# Waves and their modulations

Sylvie Benzoni-Gavage

June 26, 2020

**Abstract**

The words of the title, waves and modulations, are ubiquitous in science and technology. They are in particular the purpose of an active field of research in applied mathematics. Modulation 'theory' is fifty years old and nevertheless conceals many open questions. An overview of it is given along a journey in the landscape of a few famous equations in mathematical physics.

## 1 Oscillations

Our life is bound to many periodic phenomena, starting with the rhythm of our heart beat to the rotation on itself and the orbiting of the Earth around the Sun.

Mathematically speaking, periodic signals may have all sorts of form but the most basic ones are represented by (co)sine functions. For instance, one may consider an oscillating signal represented as a function of time $t$ by $y = \cos(\omega_0 t)$, where $\omega_0$ is a parameter determining how often the signal goes back to its original value. More precisely, the *period* of this signal is $T_0 = 2\pi/\omega_0$, and the inverse of $T_0$ is called the *frequency* of the signal. In terms of music, the larger the frequency, the more high-pitched the note.

For instance a concert A produced by a tuning fork has a - fundamental - frequency of 440 Hz, meaning that it - almost - corresponds to a sinusoidal signal of period $1/440 \simeq 0.0023$ seconds. However, a purely sinusoidal signal of 440 Hz or any other frequency is not at all pleasant to hear. Musical instruments produce much richer sounds, involving in particular 'modulations' that make for example bells ring different from guitars.

Before going to the actual meaning of the word 'modulation' in science, let us mention that it has been used in technology for decades, modulation being used to encode information in telecommunications. The word appears in particular, more or less hidden, on standard radio sets. The F that can be seen in the abbreviation FM on a radio set is indeed for frequency, the M meaning *modulation*. We can also see AM on a radio set, which means *amplitude modulation* and is slightly easier to formulate than frequency modulation.

Performing amplitude modulation of a reference signal amounts to considering a modified signal of the form $y = a(t)\cos(\omega_0 t)$, where the *amplitude* $a = a(t)$ varies over large time scales compared to the period of the reference signal. For instance we may consider $a = \cos(\omega t)$ with $\omega$ being much smaller than $\omega_0$. An example is plotted on Figure 1.

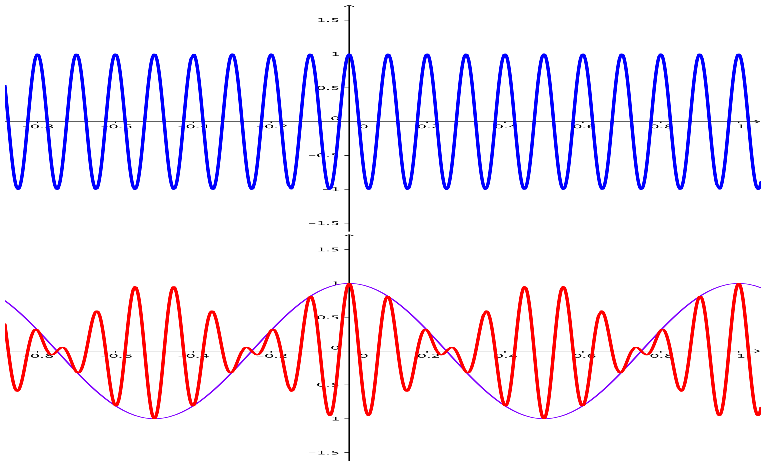

Figure 1: Plot of $y = \cos(20\pi t)$ (top), $y = \cos(2\pi t)$ (purple), and $y = \cos(2\pi t)\cos(20\pi t)$ (red).

Frequency modulation of a reference signal $y = \cos(\omega_0 t)$ amounts to modifying its linear *phase* $\omega_0 t$ into a phase function $\varphi = \varphi(t)$ in such a way that $\varphi'(t)$ oscillates around $\omega_0$ over large time scales compared to $T_0 = 2\pi/\omega_0$. For instance we may consider $\varphi = \omega_0 t + h\cos(\omega t)$ with $\omega_0$ being a multiple of $\omega$ so as to keep a periodic signal, namely of period $T = 2\pi/\omega$, which is then a multiple of $T_0$, and with $h$ a parameter tuning the height of frequency oscillations. An example is plotted on Figure 2.

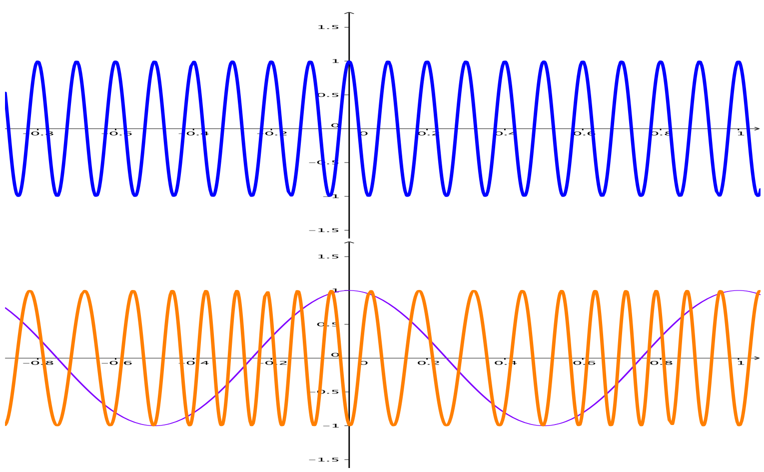

Figure 2: Plot of $y = \cos(20\pi t)$ (top), $y = \cos(2\pi t)$ (purple), and $y = \cos(20\pi t + 3\cos(2\pi t))$ (orange).

## 2 Linear waves

Let us now consider a function of both time $t$ and spatial position $x$ of the form $u = \cos(k_0 x - \omega_0 t)$, assuming for the time being that both $k_0$ and $\omega_0$ are positive, tuning parameters. Depending on whether $u$ is viewed as a function of $t$ or $x$, it has a time period $T_0 = 2\pi/\omega_0$ as before, but also another period in space called the *wavelength*, $X_0 = 2\pi/k_0$.

If we plot $u$ as a function of $x$ at each time $t$ we receive a movie of a sinusoidal signal propagating at velocity $c_0 = \omega_0/k_0$, meaning that at some time $t$ we get a signal that is just the initial one shifted by a distance $c_0 t$. This may represent for instance the pressure wave carrying - at a velocity of about 340 m per second in the air - a pure musical note of frequency $1/T_0$.

Such a *wave* is a particular solution of the partial differential equation (PDE)

$$u_t + c_0 u_x = 0, \tag{1}$$

called a *transport equation*, in which the subscripts stand for partial derivatives - as will be the case everywhere $t$ or $x$ appear in a subscript. It also solves the *wave equation*

$$u_{tt} - c_0^2 u_{xx} = 0,$$

which admits in addition solutions that propagate in the opposite direction - at the same speed $c_0$.

There are many other PDEs admitting solutions that are oscillatory traveling waves. In particular, one may consider the Airy equation[1]

$$u_t + \alpha u_{xxx} = 0, \tag{2}$$

which admits solutions of the form $u = \cos(kx - \omega t)$ provided that the *wave number* $k$ is linked to $\omega$ through the *dispersion relation*

$$\omega + \alpha k^3 = 0. \tag{3}$$

Here $\alpha$ is just a parameter introduced for the sake of physical consistency. Whether it is positive or negative determines the sign of $\omega$, and thus that of the *phase speed* $c = \omega/k$. This speed depends on $k$ and for this reason we say that the Airy equation (2) is *dispersive*, since waves of distinct wave numbers propagate at different speeds and thus disperse from each other as time goes on.

Equations (1) and (2) have in common that they are *linear*, and as a consequence any two solutions of them may be added to yield a third solution. If we stick to solutions of the form $u = \cos(kx - \omega t)$, and take two of them, $u_1 = \cos(k_1 x - \omega_1 t)$ and $u_2 = \cos(k_2 x - \omega_2 t)$, we get a third one by taking their mean value

$$u = \frac{1}{2}\left(\cos(k_1 x - \omega_1 t) + \cos(k_2 x - \omega_2 t)\right),$$

[1] Named after George Biddell Airy [1801–1892].

By trigonometry this equivalently reads

$$u = \cos(kx - \omega t)\cos(k_0 x - \omega_0 t)$$

with $k = (k_1 - k_2)/2$ (assuming for instance that $k_1 > k_2$ to ensure a positive wave number $k$), $\omega = (\omega_1 - \omega_2)/2$, $k_0 = (k_1 + k_2)/2$, $\omega_0 = (\omega_1 + \omega_2)/2$. If $k$ is much smaller than $k_0$, at each time $t$ the resulting function of $x$ is a modulation of the *carrier wave* $\cos(k_0 x - \omega_0 t)$ with amplitude $\cos(kx - \omega t)$. Note that $k$ happens to be much smaller than $k_0$ when the wave numbers $k_1$ and $k_2$ are chosen close enough to each other.

