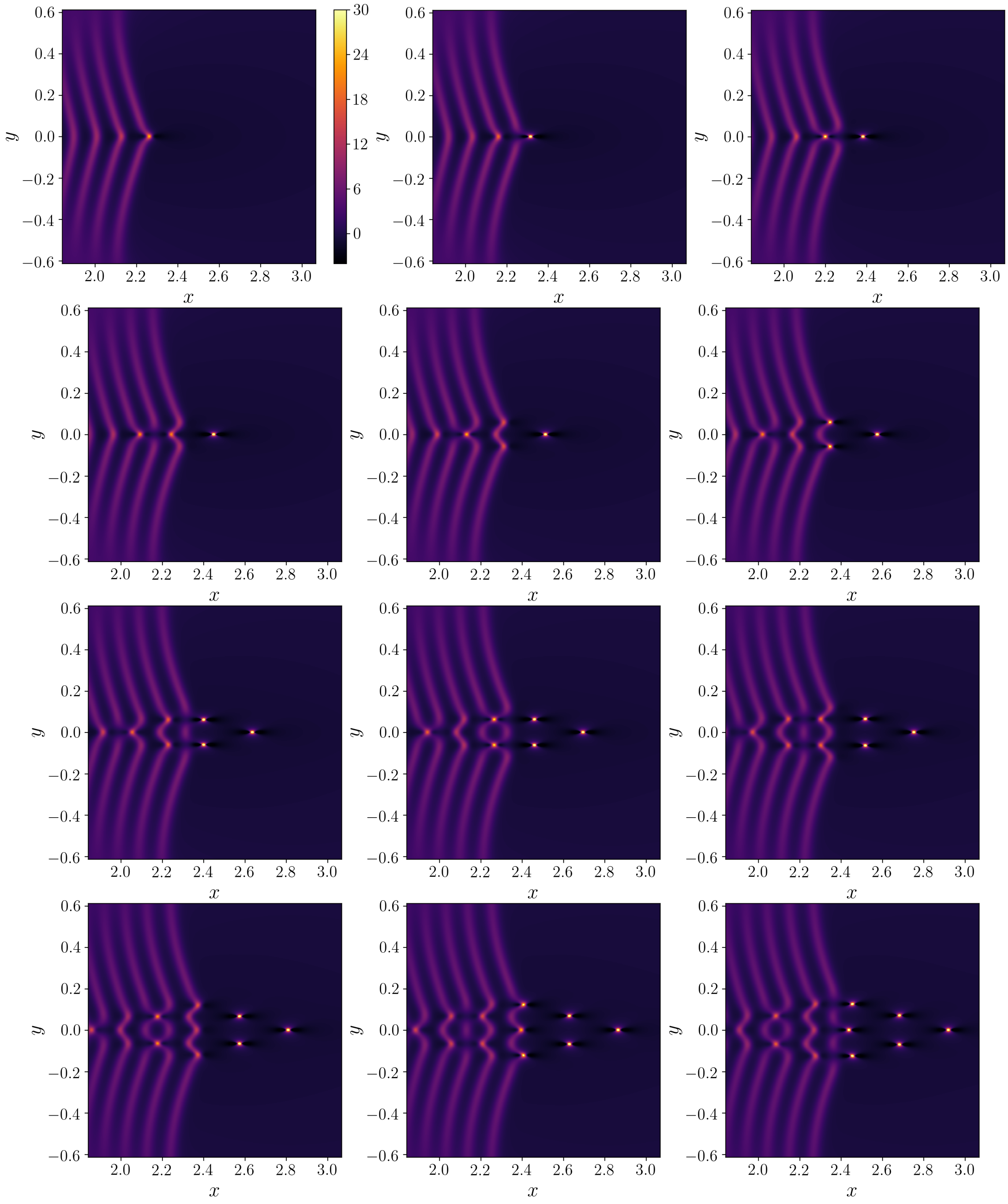


FIGURE 5.6: Detail of the lump formation mechanism for a numerical computation with $\epsilon = 0.02$ and $C_0 = 6$ (case n. 14 of Table 5.2). The upper left image corresponds to $t = 0.42$, and the others follow from left to right, from top to bottom with a time separation of $\Delta t = 0.01$. For the meaning of the colors, see the colorbar on the upper left plot.

Figure 5.7 (right) confirms that the maximum height of the lump is linearly correlated with the magnitude of the initial datum, and that the time required by the lump to achieve its maximum value is inversely proportional to C_0. A notable numerical result is that after the lump develops, its height is not constant in time, but is decreasing initially at a fast rate, later on at a much slower pace. This is in contrast with the exact solution (5.1), that describes a lump of constant height, and is the main observation that led us to the development of Whitham's modulation equations for the lump in Section 2.4.

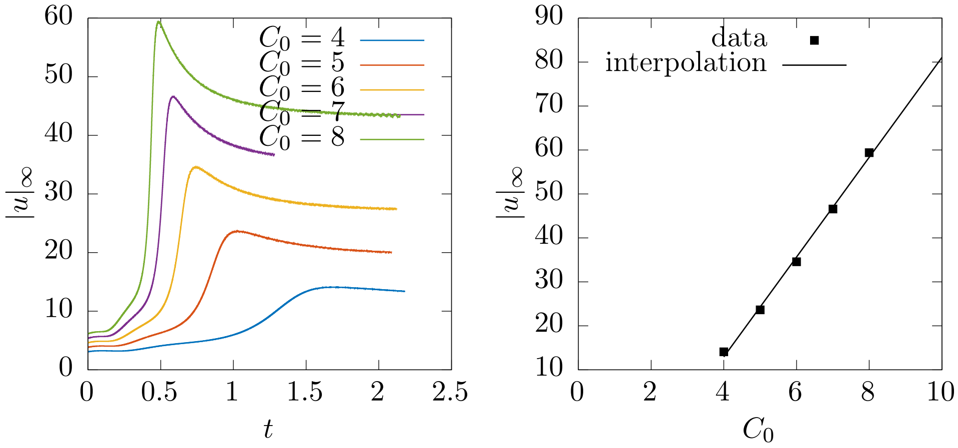


FIGURE 5.7: On the left, time evolution of the L^∞ norm computed numerically for $\epsilon = 0.02$ and increasing magnitude of the initial datum, as for Table 5.1. On the right, the maximum value of the L^∞ norm for the times shown in the left plot is regressed against C_0. The linear interpolant has the expression: $|u|_\infty = \alpha + \beta C_0$ with coefficients determined by least squares interpolation: $\alpha = -29.894$, $\beta = 10.838$.

