Dispersive Waves - Lecture Notes

Tomáš Dohnal
Department of Mathematics, Martin Luther University Halle-Wittenberg

April 11, 2024

Contents

THESE LECTURE NOTES ARE UNDER CONSTRUCTION. I WILL APPRECIATE ANY REPORTS OF TYPOS, ERRORS OR UNCLARITIES.

0.1 Introduction

Dispersion describes the effect of distinct wavelengths propagating at different velocities. This lecture deals primarily with the group velocity dispersion, which is based on the concept of group velocity. At the moment, for a rough physical picture, let us, however, use the slightly ambiguous term 'velocity'. An elementary example of dispersion is that of light in a glass prism (or generally any other medium except for vacuum). When light travels through a material, different wavelengths propagate at different velocities. Hence, when white light enters, for instance, a glass prism from air, Snell's law tells us that the different wavelength components refract in different directions. This is called chromatic dispersion. In this lecture we will deal only with waves propagating in one material or a periodic arrangement of materials and hence refraction at interfaces will not play a role.

Even in a homogenous material dispersion still causes the separation of different wavelengths. For instance, an initially localized disturbance (a wave-packet) is a combination of waves with many different wavelengths. These waves will disintegrate and the wave-packet become gradually broader because of the different wavelength components propagating at different velocities. This effect is called group velocity dispersion and is of extreme relevance, e.g., in fiber optics. In long-haul optical fibers clever techniques need to be used to counteract dispersion in order to preserve a signal. This goes under the name of dispersion management and is based on periodically alternating the material of the fiber along the propagation direction. A group velocity dispersion effect, which you can observe even at home, is that of water waves - more precisely capillary water waves. When a relatively small object is dropped into water, the nearly circular waves emanating from the disturbance will disperse and shorter wavelengths will travel faster than longer ones.

Note, however, that not all wave propagation is dispersive. For instance, sound waves in air thankfully undergo virtually no dispersion. Hence, the music we hear five or fifty meters far from a band sounds the same (up to dissipation).

Dispersion is also a smoothing mechanism. Singularities can be described by the presence of many waves of different (mainly very small) wavelengths. Because these propagate at different velocities, the singularity will get disintegrated and weaken in time.

Chapter 1

Linear Waves

The simplest form of a wave is the *plane wave*, i.e. a wave (or generally a physical quantity) for which the temporal evolution is a simple translation and the value of which at any given time is constant over any plane orthogonal to the direction of propagation.

A *wavefront* is defined as the set of points with the same phase (i.e. the same argument). For a plane wave the wavefront at any time is a plane orthogonal to the direction of propagation and it is translated with the propagation velocity.

The most general form of a plane wave in $\mathbb{R}^n$ is thus

$$u(x,t) = f(v \cdot x - st), \qquad x \in \mathbb{R}^n, t \in \mathbb{R} \tag{1.1}$$

with the direction vector $v \in \mathbb{R}^n$, $|v| = 1$ and the speed $s \in \mathbb{R}$. Any wavefront of (1.1) at time t is the set

$$\mathcal{F} = \{x \in \mathbb{R}^n : F(x) := v \cdot x - st - c = 0\},$$

with some $c \in \mathbb{R}$. Since $\mathcal{F}$ is a level set of F, and since $\nabla F = v$, we get $\mathcal{F} \perp v$. Hence, every wavefront is an $(n-1)$-dimensional hyperplane orthogonal to v.

Typically, however, one uses the term "plane wave" for a periodic wave, like $u(x,t) = \cos(k \cdot x - \omega t)$ with $k \in \mathbb{R}^n, \omega \in \mathbb{R}$.

A general periodic wave in $\mathbb{R}^n$ can be written as $\sum_{m\in\mathbb{Z}} e^{\mathrm{i}(k_m \cdot x - \omega_m t)}$ with $k_m \in \mathbb{R}^n$ and $\omega_m \in \mathbb{R}$ for all $m \in \mathbb{Z}$. We call ω_m the (temporal) *frequency* and k_m the *wave-vector* of the $m-$th component. In the case $n = 1$ the wave-vector is called the *wave-number*.

1.1 Hyperbolic Systems of First Order with Constant Coefficients

We consider the system

$$\partial_t u = \sum_{j=1}^{n} A_j \partial_{x_j} u, \qquad x \in \mathbb{R}^n, t > 0 \tag{1.2}$$

with $A_j \in \mathbb{R}^{m\times m}$ and $u : \mathbb{R}^n \times [0,\infty) \to \mathbb{R}^m, m \in \mathbb{N}$. Shortly we will formulate a condition on the matrices A_j which makes system (1.2) hyperbolic, i.e. supporting wave solutions.

System (1.2) includes the classical wave equation as well as, for instance, Maxwell's equations in vacuum in 3D as we show next.