There is a notable difference between solutions to (1) and solutions to (2). Indeed, in the former all waves propagate at speed $c_0$, so that $\omega/k = \omega_0/k_0$. Both the carrier wave and the amplitude and therefore also the modulated wave propagate at speed $c_0$, whereas for solutions to the Airy equation (2) the two speeds $\omega/k = -\alpha k^2$ and $\omega_0/k_0 = -\alpha k_0^2$ are distinct as soon as $k_1 \neq k_2$.

Would we watch the movie of a modulated wave solution to (2) we could see that the higher frequency oscillations of the carrier wave $\cos(k_0 x - \omega_0 t)$ move inside the envelope determined by the amplitude $\cos(kx - \omega t)$.

Moreover, let us point out that the speed of propagation of the amplitude

$$\omega/k = (\omega_1 - \omega_2)/(k_1 - k_2) = -\alpha(k_1^3 - k_2^3)/(k_1 - k_2)$$

goes to $-3\alpha k_0^2$ when $k_1$ and $k_2$ both approach $k_0$. This value $-3\alpha k_0^2$ is obtained by differentiating $\omega$ with respect to $k$ at $k_0$ in the dispersion relation (3). It is called the *group velocity* of the carrier wave, which is here the triple of its phase speed $\omega_0/k_0 = -\alpha k_0^2$. The fact that the phase speed differs from the group velocity, which is equivalent to the fact that the phase speed is not constant as a function of the wave number, is a manifestation of the already mentioned dispersive feature of (2).

By using the elementary solutions described above and Fourier analysis[2], one can find a representation formula for all solutions of linear equations such as (1) and (2). Alternatively, the resolution of (1) actually reduces to merely observing, without any Fourier analysis, that $u(x,t) = u(x - c_0 t, 0)$ for all $x$ and $t$ and any solution $u$ of (1). The resolution of (2) is more delicate, and involves an intriguing function named as the equation after Airy. Dwelling on this resolution is not the purpose of this paper though.

Rather, we aim at considering some *nonlinear* PDEs, which are not amenable to Fourier analysis, and that still admit periodic waves propagating at constant speed.

# 3 Nonlinear periodic waves

Generally speaking, any function of the form $u = U(x - ct)$ is called a *traveling wave*. At any time $t$, the graph of $u(\cdot, t)$ is the one of $U$ shifted by $ct$. The function $U$ is called the *profile* of the wave.

[2]Fourier analysis is a domain of mathematics with numerous applications nowadays that was funded by the 19th century scientist Joseph Fourier [1768–1830].

The perhaps most well-known nonlinear PDE that admits periodic traveling wave solutions is the *Korteweg–de Vries* equation[3]

$$u_t + uu_x + \alpha u_{xxx} = 0. \tag{4}$$

This is a model equation for long water waves, which can be viewed as a 'mixture' of the Airy equation (2) and of a nonlinear transport equation

$$u_t + uu_x = 0. \tag{5}$$

In the latter, the speed of propagation is given by the unknown $u$ itself, by contrast with the constant velocity $c$ in (1). The periodic traveling wave solutions to (4) involve Jacobi elliptic functions, which we shall not use here. Some of them are called cnoidal waves, displaying sharper crests and flatter troughs than in a sine wave (see Figure 3 for an artist view).

Figure 3: Cnoidal wave profile - plot of elliptic cosine $\mathrm{cn}(\cdot, 0.9)$. CC BY-SA 3.0 Kraaiennest.

Another well-known nonlinear PDE with similar features and even more elementary periodic traveling wave solutions is the *nonlinear Schrödinger* equation[4]

$$z_t = i\alpha z_{xx} - i\gamma z|z|^2, \tag{6}$$

in which the unknown $z$ takes complex values, and $\alpha$, $\gamma$ are real parameters. This model is used in various fields of mathematical physics, ranging from quantum mechanics to nonlinear optics and water waves on deep water.

We can see that for any positive numbers $r$ and $k$, the function $z = re^{i(kx-\omega t)}$ is a solution of (6) if and only if

$$\omega = \alpha k^2 + \gamma r^2. \tag{7}$$

This relation between the frequency $\omega$, the wave number $k$ and the amplitude $r$ is called a *nonlinear dispersion relation*. Solutions of the form $z = re^{i(kx-\omega t)}$ are called *harmonic waves*, their real and imaginary parts being as the (co)sine waves discussed in Section 2. They are not the only periodic traveling wave solutions to (6). This nonlinear PDE also admits more complicated periodic traveling wave solutions - namely, cnoidal waves again - that will not be described here.

Let us give a last example of a nonlinear PDE admitting periodic traveling wave solutions. This one can be viewed as a 'damped'[5] modification of (6). It is called the *complex Ginzburg–Landau* equation[6] and reads

$$z_t = (1 + i\alpha)z_{xx} + z - (1 + i\gamma)z|z|^2. \tag{8}$$

[3]Named after Diederik Johannes Korteweg [1848–1941] and Gustav de Vries [1866–1934]. For more information see [9].

[4]Named after Erwin Schrödinger [1887–1961].

[5]This word is not essential here and we shall not try to explain it.

[6]Named after Vitaly Ginzburg [1916–2009] and Lev Landau [1908–1968], the complex Ginzburg–Landau equation is the main example considered in the seminal work [11] on the topic we address here.

This equation arises for instance in the modeling of wakes past obstacles in viscous fluids flows. We can see that $z = r\mathrm{e}^{i(kx-\omega t)}$ is a solution of (8) if and only if

$$r^2 = 1 - k^2, \quad \omega = \gamma + (\alpha - \gamma)k^2. \tag{9}$$

We thus still find harmonic waves, but with only one *degree of freedom*, the wave number $k$ determining both $\omega$ and $r$ (up to a $\pm$ sign). By comparison, there are two degrees of freedom for harmonic wave solutions to (6), namely, the wave number $k$ and the modulus $r$, which are both free and determine $\omega$ through (7).

Because of nonlinearity, we cannot add together harmonic waves to build modulated wave solutions to those PDEs. Yet it is natural to ask whether they admit *modulated wave trains*, say of the form $z = r(x,t)\mathrm{e}^{i\varphi(x,t)}$ where the amplitude $r$, the *local wave number* $\varphi_x$ and the *local frequency* $\varphi_t$ vary significantly only over large scales. This turns out to be a tough question in general.

More precisely, this topic involves the following series of natural questions, starting from the most basic one up to the really tough one.

1. For a given nonlinear PDE, what are the equations governing the large scale variations of the amplitude and the phase of modulated wave train solutions, at least in an approximate manner?

2. Do these equations, referred to as *modulated equations*, actually have solutions on sufficiently large scales, in particular for a long enough time?

3. Once we have a positive answer to the previous question, to what extent does it give information on actual solutions of the original PDE? In other words, how can we build actual modulated wave train solutions from this approach?

The case when modulated equations have long-term solutions is the good one as regards mathematical analysis of wave trains. From an applied mathematics point of view, it can be more interesting to address the following.

4. What happens when the modulated equations fail to have long-term solutions?

Answering question 1 requires little creativity. As a matter of fact the derivation of modulated equations, which are expected to provide an asymptotic model for large scale variations, is just a matter of calculus, as we exemplify in the next two sections.

The way Question 2 can be dealt with depends on the form of modulated equations. In the simplest cases modulated equations can turn out to be classical - systems of - PDEs, as in the examples given in Section 4, so that it is not difficult to guess under which conditions(s) they have long-term solutions.

However, as explained in Section 5, modulated equations in general take the form of averaged equations involving mean values of periodic wave profiles and nonlinear functions of these profiles. Since the wave profiles are not known 'explicitly' in general - and at best known through rather complicated special functions for a few special PDEs - , determining whether modulated equations have long-term solutions is already an issue.

Question 3 has been mostly unanswered for fifty years, except for a couple of special PDEs, while Question 4 has been partly answered. Section 6 is devoted to giving the flavor of a rule of thumb about Question 4 that became a mathematical result in various frameworks in the last two decades.

To finish with this overview, let us stress that besides a rather complete theory for the Korteweg–de Vries equation and a more limited one for the nonlinear Schrödinger equation, there have been recent breakthroughs regarding Question 3 for some classes of 'dissipative' PDEs. Section 7 gives a glimpse of the known results and of the remaining issues in general.

# 4 Some explicit modulated equations

The study of modulated wave trains was initiated in the mid 1960s by Whitham [31], especially for dispersive PDEs such as (4). More precisely, Gerald Whitham [1927–2014] pointed out several methods to derive *modulated equations* that should govern, at least approximately, the evolution of modulated wave trains, should they exist. One of those methods relies on a possible underlying Lagrangian for the reference PDE, and is thus by nature not applicable to dissipative equations such as (8). Another method is what he called the 'two-timing' method. This one is more systematic and is actually based on a formal, *multi-scale expansion*. The machinery for deriving modulated equations in this way is a little bit tedious but not complicated. Let us exemplify it with two simple cases.