5.3 Lump scaling for fixed $|u_0|$

The next campaign of numerical computations aims at studying the complementary approach with respect to Section 5.2. In this case, the magnitude of the initial data is kept constant (with the same initial data of Equation (5.6), but a fixed value $C_0 = 6$), and several numerical computations are executed for a few relevant values of ϵ, see Table 5.2 for the details. Following the outline of the previous section, we report in Figure 5.8 the time behaviour of the L^∞ norm as computed numerically, for nine values of ϵ between 0.02 and 0.10. Comparing Figure 5.8 with Figure 5.7, it can be inferred that in this respect the dispersion coefficient has the opposite effect than the magnitude of the initial data. More precisely, the lump maximum height is inversely proportional to ϵ, while in the previous section it was shown that it is directly proportional to C_0. Conversely, the time required by the modulated lump to achieve its maximum height grows with ϵ for a fixed C_0, while it decreases with C_0 for a fixed ϵ. To get a scaling law from the numerical results, we consider the power-law scaling:

$$|u|_\infty = \alpha + \beta \epsilon^\gamma, \tag{5.7}$$

which gives an exponent $\gamma = 0.6437$.

A last scaling consideration that we make here regards the position and time at which the lump reaches its maximum height. In principle, it is possible to study the scaling of these two quantities separately, but to provide a stricter scaling rule we consider the following simultaneous fit:

$$x_{\max} - \frac{|u|_{\max}}{t_{\max}} = \alpha + \beta \epsilon^\gamma, \tag{5.8}$$

with

$$|u|_{\max} = \max_{t\in[0,T]} \|u(x,y,t)\|_{L^\infty(\mathbb{T}^2)} \qquad t_{\max} = \arg\max_{t\in[0,T]} \|u(x,y,t)\|_{L^\infty(\mathbb{T}^2)}. \tag{5.9}$$

Table 5.2: Relevant parameters of the numerical computations for a fixed initial datum ($C_0 = 6$), and varying dispersion coefficient (ϵ). In the table, “n” denotes a numeric identifier for each run, C_0 is the constant in Equation (5.6) defining the symmetric initial datum, h is the fixed time step chosen, and “grid” denotes the number of grid points (or, equivalently, Fourier modes) in each direction.

n	ϵ	C_0	h	grid
6	0.10	6	$2 \cdot 10^{-4}$	$2^{14} \times 2^{14}$
7	0.09	6	$2 \cdot 10^{-4}$	$2^{14} \times 2^{14}$
8	0.08	6	$2 \cdot 10^{-4}$	$2^{14} \times 2^{14}$
9	0.07	6	$2 \cdot 10^{-4}$	$2^{14} \times 2^{14}$
10	0.06	6	$1 \cdot 10^{-4}$	$2^{14} \times 2^{14}$
11	0.05	6	$1 \cdot 10^{-4}$	$2^{15} \times 2^{15}$
12	0.04	6	$8 \cdot 10^{-5}$	$2^{15} \times 2^{15}$
13	0.03	6	$4 \cdot 10^{-5}$	$2^{16} \times 2^{15}$
14	0.02	6	$4 \cdot 10^{-5}$	$2^{16} \times 2^{16}$

The regression for the quantity of Equation (5.8), as well as its numerical data are shown in Figure 5.9. The scaling given by the fit is $\gamma = 0.7820$, and this value will be discussed in Section 5.4.

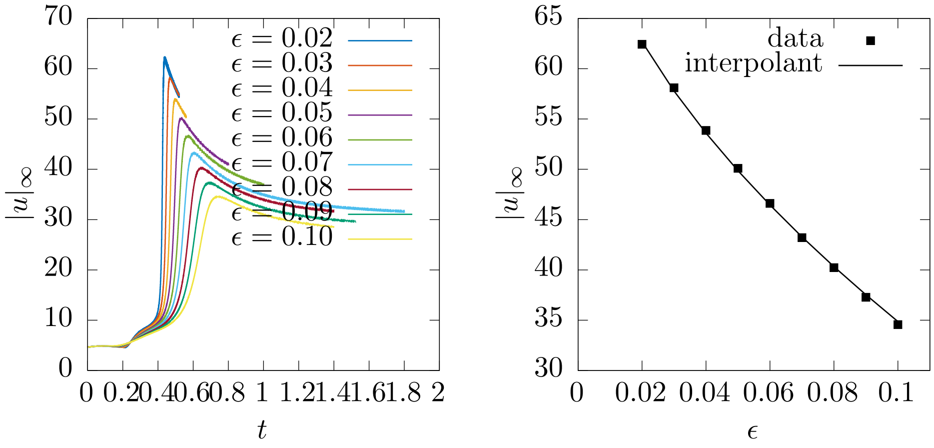


Figure 5.8: Numerical values for the L^∞ norm of the solution for a fixed value of C_0 and several values of ϵ (left), and maximum absolute value of the numerical solution computed over the time intervals considered for the numerical computations of Table 5.2 (right). The power-law fit $|u|_\infty = \alpha + \beta\epsilon^\gamma$ gives $\alpha = 77.9350$, $\beta = -181.4782$, $\gamma = 0.6437$.