Example 1.1. For the wave equation

$$\partial_t^2 \psi = \Delta\psi, \quad x \in \mathbb{R}^n, t > 0$$

we let $u := \begin{pmatrix} \partial_t \psi \\ \nabla \psi \end{pmatrix}$. Clearly $u(x,t) \in \mathbb{R}^{n+1}$ so that $m = n+1$ and we get

$$\partial_t u = \begin{pmatrix} \partial_t^2 \psi \\ \nabla \partial_t \psi \end{pmatrix} = \begin{pmatrix} \nabla \cdot \nabla \psi \\ \nabla \partial_t \psi \end{pmatrix} = \sum_{j=1}^{N} A_j \partial_{x_j} u,$$

where

$$A_j = \begin{pmatrix} 0 & \dots & 1 & \dots & 0 \\ \vdots & & & & \\ 1 & & & & \\ \vdots & & & & \\ 0 & & & & \end{pmatrix} = e_1 e_{j+1}^T + e_{j+1} e_1^T,$$

with $e_j \in \mathbb{R}^m$ being the Euclidean unit (column) vector in the $j-$th direction.

Example 1.2. Maxwell's equations in vacuum in 3D ($n = 3$) read

$$\partial_t B = -\nabla \times E, \quad \partial_t E = c^2 \nabla \times B,$$

where E and B are the electric and magnetic fields respectively, and c is the speed of light in vacuum. Letting $u := \begin{pmatrix} \tilde{B} \\ E \end{pmatrix}$, where $\tilde{B} = cB$, we get

$$\partial_t u = \begin{pmatrix} \partial_t \tilde{B} \\ \partial_t E \end{pmatrix} = c \begin{pmatrix} 0 & -\nabla \times \\ \nabla \times & 0 \end{pmatrix} \begin{pmatrix} \tilde{B} \\ E \end{pmatrix} = \sum_{j=1}^{3} A_j \partial_{x_j} u,$$

where

$$A_1 = c \begin{pmatrix} 0 & Q_1 \\ Q_1^T & 0 \end{pmatrix}, \; A_2 = c \begin{pmatrix} 0 & Q_2 \\ Q_2^T & 0 \end{pmatrix}, \; A_3 = c \begin{pmatrix} 0 & Q_3 \\ Q_3^T & 0 \end{pmatrix},$$

where $Q_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$, $Q_2 = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$, and $Q_3 = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ such that all A_j are symmetric.

Let us consider the existence of plane waves for the general system (1.2). Substituting $u(x,t) = f(v \cdot x - st)$, we get

$$-sf' = \left(\sum_{j=1}^{n} v_j A_j \right) f', \tag{1.3}$$

i.e. s has to be an eigenvalue of $\sum_{j=1}^{n} v_j A_j \in \mathbb{R}^{m \times m}$ and f' a corresponding eigenvector. This forces the vector valued *function* f' to have the form

$$f'(y) = \varphi(y)\xi, \quad \varphi : \mathbb{R} \to \mathbb{R},$$

where (s, ξ) is an eigenpair of $\sum_{j=1}^{n} v_j A_j$.

For a hyperbolic system we expect m linearly independent plane waves for each direction vector v. This requires firstly that the eigenvalues s are real (so that they describe the wave speed), and secondly the diagonalizability of $\sum_{j=1}^{n} v_j A_j$.

Definition 1.1. *System* (1.2) *is called hyperbolic if for every* $v \in \mathbb{R}^n$ *the matrix* $\sum_{j=1}^{n} v_j A_j$ *is diagonalizable and has only real eigenvalues.*

Remark 1.1. *For cosine plane waves*

$$u(x,t) = \cos(k \cdot x - \omega t)\vec{c} \quad \textit{with} \quad \vec{c} \in \mathbb{R}^m$$

we have $v = k/|k|$ *and* $s = \omega/|k|$ *so that the eigenvalue problem* (1.3) *becomes*

$$-\omega \vec{c} = \left(\sum_{j=1}^{n} k_j A_j \right) \vec{c}.$$

Here and below $|k|^2 = \sum_{j=1}^n k_j^2$ *for* $k \in \mathbb{R}^n$*. Clearly,* ω *is linear in* $|k|$

$$\omega(k) = |k|\tilde{\omega}\left(\frac{k}{|k|}\right),$$

so that $s = \tilde{\omega}(k/|k|)$*. The phase velocity* $sv = \omega k/|k|^2$*, which is also defined in Sec. 1.2, thus depends only on the direction vector* $k/|k|$

$$sv = \frac{k}{|k|}\tilde{\omega}\left(\frac{k}{|k|}\right).$$

The phase velocity is independent of $|k|$*, hence of the wavelength, which is what we understand under a non-dispersive wave propagation. Although in Sec. 1.2 we will define dispersive problems with the help of the group velocity, the two definitions coincide for the case of systems of the type* (1.2)*.*

1.1.1 Symmetric Hyperbolic Systems

Starting here (and in most sections thereafter) we use the Fourier transform as a main tool. An overview of relevant definitions and results on this topic is in Appendix A.