## 4.1 Basic example

Let us start by giving its flavor for (8) - we refer to [11] for more details. We recall that harmonic wave solutions to (8) are of the form

$$z = \sqrt{1-k^2}\,\mathrm{e}^{i(kx-(\gamma+(\alpha-\gamma)k^2)t)},$$

according to the nonlinear dispersion relation (9).

For simplicity - as in [11, § 3.3] - we concentrate on modulations of harmonic waves with $k=0$. Let us start by noticing that there are actually harmonic wave solutions to (8) with arbitrarily small wave numbers $k=\varepsilon K$, namely

$$z_\varepsilon = \sqrt{1-\varepsilon^2K^2}\,\mathrm{e}^{i(-\gamma t+\varepsilon Kx-(\alpha-\gamma)\varepsilon^2K^2t)} \tag{10}$$

for any fixed $K$ and an arbitrarily small positive parameter $\varepsilon$ . The first reason why we introduce $K$ here (instead of just considering the case $K=1$) is because we think of $\varepsilon$ as being nondimensional (that is, with no unit of measurement contrary to $k$ and $K$, which are homogeneous the inverse of a length).

By construction, (10) defines a family of solutions to (8) parametrized by $\varepsilon$ that are 'slow' modulations over 'large' spatial scales of the limiting one $z_0=\mathrm{e}^{-i\gamma t}$. Indeed, for any positive $\varepsilon$ the function $z_\varepsilon$ in (10) is equal to $z_0$ multiplied by a harmonic wave of wavelength $2\pi/(K\varepsilon)$ and time period $2\pi/((\alpha-\gamma)K^2\varepsilon^2)$, both going to infinity when $\varepsilon$ goes to zero.

A deeper reason why we have introduced $K$ is that we want to look for more general modulations of $z_0$ in terms of the *rescaled variables* $\chi = \varepsilon x$ and $\tau = \varepsilon^2 t$, in which $K$ will be varying as a *local* wave number with respect to $\chi$ and $\tau$. More precisely, we seek solutions to (8) of the form

$$z_\varepsilon = (1 + \varepsilon^2 R_\varepsilon(\varepsilon x, \varepsilon^2 t))\, \mathrm{e}^{i(-\gamma t + \Phi_\varepsilon(\varepsilon x, \varepsilon^2 t))} \tag{11}$$

We already know special solutions of this form, namely those given in (10), which correspond to $\Phi_\varepsilon(\chi, \tau) = K\chi - (\alpha - \gamma)K^2\tau$ for a fixed $K$, and

$$R_\varepsilon = (\sqrt{1 - \varepsilon^2 K^2} - 1)/\varepsilon^2 = -K^2/2 + \mathcal{O}(\varepsilon^2)$$

by Taylor expansion. For those special solutions the phase function $\Phi_\varepsilon$ is linear in $(\chi, \tau)$ and actually independent of $\varepsilon$.

Let us now seek solutions to (8) of the form (11) by just assuming that $R_\varepsilon$ and $K_\varepsilon = \Phi_{\varepsilon,\chi}$ admit asymptotic expansions

$$R_\varepsilon = R + \mathcal{O}(\varepsilon), \quad K_\varepsilon = K + \mathcal{O}(\varepsilon)$$

that depend smoothly on the variables $(\chi, \tau)$. Then, by plugging (11) in (8) we find through a calculus exercise the following equations for the lower order terms $R$, $K$

$$R = -K^2/2 - \alpha K_\chi/2\,,$$
$$K_\tau + 2(\alpha - \gamma)KK_\chi = (1 + \alpha\gamma)K_{\chi\chi}, \tag{12}$$

in which the subscripts in $\tau$ and $\chi$ stand for partial derivatives. The latter is the sought *modulated equation* associated with (8). The properties of this equation depend crucially on the sign of the right-hand side coefficient $1 + \alpha\gamma$. The 'good' case is when $1 + \alpha\gamma$ is positive. Then (12) is a *parabolic* PDE called *Burgers* equation[7], a 'mixture' of the *heat equation*

$$K_\tau = (1 + \alpha\gamma)K_{\chi\chi}$$

and the nonlinear transport equation

$$K_\tau + 2(\alpha - \gamma)KK_\chi = 0.$$

The fact that the modulated equation (12) is parabolic is reminiscent of the parabolic nature of the Ginzburg-Landau equation (8). The term *parabolic* for the latter merely refers to the fact that the set of pairs $(\omega, k)$ satisfying the second relation in (9) is a parabola. The reason for classifying the heat equation as parabolic is similar, and by extension we also say the Burgers equation is parabolic.

Parabolic PDEs are known to have solutions for all positive times. So the answer to Question 2 is positive for the modulated equation (12) when $1 + \alpha\gamma$ is positive. We will come back in Section 7 to Question 3 regarding the link between solutions to (12) and actual solutions to (8).

We have seen so far an example of a PDE (8) associated with a single modulated equation (12). This is actually the simplest situation.

[7]Named after Jan Burgers [1895–1981].

## 4.2 Another example

In general, there are several modulated equations associated with a single PDE. As an example, let us derive modulated equations associated with (6).

Keeping in mind that in view of its nonlinear dispersion relation (7) Equation (6) admits harmonic wave solutions of the form

$$z = r\,\mathrm{e}^{i(kx-(\gamma r^2+\alpha k^2)t)},$$

we may again investigate modulations of $z_0 = \mathrm{e}^{-i\gamma t}$, corresponding to $k = 0$ and $r = 1$. We thus seek solutions to (6) of the form

$$z_\varepsilon = (1 + \varepsilon R_\varepsilon(\varepsilon x, \varepsilon t))\,\mathrm{e}^{i(-\gamma t+\Phi_\varepsilon(\varepsilon x,\varepsilon t))} \tag{13}$$

for a small positive parameter $\varepsilon$. The chosen scaling is again consistent with the form of the explicit solutions we already know

$$(1 + \varepsilon R)\,\mathrm{e}^{i(\varepsilon Kx-(\gamma(1+\varepsilon R)^2+\alpha\varepsilon^2K^2)t)},$$

which are of the form (13) for any $R$ and $K$, with $R_\varepsilon = R$ and

$$\Phi_\varepsilon(\chi, \tau) = K\chi - \gamma\tau(2R + \varepsilon R^2) - \alpha\varepsilon K^2\tau = K\chi - 2\gamma\tau R + \mathcal{O}(\varepsilon),$$

where we have denoted $\chi = \varepsilon x$ as before, and here $\tau = \varepsilon t$. Then, just assuming that $R_\varepsilon$ and $K_\varepsilon = \Phi_{\varepsilon,\chi}$ admit asymptotic expansions

$$R_\varepsilon = R + \mathcal{O}(\varepsilon), \quad K_\varepsilon = K + \mathcal{O}(\varepsilon)$$

that depend smoothly on the variables $(\chi, \tau)$, by plugging (13) in (6) we find (through a calculus exercise very much similar to the derivation of (12), with actually fewer terms to deal with) the following equations for the lower order terms $R$, $K$

$$R_\tau + \alpha K_\chi = 0, \quad K_\tau + 2\gamma R_\chi = 0. \tag{14}$$

The fact that we receive a system of PDEs instead of a single PDE as in (12) comes from the fact that harmonic wave solutions to (6) have one more degree of freedom than those of (8).

We can see that smooth enough solutions of System (14) are such that $K$ satisfies a single, second order PDE. Indeed, we can eliminate $R$ by differentiating both equations, and we receive

$$K_{\tau\tau} - 2\alpha\gamma K_{\chi\chi} = 0,$$

which we recognize as a wave equation provided that $\alpha\gamma$ be positive (this one is called *hyperbolic* merely because the set of solutions of its dispersion relation is a hyperbola). The case $\alpha\gamma > 0$ is thus the 'good' one, for which the modulated equations have solutions for all times.