5.4 Connection with focusing NLS

In this last section, we put forward an observation which relates the structures emerging from the wave breaking in the KPI Equation with a rigorous asymptotic result known for the focusing NLS Equation. The consequences of this analogy, although not proven here, may be a valid step towards the precise description of the structures and the scaling seen numerically for the KPI Equation in the small dispersion or large energy regimes. Consider the semiclassical limit of the focusing nonlinear Schrödinger (NLS) Equation [77] in one space dimension ($x \in \mathbb{R}$):

$$\mathrm{i}\epsilon\psi_t + \frac{\epsilon^2}{2}\psi_{xx} + |\psi|^2\psi = 0, \tag{5.10}$$

with initial datum:

$$\psi(x,0) = A(x)\exp\left(\frac{\mathrm{i}}{\epsilon}S(x)\right), \tag{5.11}$$

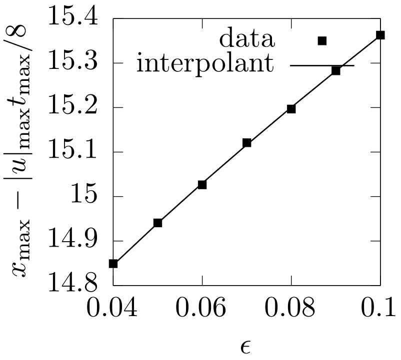


Figure 5.9: Numerical values for the quantity $x_{\max} - |u|_{\max} t_{\max}$ (see Equation (5.9) for the meaning of the symbols) and power-law fit $\alpha + \beta\epsilon^{\gamma}$, which gives the coefficients $\alpha = 14.354$, $\beta = 6.1037$, $\gamma = 0.7820$.

where $A(x)$ and $S(x)$ are real analytic functions of x, and $A(x)$ represents a localised amplitude. It is known that generically the initial datum (5.11) under the NLS flow will form a point of gradient singularity (x_0, t_0), and that in a neighbourhood of the singularity the solution will develop a region of fast oscillations both in amplitude and in phase. This behaviour is shown qualitatively in Figure 5.10, where squared modulus (left) and phase (right) of a numerical solution for the initial datum:

$$\psi(x, 0) = \operatorname{sech} x, \tag{5.12}$$

with $\epsilon = 0.02$ are plot.

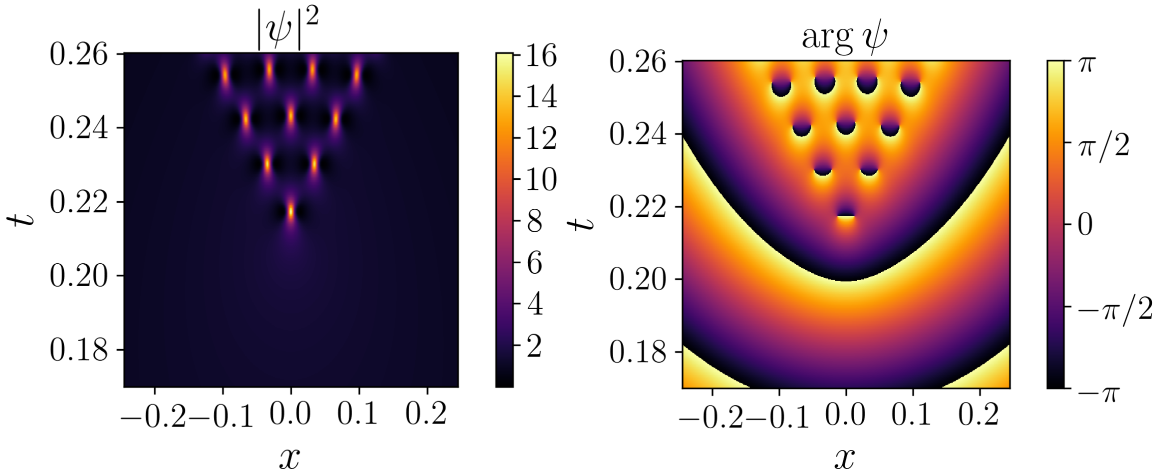


Figure 5.10: Breather formation (space-time plot) in the semiclassical limit of the NLS Equation, with the initial datum of Equation (5.12) and $\epsilon = 0.02$. The square modulus ($\rho = |\psi|^2$) is shown on the left picture, while the phase ($\arg\psi$) is shown on the right picture. The data for the plots is obtained from a numerical computation using a Fourier pseudospectral method.

Bertola and Tovbis [18] proved several results that make more precise the qualitative behaviour of Figure 5.10, among which we recall the following.

- The maximum peak is approximated by the Peregrine breather [108], defined by:

$$Q_P(X, T; a, b) = b e^{\mathrm{i}\left(aX + (a^2/2 - b^2)T\right)} \left(1 - 4\frac{1 + 2\mathrm{i}b^2 T}{1 + 4b^2(X + aT)^2 + 4b^4T^2}\right). \tag{5.13}$$

  and the asymptotic is of order $\frac{1}{5}$ in ϵ, namely:

$$\psi(x, t) = Q_P\left(\frac{x - x_p}{\epsilon}, \frac{t - t_p}{\epsilon}; a, b\right) + O(\epsilon^{\frac{1}{5}}), \tag{5.14}$$

pointwise, for appropriate values of the constants a and b that will be discussed later. This behaviour is universal, in the sense that it is independent of the exact expression of the initial datum, provided that it is in the class (5.11)

- The lattice arrangement of the peaks in the x-t plane is the same as the poles of the *tritronquée* solution of the Painlevé I Equation in the complex plane, which is defined as the unique solution $y(z)$ to:

  $$y''(z) = 6y(z)^2 + z \tag{5.15}$$

  subject to the asymptotic condition:

  $$y(z) = \sqrt{\frac{-6}{z}} \tag{5.16}$$

  as $z \to -\infty$.

The visual similarity of the NLS breathers in Figure 5.10 in the x-t plane and of the KPI lumps in Figure 5.6 in the x-y plane is striking, and seems to suggest some connection between the small dispersion asymptotics of these two equations.

It is convenient to introduce the Madelung variables:

$$\rho = |\psi|^2 \qquad\qquad w = \frac{\epsilon}{2\mathrm{i}}\left(\frac{\psi_x}{\psi} - \frac{\overline{\psi}_x}{\overline{\psi}}\right), \tag{5.17}$$

where ρ can be interpreted as a probability density, to rewrite the NLS Equation (5.10) in the hydrodynamic form:

$$\rho_t + (\rho w)_x = 0 \tag{5.18}$$

$$w_t - \rho_x + w w_x + \frac{\epsilon^2}{4}\left(\frac{1}{2}\frac{\rho_x^2}{\rho^2} - \frac{\rho_{xx}}{\rho}\right)_x = 0. \tag{5.19}$$

Equation (5.19) can be expressed as a nonlinear system of conservation laws:

$$\begin{pmatrix}\rho\\ w\end{pmatrix}_t + \begin{bmatrix} w & \rho\\ -1 & w\end{bmatrix}\begin{pmatrix}\rho\\ w\end{pmatrix}_x = \begin{pmatrix} 0\\ \frac{\epsilon^2}{4}\left(\frac{\rho_{xx}}{\rho} - \frac{1}{2}\frac{\rho_x^2}{\rho^2}\right)_x\end{pmatrix}. \tag{5.20}$$