In this section we assume $A_j = A_j^T$ for all j in (1.2). Clearly, when A_j are symmetric, the system is automatically hyperbolic. Let us now consider the Cauchy problem. Appending equation (1.2) with the initial data $u(x, 0) = u_0(x)$ with $u_0 \in L^2(\mathbb{R}^n)$, we perform the Fourier transform and obtain

$$\partial_t \widehat{u}(k,t) = \left(\sum_{j=1}^n \mathrm{i}k_j A_j\right)\widehat{u}(k,t) =: \widehat{P}(k)\widehat{u}(k,t)$$

with the initial condition $\widehat{u}(k, 0) = \widehat{u}_0(k)$. This is an ODE problem for each k and the solution is

$$\widehat{u}(k,t) = e^{\widehat{P}(k)t}\widehat{u}_0(k).$$

Writing $\widehat{P}(k) = \mathrm{i}|k|\sum_{j=1}^n v_j A_j$ with $v = \frac{k}{|k|}$ and using the symmetry of A_j, we get

$$e^{\widehat{P}(k)t} = Q(v)\begin{pmatrix} e^{-\mathrm{i}s_1(v)|k|t} & & \\ & \ddots & \\ & & e^{-\mathrm{i}s_m(v)|k|t}\end{pmatrix} Q^T(v)$$

with an orthogonal matrix $Q(v) \in \mathbb{R}^{m\times m}$ and $\{-s_1, \ldots, -s_m\}$ being the eigenvalues of $\sum_{j=1}^n v_j A_j$. In the physical $x-$space we get the superposition of plane waves

$$u(x,t) = (2\pi)^{-n/2}\sum_{j,l=1}^m \int_{\mathbb{R}^n} q_j(v(k))e^{\mathrm{i}|k|(v(k)\cdot x - s_j(v(k))t)}q_l^T(v(k))\widehat{u}_0(k)\,\mathrm{d}k.$$

This representation of the solution lets us formulate two important results, namely the conservation of the L^2-norm and the finite propagation speed for symmetric hyperbolic systems.

Theorem 1.2 (Conservation of the L^2-norm)**.** *The solution of* (1.2) *with* $A_j = A_j^T$ *for all* $j = 1, ..., n$ *and with* $u(\cdot, 0) = u_0 \in (L^2(\mathbb{R}^n))^m$ *satisfies*

$$\sum_{j=1}^m \|u_j(\cdot,t)\|^2_{L^2(\mathbb{R}^n)} = \sum_{j=1}^m \|u_{0,j}\|^2_{L^2(\mathbb{R}^n)} \qquad \textit{for all } t \geqslant 0.$$

Proof. The idea is that because Q is orthogonal and $e^{-\mathrm{i}\Lambda(k)t} := \begin{pmatrix} e^{-\mathrm{i}s_1(v)|k|t} & & \\ & \ddots & \\ & & e^{-\mathrm{i}s_n(v)|k|t} \end{pmatrix}$ is unitary, also the matrix $e^{\widehat{P}(k)t}$ is unitary. In detail, by Plancherel's identity (A.1) and orthogonality of Q

$$\begin{aligned}\sum_{j=1}^m \|u_j(\cdot,t)\|^2_{L^2(\mathbb{R}^n)} &= \int_{\mathbb{R}^n} \widehat{u}^*(k,t)\widehat{u}(k,t)dk = \int_{\mathbb{R}^n} \widehat{u}_0^*(k)Q(v)e^{\mathrm{i}\Lambda(k)t}Q^T(v)Q(v)e^{-\mathrm{i}\Lambda(k)t}Q^T(v)\widehat{u}_0(k)dk \\ &= \int_{\mathbb{R}^n} \widehat{u}_0^*(k,t)\widehat{u}_0(k,t)dk = \sum_{j=1}^m \|u_{0,j}\|^2_{L^2(\mathbb{R}^n)}\end{aligned}$$

because $Q^TQ = QQ^T = I$. □

Theorem 1.3 (Finite propagation speed)**.** *Consider* (1.2) *with* $A_j = A_j^T$ *for all* $j = 1, ..., n$ *and with* $u(\cdot, 0) = u_0 \in (L^2(\mathbb{R}^n))^m$. *If*

$$\operatorname{supp}(u_0) \subset B_R = \{x \in \mathbb{R}^n : |x| \leqslant R\}$$

for some $R > 0$, *then*

$$\operatorname{supp}(u(\cdot,t)) \subset B_{R+s_{\max}t},$$

where

$$s_{\max} = \max_{|v|=1} \lambda_{\max}\left(\sum_{j=1}^n v_jA_j\right) = \max_{|v|=1}\max_{|\xi|=1} \xi^T\left(\sum_{j=1}^n v_jA_j\right)\xi.$$

Proof. The idea is to apply the Paley-Wiener Theorem A.11. For that we need to first show that $\widehat{u}(\cdot, t)$ is holomorphic in $\mathbb{C}^n$. Because $\widehat{u}(k,t) = e^{\widehat{P}(k)t}\widehat{u}_0(k)$, we consider $\widehat{u}_0$ and $e^{\widehat{P}t}$ separately.