In [13], a slightly more complicated modulated system is considered, which governs the leading order part of more general solutions of the form

$$z_\varepsilon = r_\varepsilon(\varepsilon x, \varepsilon t)\,\mathrm{e}^{i\varphi_\varepsilon(\varepsilon x,\varepsilon t)/\varepsilon}. \tag{15}$$

One of its two equations pertains to a very general class of modulated equations, sometimes called the *conservation of waves* equation - also called *Eikonal equation* in geometric optics. To explain its derivation, let us assume that $r_\varepsilon = r + \mathcal{O}(\varepsilon)$ and that the phase function $\varphi_\varepsilon$ is such that

$$\varphi_{\varepsilon,\chi} = k + \mathcal{O}(\varepsilon)\,, \quad \varphi_{\varepsilon,\tau} = -\omega + \mathcal{O}(\varepsilon).$$

We thus find that for $z_\varepsilon$ of the form in (15) to solve (6), $r$, $k$, and $\omega$ must satisfy (7) as functions of $(\chi, \tau)$. Moreover, writing the equality of cross derivatives $\varphi_{0,\chi\tau} = \varphi_{0,\tau\chi}$ we readily obtain the modulated equation

$$k_\tau + (\alpha k^2 + \gamma r^2)_\chi = 0. \tag{16}$$

The other modulated equation is obtained by plugging (15) in (6) and equating to zero the terms of order $\varepsilon$ in (6). It turns out to read

$$r_\tau + \alpha r k_\chi + 2\alpha k r_\chi = 0. \tag{17}$$

The system (16)-(17) is now *nonlinear*. However one may observe that its linearized version about $(k, r) = (0, 1)$ - obtained by keeping only first order approximations of quadratic quantities - coincides with (14). It is interesting to note that the 'good' case for (16)-(17) is still $\alpha\gamma > 0$, as for (14). This condition ensures indeed that (16)-(17) is *strictly hyperbolic* at any point $(k, r)$, not only at $(0, 1)$, as we explain below. As a matter of fact, the hyperbolicity of (16)-(17) is to be checked on its *quasilinear* form

$$\begin{pmatrix} k_\tau \\ r_\tau \end{pmatrix} + \begin{pmatrix} 2\alpha k & 2\gamma r \\ \alpha r & 2\alpha k \end{pmatrix} \begin{pmatrix} k_\chi \\ r_\chi \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{18}$$

which is by definition equivalent to the set of equations (16)-(17) as long as we stick to smooth solutions. Remarkably enough, the transport operator $\partial_\tau + 2\alpha k \partial_\chi$ appears in both rows of (18), and the speed $2\alpha k$ happens to be the *group velocity*, obtained by differentiating $\omega$ with respect to $k$ in the nonlinear dispersion relation (7). However the nondiagonal terms in the *characteristic matrix*

$$\mathbf{M} := \begin{pmatrix} 2\alpha k & 2\gamma r \\ \alpha r & 2\alpha k \end{pmatrix}$$

are also important to determine the properties of (18). The latter is said to be strictly hyperbolic if the characteristic matrix $\mathbf{M}$ has distinct real eigenvalues[8], which is the case precisely when $\alpha\gamma$ is positive. Otherwise, and more precisely if $\alpha\gamma$ were negative the initial value problem for (18) would be *ill-posed*.

Without trying to dwell on the ill-posedness issues here, let us mention that the word 'hyperbolic' is used as a generalization of what happens for the wave equation, even though the implications of hyperbolicity are less strong for nonlinear equations than for linear ones. In general, nonlinear hyperbolic systems involve indeed finite-time blow-up of solutions. Local-in-time solutions can nevertheless give valuable information, since we are talking about the rescaled time $\tau = \varepsilon t$. See Section 6 for more details.

[8]The eigenvalues $X$ of $\mathbf{M}$ are characterized by the existence of directions in which $\mathbf{M}$ acts as the multiplication by $X$. They are found as the roots of the polynomial $X^2 - 4\alpha k X + 4\alpha^2 k^2 - 2\alpha\gamma r^2$.

## 5 More modulated equations

Modulated equations are actually not restricted to the study of wave trains of the form (15). They may serve as a tool for investigating wave trains depending on a phase $\varphi_\varepsilon(\varepsilon x, \varepsilon t)$ in a more general manner than just through cosine/sine functions as in (15). The drawback of general modulated equations is that they are not as explicit as the ones described in the previous section.

Let us exemplify this on the seminal case of the Korteweg–de Vries equation (4). We seek solutions of the form

$$u_\varepsilon = U_\varepsilon(\varepsilon x, \varepsilon t; \varphi_\varepsilon(\varepsilon x, \varepsilon t)/\varepsilon) \tag{19}$$

with $\varepsilon$ a small positive parameter and $U_\varepsilon$ being periodic - but not a priori harmonic - in its last argument, denoted by $\theta$ in what follows. We still use the notation $\chi = \varepsilon x$ and $\tau = \varepsilon t$ for the rescaled variables. Without loss of generality we may assume the period in $\theta$ to be equal to one. Denoting $k_\varepsilon = \varphi_{\varepsilon,\chi}$ and $\omega_\varepsilon = -\varphi_{\varepsilon,\tau}$, by the chain rule we have $\partial_t = \partial_\tau - \omega_\varepsilon \partial_\theta$, $\partial_x = \partial_\chi + k_\varepsilon \partial_\theta$. So, assuming that

$$k_\varepsilon = k(\chi, \tau) + \mathcal{O}(\varepsilon)\,,\ \omega_\varepsilon = \omega(\chi, \tau) + \mathcal{O}(\varepsilon),$$

$$U_\varepsilon = U(\chi, \tau; \theta) + \varepsilon V(\chi, \tau; \theta) + \mathcal{O}(\varepsilon^2),$$

we find that for $U_\varepsilon$ to solve (4) we must have

$$-\omega U_\theta + kUU_\theta + \alpha k^3 U_{\theta\theta\theta} = 0,$$

where the subscripts in $\theta$ stand once more for partial derivatives. This relation is obtained by equating to zero the leading order term in powers of $\varepsilon$ when the ansatz (19) is plugged into (4). By straightforward integration it implies the existence of $\lambda = \lambda(\chi, \tau)$ such that

$$-\,\omega U + \tfrac{1}{2} k U^2 + \alpha k^3 U_{\theta\theta} = \lambda\, k. \tag{20}$$

(The factor $k$ is introduced here above in the constant of integration for convenience, it will soon be factored out.) This second order differential equation admits $U_\theta$ as an integrating factor, which means that by multiplying (20) by $U_\theta$ we can integrate it at once. This implies that for any solution to (20) there must exist $\mu = \mu(\chi, \tau)$ such that

$$\tfrac{1}{2}\alpha k^2 U_\theta^2 = \mu + \lambda U + \tfrac{1}{2} c U^2 - \tfrac{1}{6} U^3. \tag{21}$$

with $c = \omega/k$. We readily see from the integrated form (21) of (20) that its solutions describe curves in the *phase plane* $\left\{(U, \dot{U} = kU_\theta)\right\}$ of equation

$$\tfrac{1}{2}\alpha \dot{U}^2 + \tfrac{1}{6} U^3 - \tfrac{1}{2} c U^2 - \lambda U = \mu. \tag{22}$$

We find thus find families of curves parametrized by $\mu$ at fixed $(c, \lambda)$, see Figure 4.

In particular, when the parameters $(c, \lambda)$ are such that the third order polynomial

$$\tfrac{1}{6} U^3 - \tfrac{1}{2} c U^2 - \lambda U$$

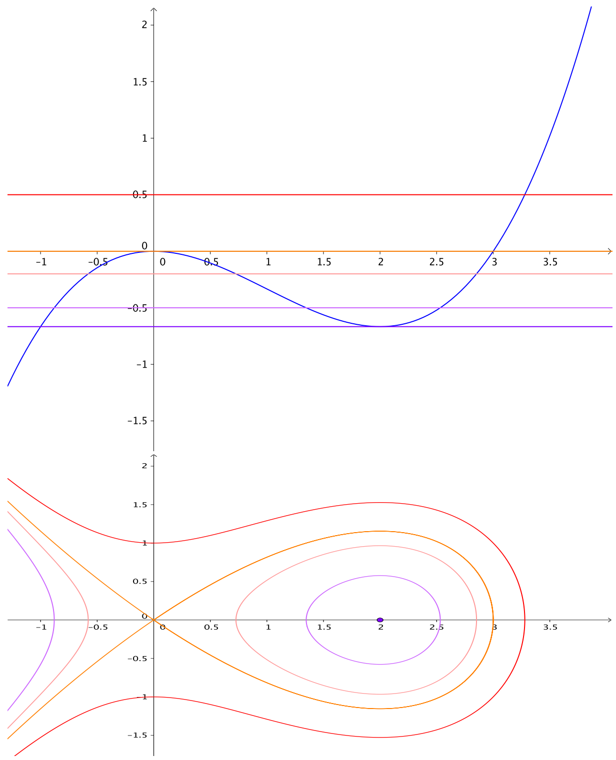

Figure 4: Top: Plot of polynomial $\frac{1}{6}U^3 - \frac{1}{2}cU^2 - \lambda U$ (in blue) and some its value levels ($\mu = -2/3, -0.5, -0.2, 0, 0.5$). Bottom: Corresponding level curves of $\frac{1}{2}\dot{U}^2 + \frac{1}{6}U^3 - \frac{1}{2}cU^2 - \lambda U$. For $c = 1$ and $\lambda = 0$.

has a strict local minimum, say $U_0 = U_0(c, \lambda)$, we find closed curves around $(U_0, 0)$ that solve (22) for values of $\mu$ greater than and sufficiently close to the minimal value of that polynomial.