It can be checked that the eigenvalues of the system (5.20) for $\epsilon = 0$ are:

$$\lambda_{1,2} = w \pm \mathrm{i}\sqrt{\rho}, \tag{5.21}$$

and that in general these values are complex numbers with nonzero imaginary part, so that the system of conservation laws (5.20) with $\epsilon = 0$ is elliptic, and generically an elliptic umbilic singularity takes place at a point (x_0, t_0). In [18], the authors prove that in the semiclassical limit, such a caustic singularity does indeed form at the point of gradient catastrophe, and that the semiclassical equation is a good approximation (to order $O(\epsilon)$ of the semiclassical case as long as the gradients do not grow too much (namely, for $t < t_0$). The effect of a nonzero dispersion is to keep the solution bounded; as a result, there is an obstruction to the formation of singularities, leading instead to the formation of high but bounded peaks, approximately described by Peregrine breathers. In the dispersive case, the peak of the first breather forms slightly ahead of the singularity in the semiclassical limit, and the distance between the lump and the singularity scales as $\epsilon^{4/5}$. The modulus of the first peak is $9\rho(x_0, t_0)$, very close to the numerical value of 10 found in our numerical computations for the KPI Equation. Bertola and Tovbis also prove that close to the position of the tritronqueé solution of the Painlevé I Equation, the cubic NLS Equation has the following asymptotic description:

$$\psi(x,t) = Q_P\left(\frac{x - x_0}{\epsilon}, \frac{t - t_0}{\epsilon}; -w(x_0, t_0), \sqrt{\rho(x_0, t_0)}\right) + O(\epsilon^{1/5}), \tag{5.22}$$

where Q_P is the *Peregrine breather*, defined in Equation (5.13). Computing the squared modulus of the Peregrine breather gives:

$$|Q_P(X,T;a,b)|^2 = b^2 - 8\frac{-4(x+aT)^2+4b^2T^2+1/b^2}{\left(4(X+aT)^2+4b^2T^2+1/b^2\right)^2}, \tag{5.23}$$

and by means of the following change of variables:

$$X = x - 3b^2t \qquad\qquad T = 2\sqrt{3}y \tag{5.24}$$

after a rescaling, the squared modulus of the Peregrine breather (5.23) gives:

$$3b^2 - 12\left|Q_P\left(x-3b^2t, 2\sqrt{3}y; 0, \frac{b}{2}\right)\right|^2 = 24\frac{-(x-3b^2t)^2+3b^2y^2+\frac{1}{b^2}}{\left((x-3b^2t)^2+3b^2y^2+\frac{1}{b^2}\right)^2}, \tag{5.25}$$

which is the symmetric[1] lump solution of the KPI Equation.

This fact, jointly with the numerically verified scaling law $\simeq \epsilon^{4/5}$, provides some justification for the similar phenomenon of peak formation in the NLS Equation and lump formation in the KPI Equation. This connection between NLS and KPI can be further developed to predict the position of the first KPI lump. In [18] it is shown that the breathers of the NLS Equation emerge at (x_b, t_b) given by the following relation:

$$z_p = \frac{c}{\epsilon^{4/5}}\left[x_b - x_0 + \left(-w(x_0,t_0) + \mathrm{i}\sqrt{\rho(x_0,t_0)}\right)(t_b - t_0)\right], \tag{5.26}$$

where z_p is a pole for the tritronqueé solution of the Painlevé I Equation, and a, b, c are constants. In particular, the first peak corresponds to the first pole on the negative real axis of the tritronqueé solution. By the change of variables (5.24), we obtain the following expression for the position (x_l, y_l, t_l) for the emergence of the first lump:

$$z_p = \frac{c}{\epsilon^{4/5}}\left[x_l - 3b^2t_l - (x_0 - 3b^2t_0)\right], \tag{5.27}$$

where (x_0, y_0, t_0) are the coordinates where the Whitham system develops the first singularity. Equation (5.27) allows to conclude that the quantity:

$$x_l = 3b^2t_l \tag{5.28}$$

scales as $\epsilon^{4/5}$. This scaling is indeed confirmed by the numerical computations in Section 5.3, see in particular Figure 5.9.

---

[1] The full lump solution, Equation (1.102) can be obtained with a slightly different change of variables:

$$3b^2 - 12\left|Q_P\left(x-(a^2+3b^2)t, 2\sqrt{3}(t+2at); \frac{a}{2\sqrt{3}}, \frac{b}{2}\right)\right|^2$$

Bibliography

[1] S. Abenda, T. Grava, and C. Klein. "Numerical Solution of the Small Dispersion Limit of the Camassa–Holm and Whitham Equations and Multiscale Expansions". In: *SIAM Journal on Applied Mathematics* 70.8 (2010).

[2] S. Abenda, T. Grava, and C. Klein. "On Whitham Equations for Camassa–Holm". In: *Waves and Stability in Continuous Media*. World Scientific, 2012, pp. 1–6. DOI: 10.1142/9789812773616_0001.

[3] M. J. Ablowitz and J. Satsuma. "Solitons and rational solutions of nonlinear evolution equations". In: *Journal of Mathematical Physics* 19.10 (1978), pp. 2180–2186. DOI: 10.1063/1.523550.

[4] M. Ablowitz and P. Clarkson. *Solitons, Nonlinear Evolution Equations and Inverse Scattering*. Cambridge University Press, 1991.

[5] M. J. Ablowitz. *Nonlinear Dispersive Waves: Asymptotic Analysis and Solitons*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2011. DOI: 10.1017/CBO9780511998324.

[6] M. J. Ablowitz, G. Biondini, and Q. Wang. "Whitham modulation theory for the Kadomtsev–Petviashvili equation". In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 473.2204 (2017). DOI: 10.1098/rspa.2016.0695.

[7] M. J. Ablowitz, A. Demirci, and Y.-P. Ma. "Dispersive shock waves in the Kadomtsev–Petviashvili and two dimensional Benjamin–Ono equations". In: *Physica D: Nonlinear Phenomena* 333 (2016). Dispersive Hydrodynamics, pp. 84–98. DOI: https://doi.org/10.1016/j.physd.2016.01.013.