That $\widehat{u}_0$ is entire on $\mathbb{C}^n$ follows from the Paley-Wiener theorem due to the compact support of u_0 and the fact that as an $L^2(\mathbb{R}^n)$ function it is certainly in $L^1_{\mathrm{loc}}(\mathbb{R}^n)$ and hence defines a tempered distribution. Moreover, the Paley-Wiener theorem guarantees that there are $C > 0, N \in \mathbb{N}$ so that

$$|\widehat{u}_0(k)| \leqslant C(1+|k|)^N e^{|\operatorname{Im}(k)|R} \quad \text{for all } k \in \mathbb{C}^n.$$

Note that the definition of $\widehat{u}_0(k)$ for $k \in \mathbb{C}^n$ is the same as for $k \in \mathbb{R}^n$, i.e. $\widehat{u}_0(k) = (2\pi)^{-n/2}\int_{\mathbb{R}^n} u_0(x)e^{-\mathrm{i}k\cdot x}\,\mathrm{d}x$.

For $e^{\widehat{P}t}$ note that

$$k \mapsto e^{\widehat{P}(k)t} = e^{\mathrm{i}t\sum_{j=1}^n k_jA_j}$$

is holomorphic on $\mathbb{C}^n$.

It remains to verify the growth condition on $\widehat{u}(k,t)$ in Theorem A.11. We have

$$\frac{d|\widehat{u}|^2}{dt} = \frac{d}{dt}(\widehat{u}^*\widehat{u}) = (\widehat{P}\widehat{u})^*\widehat{u} + \widehat{u}^*\widehat{P}\widehat{u} = \widehat{u}^*(\widehat{P}+\widehat{P}^*)\widehat{u} = -2\widehat{u}^*\left(\sum_{j=1}^n \operatorname{Im}(k_j)A_j\right)\widehat{u},$$

where in the last step the symmetry of A_j was used. We obtain

$$\frac{d|\widehat{u}|}{dt} = \frac{\widehat{u}^*\left(-\sum_{j=1}^n \operatorname{Im}(k_j)A_j\right)\widehat{u}}{|\widehat{u}|^2}|\widehat{u}| \leqslant \lambda_{\max}\left(-\sum_{j=1}^n \operatorname{Im}(k_j)A_j\right)|\widehat{u}|,$$

and thus

$$\begin{aligned}|\widehat{u}(k,t)| &\leqslant e^{t\lambda_{\max}\left(-\sum_{j=1}^n \operatorname{Im}(k_j)A_j\right)}|\widehat{u}_0(k,t)| \\ &\leqslant e^{t|\operatorname{Im}(k)|\lambda_{\max}\left(-\sum_{j=1}^n v_jA_j\right)}|\widehat{u}_0(k,t)| \\ &\leqslant e^{t|\operatorname{Im}(k)|s_{\max}}C(1+|k|)^N e^{|\operatorname{Im}(k)|R},\end{aligned}$$

where $v = k/|k|$. Finally, another application of Paley-Wiener Theorem A.11 guarantees $\operatorname{supp}(u(\cdot,t)) \subset B_{R+s_{\max}t}$. □

1.1.2 Wave Equation

Although, as shown above, the wave equation

$$\begin{aligned} \partial_t^2 u &= c^2 \Delta u, \qquad x \in \mathbb{R}^n, t > 0 \\ u(x,0) &= f(x), \partial_t u(c,0) = g(x) \end{aligned} \tag{1.4}$$

(with $c > 0$) can be written in the form of a hyperbolic first order system (1.2), we consider this important equation here separately and derive an explicit form of the solution $u(x,t)$. The results of Section 1.1.1 produce only a formula for $\partial_t u$ and ∇u. We will also be able derive more detailed information on the qualitative properties of the solution in this section.

Using the Fourier transformation, the problem becomes

$$\partial_t^2 \widehat{u} = -c^2 |k|^2 \widehat{u}, \quad \widehat{u}(k,0) = \hat{f}(k), \partial_t \widehat{u}(k,0) = \hat{g}(k).$$

This second order ODE has the solution

$$\widehat{u}(k,t) = \cos(c|k|t)\hat{f}(k) + \frac{\sin(c|k|t)}{c|k|}\hat{g}(k) = \frac{\sin(c|k|t)}{c|k|}\hat{g}(k) + \partial_t \left(\frac{\sin(c|k|t)}{c|k|}\hat{f}(k) \right).$$

We define $\widehat{W}(k,t) := \frac{\sin(c|k|t)}{c|k|}$. The difficulty is that for $n > 1$ the inverse Fourier transform of $\widehat{W}$ is no function but only a tempered distribution.