These closed curves thus yield a family of periodic profiles $U$ parametrized by $(c, \lambda, \mu)$. We also see on Figure 4 a so-called *homoclinic loop* based at the point $(0, 0)$, which corresponds to a *solitary wave* profile. Solitary waves are traveling waves $u = U(x - ct)$ of a special kind, with $U$ going exponentially fast to a same constant at both $\pm\infty$. Remarkably enough the Korteweg–de Vries equation was precisely derived in the 19th century to explain the occurrence of solitary waves on the surface of water[9]. Solitary wave profiles may be viewed as a limiting case of periodic profiles when their wavelength goes to infinity.

As to periodic profiles, they are the building blocks of modulated wave trains. Since they have three degrees of freedoms here, namely the three parameters $(c, \lambda, \mu)$, we seek three modulated equations associated with the Korteweg–de Vries equation (4).

One of the modulated equations comes somehow for free. This is the conservation of waves, as in the previous section

$$k_\tau + \omega_\chi = 0,$$

[9]These were first reported on and called 'great waves of translation' by the naval engineer John Scott Russell [1808–1882].

or equivalently

$$k_\tau + (ck)_\chi = 0. \tag{23}$$

Let us now explain where the other two modulated equations come from. They are based on averaging in the variable $\theta$.

By taking the average over one period in $\theta$, we infer from (20) that

$$-c\langle U\rangle + \langle \tfrac{1}{2}U^2\rangle = \lambda, \tag{24}$$

where the brackets $\langle\cdot\rangle$ stand for the average. This suggests that we try and find modulated equations involving $\langle U\rangle$ and $\langle U^2/2\rangle$.

Using the chain rule as before and equating to zero the terms of order $\varepsilon$ when we plug (19) into (4) we obtain

$$U_\tau + UU_\chi + (\alpha k^2 U_{\theta\theta})_\chi - \omega V_\theta + k(UV + 2\alpha k U_{\theta\chi} + \alpha k^2 V_{\theta\theta})_\theta = 0.$$

Taking the average over one period we thus get the much simpler equation

$$\langle U\rangle_\tau + \langle \tfrac{1}{2}U^2\rangle_\chi = 0. \tag{25}$$

It remains to find an equation for $\langle U^2/2\rangle$. This can be done by observing that any smooth enough solution to (4) satisfies the additional conservation law[10]

$$\left(\tfrac{1}{2}u^2\right)_t + \left(\tfrac{1}{3}u^3 + \alpha u u_{xx} - \tfrac{1}{2}\alpha u_x^2\right)_x = 0. \tag{26}$$

By the same kind of computation as for the derivation of (25), this yields the averaged equation

$$\langle \tfrac{1}{2}U^2\rangle_\tau + \langle \tfrac{1}{3}U^3 + \alpha k^2 U U_{\theta\theta} - \tfrac{1}{2}\alpha k^2 U_\theta^2\rangle_\chi = 0.$$

By using (20) and (21), we can rewrite the latter in a simpler way as

$$\langle \tfrac{1}{2}U^2\rangle_\tau + \langle \tfrac{1}{2}cU^2 - \mu\rangle_\chi = 0. \tag{27}$$

The system of modulated equations for the Korteweg–de Vries equation (4) is thus made of the three equations (23), (25), and (27). It is not clear at first glance whether it is in *closed form*, since it apparently involves more than three unknowns, namely $c$, $\mu$, $k$, $\langle U\rangle$, and $\langle \tfrac{1}{2}U^2\rangle$. However, thanks to (24) we may substitute $\lambda$ for $\langle \tfrac{1}{2}U^2\rangle$ as an unknown. In addition, as long as periodic profiles $U$ are properly parametrized by $(c,\lambda,\mu)$, their mean value $\langle U\rangle$ and wave number $k$ are well defined in terms of $(c,\lambda,\mu)$. This implies that the system (23) (25) (27) is in closed form in the 'variables' $(c,\lambda,\mu)$.

Even though in closed form, the system (23) (25) (27) is not explicit enough. It is in particular not obvious to check whether it is hyperbolic, a property that is necessary to at least have local-in-time solutions for reasonably smooth initial data. It was an astonishing achievement by Whitham not only to show that it is hyperbolic but also that this system

[10]In fact, the Korteweg–de Vries has an infinity of conservation laws, as was discovered in the mid 1960s, see for instance [12, Chap. 1] for a narrative of this series of discoveries.

admits a whole set of special quantities called *Riemann invariants*. By definition, for a first order system of PDEs Riemann invariants are quantities that satisfy a nonlinear transport equation in which the speed is an eigenvalue of the characteristic matrix - arising in the system's quasilinear form. Having a whole set of Riemann invariants is equivalent to being able to write the system in diagonal form, up to a change of coordinates in the space of unknowns.

Whitham figured out more precisely that some Riemann invariants for (23) (25) (27) just read as the sums

$$v_1 + v_2,\ v_1 + v_3,\ v_2 + v_3$$

of the three roots $v_1$, $v_2$, and $v_3$ of the polynomial

$$\mu + \lambda U + \tfrac{1}{2}cU^2 - \tfrac{1}{6}U^3$$

appearing in the right-hand side of (21) (these roots are just the abscissas of the intersection points between the blue graph and the horizontal lines, corresponding to several values of $\mu$, on the upper picture on Figure 4).

It was understood in the 1980s that the existence of Riemann invariants for the modulated equations associated with the Korteweg–de Vries equation is linked to deep algebraic properties, in connection with its infinite number of conservation laws. See for instance the survey paper by Lax, Levermore and Venakides [24] for some explanation and further references.

As to the nonlinear Schrödinger equation (6), it actually admits periodic wave profiles depending on four parameters. Therefore, the most general modulated equations for (6) include four equations. It turns out that these modulated equations also admit a complete set of Riemann invariants, see e.g. [23]. This is a very special feature linked to integrability properties of (6) analogous to those of (4).

As said in [24], "Less is known about nonintegrable cases. While they sometimes share properties exhibited by the integrable cases, new phenomena arise that have yet to be completely understood." Almost a quarter of century later, this is still the case that we need to understand the nonintegrable cases.

For general dispersive equations, modulated equations do not have any reason to admit Riemann invariants. It is not even known whether they are hyperbolic. A large range of numerical experiments has shown that this depends on the nonlinearities involved, and that it seems to hardly ever be the case that for not completely integrable PDEs modulated equations are hyperbolic in the whole range of parameters [25].

# 6 Modulational instability

At this point of the exposition, some questions about the properties of modulated equations arise naturally.

Why is hyperbolicity of modulated equations so important, and what does it imply when it fails?

In fact, the failure of hyperbolicity of modulated equations has been thought of from the very beginning of the development of modulation theory as being somehow equivalent to the instability of periodic traveling wave solutions. The story is in particular recounted by Zakharov and Ostrovsky in [32].

The importance of the so-called *modulation(al) instability* and the underlying phenomena is obvious when we search the Web, with more than 120 000 results on Google, 16 000 on Google Scholar and 600 on arXiv (among which 435 posts since 2010) as of January 3rd, 2020.

The exact meaning of the term *modulation(al) instability* may vary according to authors though. To make things clear, we use in what follows the term *modulational instability* as a synonym, for a given wave, for the modulated equations to fail to be hyperbolic at the point corresponding to this wave in the set of parameters. For instance, if we go back to § 4 and periodic traveling wave solutions to the nonlinear Schrödinger equation (6) of the form

$$z = r\,\mathrm{e}^{i(kx-(\gamma r^2+\alpha k^2)t)},$$

such a wave is said to display modulational instability if the modulated system (16)-(17) is not hyperbolic at point $(k, r)$. In this case, this equivalently means that the characteristic matrix

$$\mathbf{M} = \begin{pmatrix} 2\alpha k & 2\gamma r \\ \alpha r & 2\alpha k \end{pmatrix}$$

has nonreal eigenvalues. For non zero $r$, this happens if and only if $\alpha\gamma$ is negative, irrespective of the actual values of $r$ and $k$. Thus in this special case, the modulational instability criterion does not depend on the specific point considered in the set of wave parameters. In general it does, even though convincing examples are out of the scope of the present paper - the only other system of modulated presented here being the everywhere hyperbolic system (23) (25) (27).

The most famous example of modulational instability, which both served to explain a physical phenomenon and paved the way for further theory, is known as *Benjamin–Feir instability*. It was shown indeed by Benjamin and Feir [1, 2] that for the periodic water waves known as *Stokes waves*, over water of depth $h$ modulational instability occurs when $kh > 1.363$. Roughly speaking, this means that waves of sufficiently small wavelength compared to depth are modulationally unstable. This explained why Stokes waves are difficult to reproduce in wave tanks such as the one pictured on Figure 5.