[8] M. J. Ablowitz and V. Javier. "On the Kadomtsev–Petviashvili Equation and Associated Constraints". In: *Studies in Applied Mathematics* 85.3 (1991), pp. 195–213. DOI: 10.1002/sapm1991853195.

[9] M. J. Ablowitz and H. Segur. *Solitons and the Inverse Scattering Transform*. SIAM, 1981.

[10] M. J. Ablowitz, G. Biondini, and I. Rumanov. "Whitham modulation theory for (2+1)-dimensional equations of Kadomtsev–Petviashvili type". In: *Journal of Physics A: Mathematical and Theoretical* 51.21 (2018).

[11] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions*. Dover Publications, 1965.

[12] L. Adhianto et al. "HPCTOOLKIT: tools for performance analysis of optimized parallel programs". In: *Concurrency and Computation: Practice and Experience* 22.6 (2010), pp. 685–701. DOI: 10.1002/cpe.1553.

[13] J. Alexander, R. Pego, and R. Sachs. "On the transverse instability of solitary waves in the Kadomtsev–Petviashvili equation". In: *Physics Letters A* 226.3 (1997), pp. 187–192. DOI: https://doi.org/10.1016/S0375-9601(96)00921-8.

[14] V. Arnold. *Mathematical Methods of Classical Mechanics*. Springer, 1989.

[15] U. M. Ascher, S. J. Ruuth, and B. T. Wetton. "Implicit-Explicit Methods for Time-Dependent Partial Differential Equations". In: *SIAM Journal on Numerical Analysis* 32 (3 1995).

[16] T. B. Benjamin. "Internal waves of permanent form in fluids of great depth". In: *Journal of Fluid Mechanics* 29.3 (1967), pp. 559–592. DOI: 10.1017/S002211206700103X.

[17] T. Benjamin, J. Bona, and J. Mahony. “Model equations for long waves in nonlinear dispersive systems”. In: *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 272.1220 (1972), pp. 47–78. DOI: `10.1098/rsta.1972.0032`.

[18] M. Bertola and A. Tovbis. “Universality for the Focusing Nonlinear Schrödinger Equation at the Gradient Catastrophe Point: Rational Breathers and Poles of the Tritronquée Solution to Painlevé I”. In: *Communications on Pure and Applied Mathematics* 66.5 (2013), pp. 678–752. DOI: `10.1002/cpa.21445`.

[19] M. Blaszak. *Multi-Hamiltonian Theory of Dynamical Systems*. Springer, 1998.

[20] M. Boiti, F. Pempinelli, and A. Pogrebkov. “The KPI equation with unconstrained initial data”. In: *Acta Applicandae Mathematica* 39.1 (1995), pp. 175–192. DOI: `10.1007/BF00994633`.

[21] A. E. Borzì. “Fourier Spectral Methods for Soliton Equations”. PhD thesis. SISSA, 1990.

[22] J. Bourgain. “On the Cauchy Problem for the Kadomstev–Petviashvili Equation.” In: *Geometric and functional analysis* 3.4 (1993), pp. 315–341.

[23] J. Boyd. *Chebyshev and Fourier Spectral Methods: Second Revised Edition*. Dover Publications, 2001.

[24] J. Butcher. “A history of Runge-Kutta methods”. In: *Applied Numerical Mathematics* 20.3 (1996), pp. 247–260. DOI: `https://doi.org/10.1016/0168-9274(95)00108-5`.

[25] M. Calvo, J. de Frutos, and J. Novo. “Linearly implicit Runge–Kutta methods for advection–reaction–diffusion equations”. In: *Applied Numerical Mathematics* 37.4 (2001), pp. 535–549. DOI: `https://doi.org/10.1016/S0168-9274(00)00061-1`.

[26] R. Camassa and D. D. Holm. “An integrable shallow water equation with peaked solitons”. In: *Physical Review Letters* 71 (11 1993), pp. 1661–1664. DOI: `10.1103/PhysRevLett.71.1661`.

[27] C. Canuto et al. *Spectral Methods: Fundamentals in Single Domains*. Springer Berlin Heidelberg, 2007.

[28] S. Chakravarty, T. McDowell, and M. Osborne. “Numerical studies of the KP line-solitons”. In: *Communications in Nonlinear Science and Numerical Simulation* 44 (2017), pp. 37–51. DOI: `https://doi.org/10.1016/j.cnsns.2016.07.026`.

[29] J.-P. Chehab, P. Garnier, and Y. Mammeri. “Numerical solution of the generalized Kadomtsev-Petviashvili equations with compact finite difference schemes”. In: *ArXiv e-prints* (2016). arXiv: `1605.03213`.

[30] A. J. Christlieb et al. “Parallel high-order integrators”. In: *SIAM Journal on Scientific Computing* 32.2 (2010), pp. 818–835.

[31] A. Christlieb, B. Ong, and J.-M. Qiu. “Integral deferred correction methods constructed with high order Runge-Kutta integrators”. In: *Math. Comp.* 79.270 (2010), pp. 761–783. DOI: `10.1090/S0025-5718-09-02276-5`.

[32] A. Christlieb et al. “Semi-implicit integral deferred correction constructed with additive Runge-Kutta methods”. In: *Commun. Math. Sci.* 9.3 (2011), pp. 879–902. DOI: `10.4310/CMS.2011.v9.n3.a10`.

[33] P. A. Clarkson and S. Hood. “Nonclassical symmetry reductions and exact solutions of the Zabolotskaya–Khokhlov equation”. In: *European Journal of Applied Mathematics* 3.4 (1992), pp. 381–414. DOI: `10.1017/S0956792500000929`.

[34] S. Cox and P. Matthews. “Exponential Time Differencing for Stiff Systems”. In: *Journal of Computational Physics* 176.2 (2002), pp. 430–455. DOI: `https://doi.org/10.1006/jcph.2002.6995`.

[35] J. J. Dongarra and A. J. van der Steen. “High-performance computing systems: Status and outlook”. In: *Acta Numerica* 21 (2012), pp. 379–474. DOI: `10.1017/S0962492912000050`.

[36] P. Drazin and R. Johnson. *Solitons: an introduction*. Cambridge University Press, 1989.