We begin with the case $n = 3$. It is left as an exercise to show that W is given by

$$W : \psi \mapsto W(\psi)(t) = \frac{1}{4\pi c^2 t} \int_{|x|=ct} \psi(x) dx \quad \text{for all } \psi \in S(\mathbb{R}^3).$$

Hence, using the definition of the convolution of a distribution T and a Schwartz function, cf. Lemma A.7, we get for $f, g \in S(\mathbb{R}^3)$ the *Kirchhoff's formula*

$$u(x,t) = \frac{1}{4\pi c^2 t} \int_{|x-y|=ct} g(y) dy + \partial_t \left(\frac{1}{4\pi c^2 t} \int_{|x-y|=ct} f(y) dy \right). \tag{1.5}$$

Note that (1.5) is a solution of (1.4) (with $n = 3$) also for $f \in \mathbb{C}^3(\mathbb{R}^3), g \in C^2(\mathbb{R}^3)$ as one can check by differentiation, see [9].

From formula (1.5) one can deduce several important properties of the solution. Let us assume that the initial data are compactly supported, i.e.

$$\text{supp} f \cup \text{supp} g \subset B_R(0) \quad \text{for some } R > 0.$$

1. There is a finite speed of propagation, i.e. for each $t > 0$ the solution $u(\cdot, t)$ is compactly supported.
2. The signal has a finite lifetime at any point x_0, i.e. an observer can see/hear/feel the signal only for a bounded interval of time. At a point x_0 with $|x_0| > R$ the solution $u(x_0, t)$ is zero if $R + ct < |x_0|$, i.e. $t < \frac{|x_0| - R}{c}$ and if $-R + ct > |x_0|$, i.e. $t > \frac{|x_0| + R}{c}$, see Fig. 1.1.2.
3. At any time $t > R/c$ is $\text{supp} u(\cdot, t)$ inside a spherical shell $B_{R+ct}(0) \backslash B_{-R+ct}(0)$, see Fig. 1.1.2.

Let us now study the case $n = 2$. For $f \in C^3(\mathbb{R}^2), g \in C^2(\mathbb{R}^2)$ we can use the 3D solution formula (1.5) with $f^{(3)}(x_1, x_2, x_3) := f(x_1, x_2)$ and $g^{(3)}(x_1, x_2, x_3) := g(x_1, x_2)$. Because also the resulting u is independent of x_3, we can set e.g. $x_3 = 0$ in (1.5) and obtain

$$u(x,t) = \frac{1}{4\pi c^2 t} \int_{|\tilde{x}-y|=ct} g^{(3)}(y) dy + \partial_t \left(\frac{1}{4\pi c^2 t} \int_{|\tilde{x}-y|=ct} f^{(3)}(y) dy \right) \quad \text{for } x \in \mathbb{R}^2, \tag{1.6}$$

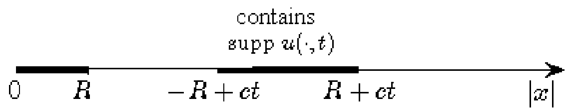

Figure 1.1: Schematic for the signal lifetime and propagation speed of the wave equation in 3D ($n = 3$).

where $\tilde{x} := (x^T, 0)^T \in \mathbb{R}^3$. Clearly, if supp$f$, supp$g \subset \Omega \subset \mathbb{R}^2$, then supp$f^{(3)}$, supp$g^{(3)} \subset \Omega \times \mathbb{R}$. We suppose that Ω is compact. An observer at $x_0 \in \mathbb{R}^2$ feels the wave for all times $t \geqslant t_{\min}(x_0) := \frac{\mathrm{dist}(x_0,\Omega)}{c}$ because the sphere $\{y : |x_0 - y| = ct\}$ cuts the support $\Omega \times \mathbb{R}$ for all such t, see Fig. 1.1.2. Hence the signal never dies down. In a 2D world everyone has an earworm - always! After a variable transformation (see [9]) solution (1.6) can be written as

$$u(x,t) = \frac{1}{2\pi c}\left(\int_{|x-y|\leqslant ct} \frac{g(y)}{(c^2t^2 - |x-y|^2)^{1/2}}dy + \partial_t\left(\int_{|x-y|\leqslant ct} \frac{f(y)}{(c^2t^2 - |x-y|^2)^{1/2}}dy\right)\right) \quad \text{for } x \in \mathbb{R}^2. \quad (1.7)$$