In practice, modulational instability is expected to imply the eventual breakup of the waveform into a train of pulses. More precisely, modulational instability is linked to what is called *sideband instability*. As said in the survey paper by Zakharov and Ostrovsky [32], "In its simplistic version, the effect of modulation instability is the result of interaction between a strong carrier harmonic wave at a frequency $\omega$, and small sidebands $\omega \pm \Omega$."

It is only rather recently that rigorous results confirmed this point of view, in various frameworks - starting with dissipative ones in [26, 27, 30], and going on with dispersive PDEs in [22, 7, 21].

Before saying a bit more on these results, let us exemplify sideband instability by considering harmonic wave solutions to (6)

$$z_r = r\,\mathrm{e}^{i(kx-\omega t)}.$$

Figure 5: Model testing with periodic waves at the Wave/Tow Tank of the Jere A. Chase Ocean Engineering Laboratory, University of New Hampshire.

Recall that the amplitude $r$, the wavenumber $k$ and the time frequency $\omega$ must satisfy the nonlinear dispersion relation in (7) for such a wave to solve (6). (The harmonic wave $z_r$ thus also depends on $k$ but for the sake of readability we refrain from introducing a double subscript.)

Sideband instability is actually a specific case of *linear instability*. Linear stability or instability have to do with perturbations $z_r + w$ of the reference solution $z_r$ that are approximate solutions up to $\mathcal{O}(w^2)$ terms. As regards sideband instability, it is both more convenient and more insightful to seek the perturbation with $z_r$ factorized, which amounts to looking for an approximate solution of the form $(1 + W)z_r$ for small $|W|$.

Plugging $z = (1 + W)z_r$ in (6), using (7) and dropping quadratic terms in $W$ or its conjugate $W^*$, we receive the following linear equation

$$W_t + 2\alpha k W_x = i\alpha W_{xx} - i\gamma r^2 (W + W^*). \tag{28}$$

A first interesting remark is that the group velocity $2\alpha k$ associated with $z_r$ shows up again, namely in the transport operator $\partial_t + 2\alpha k \partial_x$ in the left-hand side. We say that $z_r$ is *linearly unstable* if there are solutions to (28) that are unbounded when $t$ grows.

Let us in particular seek solutions to (28) of the form

$$W(x, t) = E(t)\, \mathrm{e}^{iK(x - 2\alpha k t)}$$

for real numbers $K$. The complex amplitude $E = U + iV$ must satisfy

$$E_t = -i\alpha K^2 E - i\gamma r^2 (E + E^*), \tag{29}$$

which is equivalent to the system of real ordinary differential equations (ODE)

$$\begin{cases} U_t = \alpha K^2 V \\ V_t = -(\alpha K^2 + 2\gamma r^2) U, \end{cases}$$

The behavior of its solutions critically depends on the sign of

$$D := \alpha K^2 (\alpha K^2 + 2\gamma r^2).$$

More precisely, if $D$ is positive, all solutions of (29) are bounded, but if $D$ is negative the system above admits exponentially growing solutions

$$\begin{pmatrix} U \\ V \end{pmatrix} = \mathrm{e}^{t\sqrt{-D}} \begin{pmatrix} \alpha K^2 \\ \sqrt{-D} \end{pmatrix}.$$

This implies that for negative $D$ Equation (28) has solutions of the form

$$W(x,t) = E_0\, \mathrm{e}^{t\sqrt{-D}}\, \mathrm{e}^{iK(x-2\alpha kt)}, \tag{30}$$

hence the linear instability of the wave $z_r$.

When $\alpha\gamma$ is positive, $D$ is positive as soon as $K$ is non zero. The interesting case regarding instability is thus $\alpha\gamma < 0$. For in this case $D < 0$ for non zero wave numbers $K$ such that

$$\alpha^2 K^2 < -2\alpha\gamma r^2.$$

In other words, if $\alpha\gamma$ is negative all non zero $K$ such that $|K| \in (0, r\sqrt{-2\gamma/\alpha})$ trigger instabilities of $z_r$. This is why we speak of *sideband instability*, the perturbed approximate solution $(1+W)z_r$ involving the perturbed wave number $k+K$ with non zero $K$ in the 'band' $(-r\sqrt{-2\gamma/\alpha}, r\sqrt{-2\gamma/\alpha})$.

Of course, the fact that $W$ as given by (30) is exponentially growing with $t$ implies that $(1+W)z_r$ cannot be a valid approximate solution to the nonlinear equation (6) for all times. To say it more simply, linear instability does not necessarily imply nonlinear instability for solutions of PDEs - unlike what happens for ODEs - but we are not going to dwell on this topic here.

What we are going to explain is the link with modulational instability. This has to do with small values of $|K|$, for which we note that the growth rate in the exponential $\sqrt{-D}$ is approximately equal to $r\sqrt{-2\alpha\gamma}|K|$. Thus for small $|K|$ the total factor of $t$ in (30) is approximately equal to the complex number $-iKC_\pm$ where $C_\pm = 2\alpha k \pm ir\sqrt{-2\alpha\gamma}$ and the $\pm$ sign is that of $K$.

Interestingly enough, the values $C_\pm$ happen to be precisely the eigenvalues of the characteristic matrix

$$\mathbf{M} = \begin{pmatrix} 2\alpha k & 2\gamma r \\ \alpha r & 2\alpha k \end{pmatrix},$$

which are complex conjugate when $\alpha\gamma$ is negative.

This is the sought link between modulational instability and sideband instability. As a matter of fact, the calculation above shows that the complex conjugate eigenvalues $C_\pm$ of $\mathbf{M}$ are associated with approximate solutions of the original equation (6) of the form

$$z_K(x,t) = r\,(1 + E_0\, \mathrm{e}^{i(Kx-\Omega t)})\mathrm{e}^{i(kx-\omega t)}$$

with $\Omega = i\sqrt{-D}$ so that

$$\Omega \sim C_+ K \;\text{ when } K \searrow 0\,, \quad \Omega \sim C_- K \;\text{ when } K \nearrow 0\,.$$

In the physics literature the two notions of modulational instability and sideband instability are most often used as synonyms. Mathematically speaking there is actually a single implication, which is not obvious at all to prove. This one says that modulational instability implies sideband instability in the same way as in the example here above. Namely, whenever the characteristic matrix of the modulated equations has complex eigenvalues $C_\pm$, with $\pm$ denoting the sign of their imaginary parts, the original equation linearized about the underlying wave has solutions involving a growth factor of the form

$$\mathrm{e}^{i(Kx-\Omega t)} \quad \text{with } \Omega \sim C_+ K \ \text{ when } K \searrow 0\,, \quad \Omega \sim C_- K \ \text{ when } K \nearrow 0\,.$$

The actual proof of this result depends on the framework, and has been achieved rather recently [7, 21, 22, 26, 27, 30].

Let us be a little more precise about the statement. For the sake of clarity, we stick to the example of Equation (6). Given a reference wave solution $z_r$, the linearized equation about $z_r$ is obtained by looking for perturbed solutions $z_r + Z$ and by dropping quadratic terms in $Z$ or its conjugate $Z^*$. This yields the equation

$$Z_t = i\alpha Z_{xx} - 2i\gamma |z_r|^2 Z - i\gamma z_r^2 Z^*. \tag{31}$$

What has been found here above is that for a harmonic wave $z_r = r\,\mathrm{e}^{i(kx-\omega t)}$, in the case $\alpha\gamma < 0$, Equation (31) has solutions of the form

$$Z(x,t) = \mathrm{e}^{i(Kx-\Omega t)}\mathrm{e}^{i(kx-\omega t)} \tag{32}$$

where $\Omega$ depends on $K$ and is of positive imaginary part when $|K| \searrow 0$. This $Z$ is just the product $W z_r$ with $W$ solution to (28). The advantage of (28) over (31) is merely that the former has constant coefficients, so that we could more easily find $W$ than directly $Z$.

Let us now reinterpret the existence of these solutions (32) to (31) in a more abstract manner, in order to appreciate the analogous results proved in the aforementioned references.