[37] T. A. Driscoll. "A Composite Runge–Kutta Method for the Spectral Solution of Semilinear PDEs". In: *Journal of Computational Physics* 182.2 (2002), pp. 357–367. DOI: `https://doi.org/10.1006/jcph.2002.7127`.

[38] V. Dryuma. "Analytic solution of the two-dimensional Korteweg–de Vries (KdV) equation". In: *JETP Letters* 19 (1974).

[39] A. Dubey. *PFFT : A Parallel Portable Library for Computing FFT*. URL: `http://flash.uchicago.edu/~dubey/pfft/` (visited on 05/25/2018).

[40] B. Dubrovin. "Theta functions and non-linear equations". In: *Russian Mathematical Surveys* 36.2 (1981), p. 11.

[41] B. Dubrovin, T. Grava, and C. Klein. "Numerical study of breakup in generalized Korteweg–De Vries and Kawahara equations". In: *SIAM Journal on Applied Mathematics* 71.4 (2011), pp. 983–1008.

[42] B. Dubrovin, T. Grava, and C. Klein. "On critical behaviour in generalized Kadomtsev–Petviashvili equations". In: *Physica D: Nonlinear Phenomena* 333 (2016). Dispersive Hydrodynamics, pp. 157–170. DOI: `https://doi.org/10.1016/j.physd.2016.01.011`.

[43] B. Dubrovin. *Lecture notes on Integrable Systems and Riemann Surfaces*. Apr. 2009.

[44] A. Dutt, L. Greengard, and V. Rokhlin. "Spectral deferred correction methods for ordinary differential equations". In: *BIT* 40.2 (2000), pp. 241–266. DOI: `10.1023/A:1022338906936`.

[45] L. Einkemmer and A. Ostermann. "A split step Fourier/discontinuous Galerkin scheme for the Kadomtsev–Petviashvili equation". In: *Applied Mathematics and Computation* 334 (2018), pp. 311–325. DOI: `https://doi.org/10.1016/j.amc.2018.04.013`.

[46] G. El and M. Hoefer. "Dispersive shock waves and modulation theory". In: *Physica D: Nonlinear Phenomena* 333 (2016). Dispersive Hydrodynamics, pp. 11–65. DOI: `https://doi.org/10.1016/j.physd.2016.04.006`.

[47] M. Emmett and M. Minion. "Toward an efficient parallel in time method for partial differential equations". In: *Communications in Applied Mathematics and Computational Science* 7.1 (2012).

[48] E. Ferapontov and K. Khusnutdinova. "The Haantjes tensor and double waves for multi-dimensional systems of hydrodynamic type: a necessary condition for integrability". In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 462.2068 (2006), pp. 1197–1219. DOI: `10.1098/rspa.2005.1627`.

[49] H. Flaschka, M. G. Forest, and D. W. McLaughlin. "Multiphase averaging and the inverse spectral solution of the Korteweg–de Vries equation". In: *Communications on Pure and Applied Mathematics* 33.6 (1980), pp. 739–784. DOI: `10.1002/cpa.3160330605`.

[50] M. G. Forest and J.-E. Lee. "Geometry and Modulation Theory for the Periodic Nonlinear Schrodinger Equation". In: *Oscillation Theory, Computation, and Methods of Compensated Compactness*. Ed. by C. Dafermos et al. Springer New York, 1986, pp. 35–69.

[51] B. Fornberg and T. A. Driscoll. "A Fast Spectral Algorithm for Nonlinear Wave Equations with Linear Dispersion". In: *Journal of Computational Physics* 155.2 (1999), pp. 456–467. DOI: `https://doi.org/10.1006/jcph.1999.6351`.

[52] J. Frauendiener and C. Klein. "Hyperelliptic Theta-Functions and Spectral Methods: KdV and KP Solutions". In: *Letters in Mathematical Physics* 76.2 (2006), pp. 249–267. DOI: `10.1007/s11005-006-0068-4`.

[53] M. Frigo and S. Johnson. "The Design and Implementation of FFTW3". In: *Proceedings of the IEEE* 93.2 (2005). Special issue on "Program Generation, Optimization, and Platform Adaptation", pp. 216–231.

[54] L. Gatto and P. Salehyan. *Hasse-Schmidt Derivations on Grassmann Algebras*. Springer, 2016.

[55] G. Golub and C. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 2013.

[56] D. Gottlieb and S. Orszag. *Numerical Analysis of Spectral Methods: Theory and Applications*. Society for Industrial and Applied Mathematics, 1977.

[57] T. Grava, C. Klein, and G. Pitton. "Numerical study of the Kadomtsev–Petviashvili equation and dispersive shock waves". In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 474.2210 (2018). DOI: `10.1098/rspa.2017.0458`.

[58] T. Grava, C. Klein, and J. Eggers. "Shock formation in the dispersionless Kadomtsev–Petviashvili equation". In: *Nonlinearity* 29.4 (2016), p. 1384.

[59] T. Grava. "Whitham Modulation Equations and Application to Small Dispersion Asymptotics and Long Time Asymptotics of Nonlinear Dispersive Equations". In: *Rogue and Shock Waves in Nonlinear Dispersive Media*. Ed. by M. Onorato, S. Resitori, and F. Baronio. Springer International Publishing, 2016, pp. 309–335. DOI: `10.1007/978-3-319-39214-1_10`.

[60] E. Hairer, S. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I: Nonstiff Problems*. Springer Berlin Heidelberg, 2008.

[61] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*. Springer Berlin Heidelberg, 2010.

[62] E. Hairer and G. Wanner. "Order Stars and Stability Domains". In: *Encyclopedia of Applied and Computational Mathematics*. Ed. by B. Engquist. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, pp. 1108–1112. DOI: `10.1007/978-3-540-70529-1_134`.

[63] R. Hirota. "Exact Solution of the Korteweg—de Vries Equation for Multiple Collisions of Solitons". In: *Physical Review Letters* 27 (18 1971), pp. 1192–1194.

[64] M. Hochbruck and A. Ostermann. "Exponential integrators". In: *Acta Numerica* 19 (2010), pp. 209–286. DOI: `10.1017/S0962492910000048`.

[65] D. D. Holm, T. Schmah, and C. Stoica. *Geometric Mechanics and Symmetry*. Oxford University Press, 2009.

[66] T. Y. Hou. "Blow-up or no blow-up? A unified computational and analytic approach to 3D incompressible Euler and Navier–Stokes equations". In: *Acta Numerica* 18 (2009), pp. 277–346. DOI: `10.1017/S0962492906420018`.