Finally, we inspect the case $n = 1$. Here the inverse Fourier transform of $\widehat{W}$ is, in fact, a function. We have

$$\widehat{W}(k,t) = \frac{\sin(ckt)}{ck}, \quad W(x,t) = \frac{1}{c}\left(\frac{\pi}{2}\right)^{1/2} \chi_{[-ct,ct]}(x)$$

as one easily checks. Hence

$$u(x,t) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} W(x-y,t)g(y)dy + \frac{1}{\sqrt{2\pi}}\partial_t \int_{\mathbb{R}} W(x-y,t)f(y)dy = \frac{1}{2c}\int_{|x-y|\leqslant ct} g(y)dy + \frac{1}{2c}\partial_t \int_{|x-y|\leqslant ct} f(y)dy.$$

By evaluating the derivatives, we arrive at the *d'Alembert's formula*

$$u(x,t) = \frac{1}{2c}\int_{x-ct}^{x+ct} g(y)dy + \frac{1}{2}\left(f(x+ct) - f(x-ct)\right), \quad (1.8)$$

which is the classical solution for any $f \in C^2(\mathbb{R}), g \in C^1(\mathbb{R})$.

We conclude that for $n = 1$ there is a finite propagation speed but the life time of a generic signal is infinite as the support of g remains in the integration domain of the integral $\int_{x-ct}^{x+ct} g(y)dy$ for all times after the time $t_* = \frac{1}{c}\mathrm{dist}(x, \mathrm{supp}(g))$.

1.2 The Fundamentals of Dispersive Waves

The definition of a dispersive problem will be given a bit later but we can give away some classical examples of dispersive equations already now.

- Schrödinger equation

$$\mathrm{i}\partial_t u + \Delta u = 0, \qquad x \in \mathbb{R}^n \quad (1.9)$$

- linear Korteweg-de Vries equation

$$\partial_t u + a\partial_x u + \partial_x^3 u = 0, \qquad a \in \mathbb{R}, x \in \mathbb{R} \quad (1.10)$$

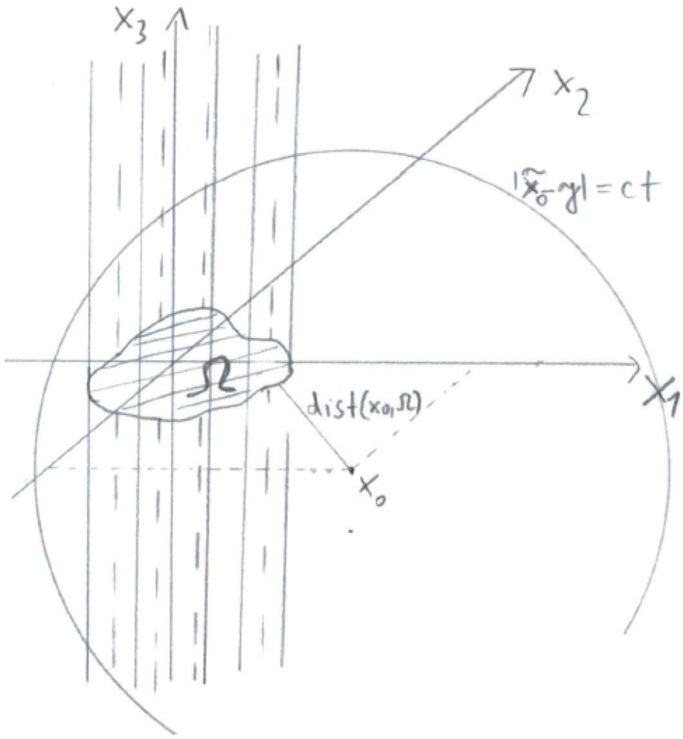

Figure 1.2: Schematic for the signal lifetime of the wave equation in 2D ($n = 2$).

- Klein-Gordon equation

$$\partial_t^2 u = \partial_x^2 u - au, \qquad a \geqslant 0, x \in \mathbb{R} \tag{1.11}$$

The Schrödinger equation describes the wave function (as the quantum state) of a particle in free space but also the envelope of a classical wave packet in a homogenous medium. The linear Korteweg-de Vries equation describes shallow water waves in the linear limit. For a full discussion see Section 2.1.1. The Klein-Gordon equation is a relativistic version of the Schrödinger equation but it can be also derived from the wave equation in a waveguide as the equation describing the longitudinal dynamics. This derivation follows.

Consider the wave equation

$$\partial_t^2 u = \Delta u, \qquad x \in \Omega = \Sigma \times \mathbb{R} \quad \text{with } \Sigma \subset \mathbb{R}^2 \text{ open and bounded}$$

and with homogenous Dirichlet or homogenous Neumann boundary conditions on $\partial\Omega = \partial\Sigma \times \mathbb{R}$, denoted by

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

In addition we have initial conditions

$$u(x, 0) = f(x), \quad \partial_t u(x, 0) = g(x).$$

The waveguide Ω is schematically depicted in Fig. 1.2. We are going to expand the solution at each (x_3, t) in the eigenfunctions of the corresponding 2D eigenvalue problem in (x_1, x_2). The Klein-Gordon equation will arise as the equation governing the (x_3, t)-dependent coefficients in the expansion.