We first invoke a usual trick when dealing with traveling waves. We make a *change of frame* so that the wave of interest becomes steady. We set $y = x - ct$ - recalling that $\omega = ck$ - and consider

$$\widetilde{Z}(y,t) := Z(x,t).$$

Then

$$Z_t = \widetilde{Z}_t - c\widetilde{Z}_y, \quad Z_x = \widetilde{Z}_y,$$

so that $\widetilde{Z}$ solves the new equation

$$\widetilde{Z}_t = i\alpha \widetilde{Z}_{yy} + c\widetilde{Z}_y - 2i\gamma |\widetilde{z_r}|^2 \widetilde{Z} - i\gamma \widetilde{z_r}^2 \widetilde{Z}^* \tag{33}$$

with $\widetilde{z_r} = r\,\mathrm{e}^{iky}$. Furthermore, (32) reads in the new frame

$$\widetilde{Z}(y,t) = \mathrm{e}^{iK(y+(c-C)t)}\mathrm{e}^{iky},$$

with $C := \Omega/K$, so that (33) implies

$$iK(c-C)\widetilde{Z} = \mathcal{M}\widetilde{Z},$$

where $\mathcal{M}$ is the *differential operator* in the $y$ variable appearing in the right-hand side of (33). To be more specific, $\mathcal{M}$ is the operator associating to any smooth enough complex valued function $Y$ of $y$ the new function

$$\mathcal{M}Y := i\alpha Y_{yy} + cY_y - 2i\gamma|\widetilde{z}_r|^2 Y - i\gamma \widetilde{z}_r^{\,2} Y^*.$$

The equality deduced from (33) is true in particular at $t = 0$ so that we have

$$iK(c - C)Y = \mathcal{M}Y$$

for $Y(y) := \mathrm{e}^{i(k+K)y}$. At this stage it is tempting to say that $\lambda := iK(c - C)$ is an *eigenvalue* of $\mathcal{M}$ associated with the direction $Y$. However, the notion of eigenvalues is more subtle for differential operators than for matrices. It depends on the chosen functional setting.

Recalling that both $k$ and $K$ are real numbers, we may say indeed that $\lambda$ is an eigenvalue of $\mathcal{M}$ in the space of $\mathcal{C}^\infty$ functions that are bounded as well as all their derivatives, since $Y$ pertains to this class of functions. Moreover, recalling that $c$ is real and $KC$ is of positive imaginary part for small $|K|$, we infer that $\lambda$ is of positive real part for such $K$. The operator $\mathcal{M}$ having an eigenvalue of positive real part implies what is called *spectral instability*[11] for the wave $z_r$.

However, we would still like to say a little bit more, which will lead us back to the notion of sideband instability. For $K = 0$ the function $Y$ reduces to $y \mapsto \mathrm{e}^{iky}$, which is periodic of period $\ell := 2\pi/k$. This period $\ell$ happens to be - and this is no chance - the wavelength of the original wave $z_r$, and it is also the period of $\widetilde{z}_r$. In other words, the differential operator $\mathcal{M}$ has periodic coefficients of period $\ell$.

For general $K$ we have $Y(\ell) = \mathrm{e}^{i\ell K} = \mathrm{e}^{i\ell K} Y(0)$. Since the operator $\mathcal{M}$ has $\ell$-periodic coefficients, this implies together with the fact that $\lambda Y = \mathcal{M}Y$ that $\lambda$ belongs to the *spectrum* of $\mathcal{M}$ on the space of square integrable functions. It would lead us too far to give a precise meaning to this statement, not even speaking of proving it - the interested reader may refer for instance to [28, Chap. I] and [29, Theorem XIII.89].

Let us just mention that in this framework $K$ can be called a *Floquet exponent*. Since $\lambda$ is of positive real part for small $|K|$, we can summarize the meaning of the existence of $Y$ by saying that *the operator $\mathcal{M}$ has unstable spectrum associated with small Floquet exponents*. This can be considered as a mathematical definition for *sideband instability*.

Now we are ready to go back to the rigorous link between modulational instability and sideband instability. It has been proved indeed in the various frameworks quoted above [7, 21, 22, 26, 27, 30] that, as for our chosen simple example, modulational instability due to complex eigenvalues $C_\pm$ of the modulated equations implies sideband instability due to some unstable spectrum

$$\lambda = iK(c - C_\pm) + \mathcal{O}(K^2)$$

associated with small Floquet exponent $K$.

[11] The word 'spectral' comes from the notion of *spectrum*, which contains eigenvalues and possibly other complex numbers for differential operators, see for instance for definitions [28, Chap. I].

# 7 Modulation theory, fifty years on and beyond

Since its inception modulation theory has been investigated in quite different ways by mathematicians and mathematical physicists, for various types of PDEs.

The earliest and most complete theory is for the Korteweg–de Vries equation (4). Thanks to deep geometric properties of this equation that cannot be described here (see [16]) both Question 3 and Question 4 were answered through a series of work in the 1980s-1990s (see [24], references therein and [10] as well), in connection with the related topic of *zero dispersion limit*.

The link between modulation theory and the zero dispersion limit comes from a rescaling argument, a function $u$ being solution to (4) if and only if the rescaled function $\widetilde{u}(\chi, \tau) = u(\chi/\varepsilon, \tau/\varepsilon)$ solves the Korteweg–de Vries equation with parameter $\varepsilon^2\alpha$.

Formally, Equation (4) goes to the nonlinear transport equation (5) when $\alpha$ goes to zero. This is a singular limit though, the nature of the two PDEs (4) and (5) being very different. One of the important features of (5) is that it admits *shock wave* solutions. A shock wave is a *discontinuous* traveling wave $u = U(x - ct)$ whose profile $U$ is a *step function*. Shock waves are to be sought as *weak solutions* to (5), defined as satisfying the integral equation

$$\int_R 2u \, \mathrm{d}x - u^2 \, \mathrm{d}t = 0 \tag{34}$$

for all rectangle $R$ of the space-time plane, Equation (34) being formally obtained from (5) by Green's theorem. By considering a rectangle $R = [0, T/c] \times [0, T]$, we find that for the profile

$$U(x) = \begin{cases} u_-, & x < 0, \\ u_+, & x > 0, \end{cases}$$

to yield a weak solution to (5), $(u_-, u_+, c)$ must satisfy the *Rankine–Hugoniot* condition

$$2c = u_- + u_+.$$

A natural question is whether shock wave solutions to (5) can be achieved as limits of solutions to (4) when $\alpha$ goes to zero. The answer is no, and this can be explained through the concept of *dispersive shocks*.

Dispersive shocks can be loosely defined as unsteady patterns

- that are close to the underlying shock wave for large $|x|$,
- that are oscillatory in a region that expands proportionally to time $t$.

Dispersive shocks are thus more complicated wave trains than traveling waves. A concrete example of such waves is given by tidal bores. These are known to happen on dozens of rivers in the world and can be fun for surfers, but can also dangerous when they are too strong - as for instance on the Qiantang River in China. For the Korteweg–de Vries equation it has been shown that such wave trains do exist and that they are well approximated by the wave trains based on modulated equations constructed by Gurevich and Pitaevskii [17].

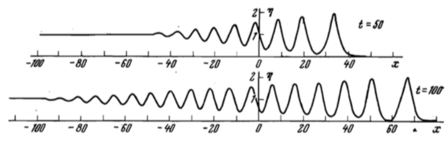


Figure 6: From [17], plots at two different times of a dispersive shock.

The oscillations in these wave trains have shorter and shorter wavelengths when $\alpha$ goes to zero, in such a way that they eventually fill a whole region within an envelope that can be computed from the modulated equations. It can be shown that the *weak limit*[12] of a dispersive shock is not a mere step function, and that is not a solution of (5). Fore more details see for instance [15, Section 3.1] and references therein.

If modulation theory is well understood for the Korteweg–de Vries equation, which has received a lot of attention in the last decades, this is not so for other types of PDEs. Even for the nonlinear Schrödinger equation, which has also been receiving a lot of attention, modulation theory has not completely been clarified yet. It is of course also linked to the zero dispersion limit ($\alpha \to 0$ in (6)), better called *semiclassical limit* in that framework. An important contribution regarding the semiclassical limit in the defocusing case ($\alpha\gamma > 0$) was brought in the 1990s by Jin, Levermore and Mc Laughlin [19], but it does not really make modulation theory a rigorous way of characterizing wave train solutions to (6). One rigorous result was obtained regarding modulation of harmonic waves - corresponding to the second example introduced in Section 4 - by Düll and Schneider in 2009 [13]. Irrespective of the sign of $\alpha\gamma$ they proved "that slow modulations in time and space of periodic wave trains of the NLS [nonlinear Schrödinger] equation can be approximated via solutions of Whithams equations associated with the wave train". Their approach is based on a change of variables and the use of the *Cauchy-Kowalewskaya theorem*[13] to circumvent the ill-posedness of modulated equations in the case $\alpha\gamma < 0$.

Remarkably enough, both the Korteweg–de Vries equation and the nonlinear Schrödinger equation themselves can be viewed as modulation equations for the *water wave equations*[14]. A rigorous justification was achieved in particular by Craig, Sulem and Sulem [8], who proved that modulations of harmonic waves as governed by the nonlinear Schrödinger equation do yield approximate solutions of the water wave equations.

A thorough study of modulated equations has been undertaken in a series of work [5, 6, 7] for a large class of nonlinear dispersive PDEs that includes the Korteweg–de Vries equation, a fluid formulation of the nonlinear Schrödinger equation, and more generally Euler–Korteweg equations, which are dispersive modifications of the usual equations for compressible fluids that are used in various fields of mathematical physics. It is in particular motivated by the study of dispersive shock waves in non integrable systems, as initiated by El in [14]. It has

[12] We say that $u$ is the weak limit of a sequence of functions $(u_n)$ if for any infinitely smooth function $\varphi$ that vanishes outside a finite interval we have $\int u_n\varphi \to \int u\varphi$.