[67] *How to square two complex doubles with 256-bit AVX vectors*. URL: `https://stackoverflow.com/questions/39509746/how-to-square-two-complex-doubles-with-256-bit-avx-vectors` (visited on 06/06/2018).

[68] E. Infeld, G. Rowlands, and A. Senatorski. "Instabilities and oscillations of one–and two–dimensional Kadomtsev—Petviashvili waves and solitons". In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 455.1992 (1999), pp. 4363–4381. DOI: `10.1098/rspa.1999.0505`.

[69] E. Infeld, A. Senatorski, and A. A. Skorupski. "Decay of Kadomtsev-Petviashvili solitons". In: *Physical Review Letters* 72 (9 1994), pp. 1345–1347. DOI: `10.1103/PhysRevLett.72.1345`.

[70] E. Infeld, A. Senatorski, and A. A. Skorupski. "Numerical simulations of Kadomtsev–Petviashvili soliton interactions". In: *Physical Review E* 51 (4 1995), pp. 3183–3191. DOI: `10.1103/PhysRevE.51.3183`.

[71] *Intrinsics for Intel Advanced Vector Extensions*. URL: `https://software.intel.com/en-us/node/524038` (visited on 06/06/2018).

[72] A. Ionescu, C. Kenig, and D. Tataru. "Global well-posedness of the KP-I initial-value problem in the energy space". In: *Inventiones mathematicae* 173.2 (2008), pp. 265–304. DOI: `10.1007/s00222-008-0115-0`.

[73] J. V. José and E. J. Saletan. *Classical Dynamics*. Cambridge University Press, 1998.

[74] C. Julien and F. Charbel. “A time-parallel implicit method for accelerating the solution of non-linear structural dynamics problems”. In: *International Journal for Numerical Methods in Engineering* 77.4 (), pp. 451–470. DOI: `10.1002/nme.2418`.

[75] B. B. Kadomtsev and V. I. Petviashvili. “On the Stability of Solitary Waves in Weakly Dispersing Media”. In: *Soviet Physics Doklady* 15 (1970), p. 539.

[76] A. M. Kamchatnov. *Nonlinear Periodic Waves and their Modulations: an Introductory Course*. World Scientific Publishing Co, 2000. DOI: `10.1142/4513`.

[77] S. Kamvissis, K. D. T.-R. McLaughlin, and P. D. Miller. *Semiclassical Soliton Ensembles for the Focusing Nonlinear Schrodinger Equation*. Princeton University Press, 2003.

[78] A. Kasman. *Glimpses of Soliton Theory*. American Mathematical Society, 2010.

[79] D. I. Ketcheson and U. b. Waheed. “A comparison of high-order explicit Runge–Kutta, extrapolation, and deferred correction methods in serial and parallel”. In: *Communications in Applied Mathematics and Computational Science* 9 (2014), p. 2.

[80] C. Klein. “Fourth Order Time-Stepping for Low Dispersion Korteweg-De Vries and Nonlinear Schrödinger Equations”. In: *Electronic Transactions on Numerical Analysis* 29 (2008), pp. 116–135.

[81] C. Klein and K. Roidot. “Fourth Order Time-Stepping for Kadomtsev–Petviashvili and Davey–Stewartson Equations”. In: *SIAM Journal on Scientific Computing* 33 (6 2011), pp. 3333–3356.

[82] C. Klein and K. Roidot. “Numerical study of shock formation in the dispersionless Kadomtsev–Petviashvili equation and dispersive regularizations”. In: *Physica D: Nonlinear Phenomena* 265 (2013), pp. 1–25. DOI: `https://doi.org/10.1016/j.physd.2013.09.005`.

[83] C. Klein and J.-C. Saut. “Numerical Study of Blow up and Stability of Solutions of Generalized Kadomtsev–Petviashvili Equations”. In: *Journal of Nonlinear Science* 22.5 (2012), pp. 763–811. DOI: `10.1007/s00332-012-9127-4`.

[84] C. Klein and R. Peter. “Numerical study of blow-up in solutions to generalized Kadomtsev-Petviashvili equations”. In: *Discrete & Continuous Dynamical Systems — B* 19 (2014), p. 1689. DOI: `10.3934/dcdsb.2014.19.1689`.

[85] C. Klein, C. Sparber, and P. Markowich. “Numerical Study of Oscillatory Regimes in the Kadomtsev–Petviashvili Equation”. In: *Journal of Nonlinear Science* 17.5 (2007), pp. 429–470. DOI: `10.1007/s00332-007-9001-y`.

[86] Y. Kodama. “KP solitons in shallow water”. In: *Journal of Physics A: Mathematical and Theoretical* 43 (2010).

[87] R. Krasny. “A study of singularity formation in a vortex sheet by the point-vortex approximation”. In: *Journal of Fluid Mechanics* 167 (1986), pp. 65–93. DOI: `10.1017/S0022112086002732`.

[88] I. M. Krichever. “Integration of nonlinear equations by the methods of algebraic geometry”. In: *Functional Analysis and Its Applications* 11.1 (1977), pp. 12–26. DOI: `10.1007/BF01135528`.

[89] I. Krichever. “Methods of Algebraic Geometry in the Theory of non-linear Equations”. In: *Russian Mathematical Surveys* 32.6 (1977), p. 185.

[90] S. Krogstad. “Generalized integrating factor methods for stiff PDEs”. In: *Journal of Computational Physics* 203.1 (2005), pp. 72–88. DOI: `https://doi.org/10.1016/j.jcp.2004.08.006`.

[91] D. Lawden. *Elliptic Functions and Applications*. Springer, 2014.

[92] N. Li and S. Laizet. “2DECOMP&FFT – A highly scalable 2D decomposition library and FFT interface”. In: *Cray User Group 2010 conference*. Edinburgh, 2010.

[93] J.-E. Lin and H. Chen. “Constraints and conserved quantities of the Kadomtsev–Petviashvili equations”. In: *Physics Letters A* 89.4 (1982), pp. 163–167. DOI: `https://doi.org/10.1016/0375-9601(82)90198-0`.