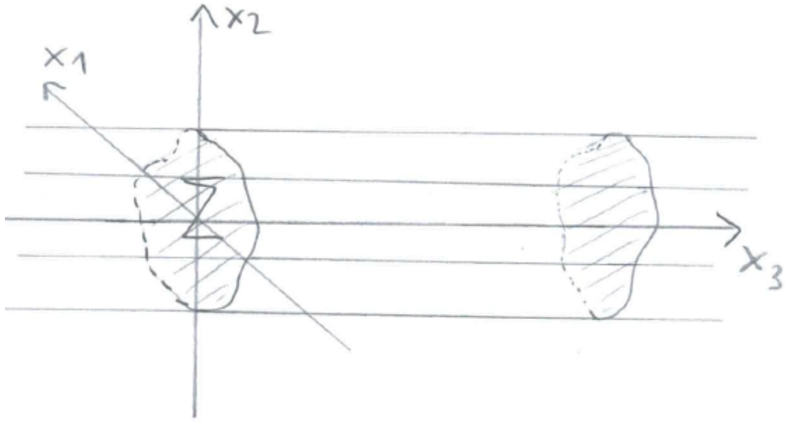

Figure 1.3: A waveguide for the Klein-Gordon example.

The (symmetric elliptic) eigenvalue problem

$$-\Delta_{x_1,x_2}\psi = \lambda\psi \qquad \text{on} \quad \Sigma$$
$$R\psi = 0 \qquad \text{on} \quad \partial\Sigma$$

provides an orthonormal basis of $L^2(\Sigma)$ via its eigenfunctions $(w_k)_{k\in\mathbb{N}}$ with $w_k \in H_0^1(\Sigma)$ for Dirichlet boundary conditions and $w_k \in H_1(\Sigma)$ for Neumann boundary conditions. The eigenvalues are non-negative (positive in the Dirichlet case)

$$\lambda_k = \mu_k^2 \geqslant 0 \qquad \text{for all} \quad k \in \mathbb{N}.$$

We expand the initial data at each point $x_3 \in \mathbb{R}$ and expand the solution at each $(x_3, t) \in \mathbb{R} \times [0, \infty)$ in the eigenfunctions:

$$f(x) = \sum_{k\in\mathbb{N}} f_k(x_3)w_k(x_1, x_2), \qquad g(x) = \sum_{k\in\mathbb{N}} g_k(x_3)w_k(x_1, x_2),$$
$$u(x,t) = \sum_{k\in\mathbb{N}} u_k(x_3, t)w_k(x_1, x_2).$$

The coefficient u_k then satisfies the Klein-Gordon equation with $a = \mu_k^2$

$$\partial_t^2 u_k = \partial_{x_3}^2 u_k - \mu_k^2 u_k,$$
$$u_k(x_3, 0) = f_k(x_3), \quad \partial_t u_k(x_3, 0) = g_k(x_3).$$

1.2.1 Dispersion Relation, Phase Velocity, Group Velocity

We provide here a rather general theory of dispersive PDEs and explain the roles of the dispersion relation as well as the phase and group velocities. This exposition is along the lines of Section 11.1 in [26] but we are a bit more restrictive in the definition of dispersive systems in order to be able to justify the use of elementary plane wave solutions.

1.2.1.1 Dispersion Relation and Dispersive PDEs

Systems of $m \in \mathbb{N}$ linear PDEs with constant coefficients on $\mathbb{R}^n$ can be written compactly as

$$P(\partial_t, \partial_{x_1}, \ldots, \partial_{x_n})\vec{u} = 0, \qquad x \in \mathbb{R}^n, t > 0, \tag{1.12}$$

where $\vec{u} \in \mathbb{C}^m$ and $P : \mathbb{R}^{n+1} \to \mathbb{C}^{m\times m}$ is a matrix valued polynomial. We will limit our attention mainly to systems which have the form

$$P(\partial_t, \partial_{x_1}, \ldots, \partial_{x_n}) = A(\partial_{x_1}, \ldots, \partial_{x_n})\partial_t - Q(\partial_{x_1}, \ldots, \partial_{x_n}),$$

i.e.

$$A(\partial_{x_1}, \ldots, \partial_{x_n})\partial_t\vec{u} - Q(\partial_{x_1}, \ldots, \partial_{x_n})\vec{u} = 0, \qquad x \in \mathbb{R}^n, t > 0, \tag{1.13}$$

where $A, Q : \mathbb{C}^n \to \mathbb{C}^{m\times m}$ are polynomials and $A(l)$ is non-singular for each $l \in \mathbb{C}^n$. In other words, the equations first order in ∂_t and the time derivative appears in each equation. Of course, if a higher order derivative $\partial_t^p, p > 1$ appears in one of the equations, the system can be transformed to a system of $m + p - 1$ equations of first order. Examples of (1.13) are numerous and include all scalar linear equations of first order in t, for instance the linear Schrödinger and the linear KdV equations or more exotic problems like the linear Kadomtsev-Petviashvili equation

$$\partial_{x_1}(\partial_t u + \partial_{x_1}^3 u) + \partial_{x_2}^2 u = 0, \qquad x \in \mathbb{R}^2.$$