[13] This theorem shows the existence of analytic solutions for general classes of PDEs with analytic data.

[14] The water wave equations consist of a free interface version of the Euler equations for incompressible fluids.

already turned out that the Gurevich-Pitaevskii problem is in general much more singular than expected.

Another field of research in connection with modulation theory is the stability of periodic traveling waves with respect to 'localized' perturbations. If we think of a water wave, a localized perturbation could be induced for instance by a pebble thrown into the water. As regards dispersive models, the stability of periodic traveling waves with respect to 'localized' perturbations is a widely open problem[15], even though we can find numerical evidence - by computing approximate solutions through carefully chosen algorithms - that modulation equations do play a key role in describing the perturbed solutions.

Some breakthroughs have been achieved rather recently on this topic for dissipative models. This started with the work by Doelman, Sandstede, Scheel, and Schneider [11], who addressed the case of *reaction-diffusion equations*[16], and dealt in particular with the first example described in Section 4. Then Johnson, Noble, Rodrigues, and Zumbrun [20] managed to develop a similar theory for systems of *viscous conservation laws*, which are generalized versions of the Burgers equation arising for instance in the modeling of viscous compressible fluids. Their analysis is in particular based on higher order modulated equations, like (12) for the Ginzburg-Landau equation (8). The counterpart of these higher order modulated equations for dispersive PDEs like (6) would be third order systems of PDEs instead of just first order systems like (18).

To conclude this overview, let us stress that all the aforementioned work is basically dealing with waves in one space dimension. Of course most waves in practical applications are not unidimensional, as we can see for instance on the picture of a tidal bore on Figure 7.

Figure 7: Tidal bore in Alaska.

Research is going on. In particular, a semester will be devoted to 'Dispersive hydrodynamics: mathematics, simulation and experiments, with applications in nonlinear waves' at the Isaac Newton Institute (Cambridge, UK) - originally scheduled for 2020, this thematic semester has unfortunately been postponed to 2022 due to the health crisis.

[15]By contrast, numerous results have been obtained regarding the stability of periodic traveling waves with respect to perturbations of the same period, see [3] and references therein.

[16]A reaction-diffusion equation is a 'mixture' of the ordinary differential equation $u_t = f(u)$ (the reaction part) and of the heat equation $u_t = \alpha u_{xx}$ (with $\alpha > 0$ being linked to heat conductivity), and thus reads $u_t = \alpha u_{xx} + f(u)$. Such equations are widely used in mathematical biology.

**Acknowledgment.** L. Miguel Rodrigues is warmly thanked for sharing his insight, for his careful reading of successive versions of the manuscript, and for all the great discussions we have been having on this topic.

# References

[1] T.B. Benjamin. Instability of periodic wavetrains in nonlinear dispersive systems. *Proc. Roy. Soc. A.* 299 (1967) 5975.

[2] T. B. Benjamin, J. E. Feir. The disintegration of wave trains on deep water. Part 1. Theory. *J. Fluid Mech.* 27 (1967) 417430.

[3] S. Benzoni-Gavage, C. Mietka, and L. M. Rodrigues. Co-periodic stability of periodic waves in some hamiltonian PDEs. *Nonlinearity*, 29(11):3241, 2016.

[4] S. Benzoni-Gavage, C. Mietka, and L.M. Rodrigues. Stability of periodic waves in hamiltonian PDEs of either long wavelength or small amplitude. *Indiana Univ Math J.*, To appear.

[5] S. Benzoni-Gavage, C. Mietka, and L.M. Rodrigues. Modulated equations of Hamiltonian PDEs and dispersive shocks. arXiv 1911.10067, 2019.

[6] S. Benzoni-Gavage, P. Noble, and L.M. Rodrigues. Stability of periodic waves in Hamiltonian PDEs. In *GDR Analyse des EDP*. cedram, 2013.

[7] S. Benzoni-Gavage, P. Noble, and L. M. Rodrigues. Slow modulations of periodic waves in Hamiltonian PDEs, with application to capillary fluids. *J. Nonlinear Sci.*, 24(4):711–768, 2014.

[8] W. Craig, C. Sulem, P.-L. Sulem. Nonlinear modulation of gravity waves: a rigorous approach. *Nonlinearity*, 5(2):497–522, 1992.

[9] E.M. de Jager. On the origin of the Korteweg–de Vries equation. *Forum der Berliner Mathematischen Gesellschaft*, Band 19, Dezember 2011, pp. 171-195.

[10] P. Deift, S. Venakides, and X. Zhou. New results in small dispersion KdV by an extension of the steepest descent method for Riemann-Hilbert problems. *Internat. Math. Res. Notices*, (6):286–299, 1997.

[11] A. Doelman, B. Sandstede, A. Scheel, and G. Schneider. The dynamics of modulated wave trains. *Mem. Amer. Math. Soc.*, 199(934):viii+105, 2009.

[12] P.G. Drazin and R.S. Johnson. Solitons: an introduction. Cambridge University Press, 1996.

[13] W.-P. Düll, and G. Schneider. Validity of Whithams Equations for the Modulation of Periodic Traveling Waves in the NLS Equation. J Nonlinear Sci (2009) 19: 453466.

[14] G. A. El. Resolution of a shock in hyperbolic systems modified by weak dispersion. *Chaos*, 15(3):037103, 21, 2005.

[15] G. A. El. and M. A. Hoefer. Dispersive shock waves and modulation theory. *Physica D*, 333:11–65, 2016.

[16] H. Flaschka, M. G. Forest, and D. W. McLaughlin. Multiphase averaging and the inverse spectral solution of the Korteweg-de Vries equation. *Comm. Pure Appl. Math.*, 33(6):739–784, 1980.

[17] A. V. Gurevich and L. P. Pitaevskii. Nonstationary structure of a collisionless shock wave. *Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki*, 65:590–604, August 1973.

[18] M. A. Hoefer. Shock waves in dispersive Eulerian fluids. *J. Nonlinear Sci.*, 24(3):525–577, 2014.

[19] S. Jin, C. D. Levermore, and D. W. McLaughlin. The semiclassical limit of the defocusing NLS hierarchy. *Comm. Pure Appl. Math.*, 52(5):613–654, 1999.

[20] M. A. Johnson, P. Noble, L .M. Rodrigues, and K. Zumbrun. Behavior of periodic solutions of viscous conservation laws under localized and nonlocalized perturbations. *Inventiones mathematicae*, pages 1–99, 2013.

[21] M. A. Johnson and W. R. Perkins. Modulational Instability of Viscous Fluid Conduit Periodic Waves. arXiv 1904.06348, 2019.

[22] M. A. Johnson, K. Zumbrun, and J. C. Bronski. On the modulation equations and stability of periodic generalized Korteweg-de Vries waves via Bloch decompositions. *Phys. D*, 239(23-24):2057–2065, 2010.

[23] A. M. Kamchatnov. *Nonlinear periodic waves and their modulations.* World Scientific Publishing Co., Inc., River Edge, NJ, 2000. An introductory course.

[24] P. D. Lax, C. D. Levermore, and S. Venakides. The generation and propagation of oscillations in dispersive initial value problems and their limiting behavior. In Important developments in soliton theory, 205241, Springer Ser. Nonlinear Dynam., Springer, Berlin, 1993.

[25] C. Mietka. *Periodic waves in some Hamiltonian PDEs : stability, modulations and dispersive shocks.* Theses, Université de Lyon, February 2017.

[26] M. Oh and K. Zumbrun. Stability of periodic solutions of conservation laws with viscosity: analysis of the Evans function. *Arch. Ration. Mech. Anal.*, 166(2):99–166, 2003.

[27] M. Oh and K. Zumbrun. Stability of periodic solutions of conservation laws with viscosity: pointwise bounds on the Green function. *Arch. Ration. Mech. Anal.*, 166(2):167–196, 2003.

[28] M. Reed and B. Simon. Methods of modern mathematical physics. I Functional analysis. Academic Press, 1972.

[29] M. Reed and B. Simon. Methods of modern mathematical physics. IV Analysis of operators. Academic Press, 1978.

[30] D. Serre. Spectral stability of periodic solutions of viscous conservation laws: large wavelength analysis. *Comm. Partial Differential Equations*, 30(1-3):259–282, 2005.

[31] G. B. Whitham. *Linear and nonlinear waves.* John Wiley & Sons Inc., New York, 1999. Reprint of the 1974 original, A Wiley-Interscience Publication.

[32] V. E. Zakharov, L. A. Ostrovsky. Modulation instability: The beginning. Physica D 238 (2009) 540548.