[94] J.-L. Lions, Y. Maday, and G. Turinici. “Résolution d’EDP par un schéma en temps «pararéel»”. In: *Comptes Rendus de l’Académie des Sciences - Series I - Mathematics* 332.7 (2001), pp. 661–668. DOI: `https://doi.org/10.1016/S0764-4442(00)01793-6`.

[95] V. T. Luan and A. Ostermann. “Explicit exponential Runge–Kutta methods of high order for parabolic problems”. In: *Journal of Computational and Applied Mathematics* 256 (2014), pp. 168–179. DOI: `https://doi.org/10.1016/j.cam.2013.07.027`.

[96] S. Manakov and P. Santini. “On the solutions of the dKP equation: the nonlinear Riemann Hilbert problem, longtime behaviour, implicit solutions and wave breaking”. In: *Journal of Physics A: Mathematical and Theoretical* 41.5 (2008).

[97] S. Manakov and P. Santini. “Wave breaking in solutions of the dispersionless kadomtsev-petviashvili equation at a finite time”. In: *Theoretical and Mathematical Physics* 172.2 (2012), pp. 1118–1126.

[98] S. Manakov et al. “Two-dimensional solitons of the Kadomtsev-Petviashvili equation and their interaction”. In: *Physics Letters A* 63.3 (1977), pp. 205–206. DOI: `https://doi.org/10.1016/0375-9601(77)90875-1`.

[99] G. Markus et al. “The Scalasca performance toolset architecture”. In: *Concurrency and Computation: Practice and Experience* 22.6 (), pp. 702–719. DOI: `10.1002/cpe.1556`.

[100] H. McKean and V. Moll. *Elliptic Curves: Function Theory, Geometry, Arithmetic.* Cambridge University Press, 1999.

[101] A. Minzoni and N. Smyth. “Evolution of lump solutions for the KP equation”. In: *Wave Motion* 24.3 (1996), pp. 291–305. DOI: `https://doi.org/10.1016/S0165-2125(96)00023-6`.

[102] L. Molinet, J. Saut, and N. Tzvetkov. “Global well-posedness for the KP-I equation”. In: *Mathematische Annalen* 324.2 (2002), pp. 255–275. DOI: `10.1007/s00208-002-0338-0`.

[103] M. Mulase. “Algebraic theory of the KP equations”. In: *Perspectives in mathematical physics*. Conf. Proc. Lecture Notes Math. Phys., III. Int. Press, Cambridge, MA, 1994, pp. 151–217.

[104] H. Ono. “Algebraic Solitary Waves in Stratified Fluids”. In: *Journal of the Physical Society of Japan* 39.4 (1975), pp. 1082–1091. DOI: `10.1143/JPSJ.39.1082`.

[105] M. Pavlov. “Nonlinear Schrödinger equation and the bogolyubov-whitham method of averaging”. In: *Theoret. and Math. Phys.* 71 (3 1987). DOI: `https://doi.org/10.1007/BF01017090`.

[106] D. Pekurovsky. “P3DFFT: A Framework for Parallel Computations of Fourier Transforms in Three Dimensions”. In: *SIAM Journal on Scientific Computing* 34.4 (2012).

[107] D. Pelinovsky and Y. Stepanyants. “Self-focusing instability of plane solitons and chains of two-dimensional solitons in positive-dispersion media”. In: *Journal of Experimental and Theoretical Physics* 77 (4 1993).

[108] D. H. Peregrine. “Water waves, nonlinear Schrödinger equations and their solutions”. In: *The Journal of the Australian Mathematical Society. Series B. Applied Mathematics* 25.1 (1983), pp. 16–43. DOI: `10.1017/S0334270000003891`.

[109] K. Roidot. “Numerical study of non-linear dispersive partial differential equations”. PhD thesis. Université de Bourgogne, 2011.

[110] F. Rousset and N. Tzvetkov. “Transverse nonlinear instability for two-dimensional dispersive models”. In: *Annales de l’Institut Henri Poincare (C) Non Linear Analysis* 26.2 (2009), pp. 477–496. DOI: `https://doi.org/10.1016/j.anihpc.2007.09.006`.

[111] F. Santucci and P. Santini. “On the dispersionless Kadomtsev–Petviashvili equation with arbitrary nonlinearity and dimensionality: exact solutions, longtime asymptotics of the Cauchy problem, wave breaking and shocks”. In: *Journal of Physics A: Mathematical and Theoretical* 49 (2016).

[112] J. Satsuma and M. J. Ablowitz. “Two-dimensional lumps in nonlinear dispersive systems”. In: *Journal of Mathematical Physics* 20.7 (1979), pp. 1496–1503. DOI: `10.1063/1.524208`.

[113] J. Shen, T. Tang, and L. Wang. *Spectral Methods: Algorithms, Analysis and Applications*. Springer Berlin Heidelberg, 2011.

[114] S. Shende and A. Malony. “The Tau Parallel Performance System”. In: *International Journal of High Performance Computing Application* 20.2 (2006), pp. 287–311.

[115] P. Sionóid and A. Cates. “The generalized Burgers and Zabolotskaya–Khokhlov equations”. In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 447.1930 (1994), pp. 253–270. DOI: `10.1098/rspa.1994.0139`.

[116] G. Whitham. *Linear and Nonlinear Waves*. Wiley, 1974.

[117] G. Whitham. “Non-linear dispersive waves”. In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 283.1393 (1965), pp. 238–261. DOI: `10.1098/rspa.1965.0019`.

[118] N. J. Zabusky and M. D. Kruskal. “Interaction of “Solitons” in a Collisionless Plasma and the Recurrence of Initial States”. In: *Physical Review Letters* 15 (6 1965), pp. 240–243. DOI: `10.1103/PhysRevLett.15.240`.

[119] A. A. Zaitsev. “Formation of stationary nonlinear waves by superposition of solitons”. In: *Soviet Physics Doklady* 28 (1983), p. 720.

[120] V. Zakharov. “Instability and nonlinear oscillations of solitons”. In: *Journal of Experimental and Theoretical Physics Letters* 72 (1975).

[121] V. Zakharov and E. Kuznetsov. “Three-dimensional solitons”. In: *Soviet Physics-JETP* 39.2 (1974).

[122] V. E. Zakharov and E. A. Kuznetsov. “Solitons and collapses: two evolution scenarios of nonlinear wave systems”. In: *Physics-Uspekhi* 55.6 (2012), p. 535.