It also includes many systems, e.g. the linear coupled mode equations in $x \in \mathbb{R}$

$$\begin{aligned} \mathrm{i}(\partial_t u + \partial_x u) + \kappa v &= 0 \\ \mathrm{i}(\partial_t v - \partial_x v) + \kappa u &= 0, \end{aligned}$$

with $\kappa \in \mathbb{R}$.

A large subclass of problems satisfy $A = I$, i.e. have the simple form

$$\partial_t\vec{u} = Q(\partial_{x_1}, \ldots, \partial_{x_n})\vec{u}, \qquad x \in \mathbb{R}^n, t > 0. \tag{1.14}$$

As an example the Klein-Gordon equation (1.11) can be written in the form $\partial_t\vec{v} = Q(\partial_x)\vec{v}$ for $\vec{v} = (u, \partial_t u)^T$ and $Q = \begin{pmatrix} 0 & 1 \\ \partial_x^2 - a & 0 \end{pmatrix}$.

Applying the Fourier transform to (1.14) produces

$$\partial_t\widehat{\vec{u}} = Q(\mathrm{i}k_1, \ldots, \mathrm{i}k_n)\widehat{\vec{u}}$$

for each $k \in \mathbb{R}^n$. If $Q(\mathrm{i}k)$ is diagonalizable, i.e.

$$Q(\mathrm{i}k) = X(k)\Lambda(k)X^{-1}(k), \qquad \Lambda(k) = \begin{pmatrix} \lambda_1(k) & & \\ & \ddots & \\ & & \lambda_m(k) \end{pmatrix}, \; X(k) = \left(\xi^{(1)}(k), \ldots, \xi^{(m)}(k)\right),$$

where $\xi^{(1)}, \ldots, \xi^{(m)}$ are eigenvectors to the eigenvalues $\lambda_1, \ldots, \lambda_m$ of $Q(\mathrm{i}k)$, then solutions have the form

$$\widehat{\vec{u}}(k, t) = X(k)e^{\Lambda(k)t}\vec{c}(k)$$

with some vector $\vec{c} \in \mathbb{C}^m$. This is a linear combination of the *solution modes* $e^{\lambda_j(k)t}\xi^{(j)}(k)$, $j = 1, \ldots, m$. In the physical x-space the solution $u(x, t)$ is then a $k-$integral of the *elementary solutions*

$$e^{\lambda_j(k)t + \mathrm{i}k\cdot x}\xi^{(j)}(k) \tag{1.15}$$

weighted by the Fourier transform of the initial data. For pure wave propagation problems we expect and require that the eigenvalues λ_j are imaginary so that no damping or gain appears. Therefore, we denote from now on the spectral unknown λ by $-\mathrm{i}\omega$. We have thus shown that solutions of (1.14) are built out of elementary solutions which have the form of plane-waves

$$e^{\mathrm{i}(k\cdot x - \omega(k)t)}\vec{c}. \tag{1.16}$$

As we will see below, the nature of the wave propagation is largely determined by the form of the function $\omega(k)$. This function, being an eigenvalue of $\mathrm{i}Q(\mathrm{i}k)$, is necessarily a solution of

$$\det\left(\omega I - \mathrm{i}Q(\mathrm{i}k_1,\ldots,\mathrm{i}k_n)\right) = 0.$$

For the more general systems (1.13) this equation becomes

$$\det\left(\omega A(\mathrm{i}k_1,\ldots,\mathrm{i}k_n) - \mathrm{i}Q(\mathrm{i}k_1,\ldots,\mathrm{i}k_n)\right) = 0,$$

and $\xi^{(j)}(k)$ are eigenvectors of the generalized eigenvalue problem $\omega A(\mathrm{i}k_1,\ldots,\mathrm{i}k_n)\xi(k) = \mathrm{i}Q(\mathrm{i}k_1,\ldots,\mathrm{i}k_n)\xi(k)$. We assume again that the eigenspace is m-dimensional. The solution is again an integral of elementary solutions (1.15).

In the general case (1.12) after the Fourier transform we get

$$P(\partial_t,\mathrm{i}k_1,\ldots,\mathrm{i}k_n)\widehat{\vec{u}}(k,t) = 0.$$

We assume that there are again m solutions $e^{\mu_j(k)t}\zeta^{(j)}(k)$ for each $k\in\mathbb{R}^n$ with $\zeta^{(1)}(k),\ldots,\zeta^{(m)}(k)$ as solutions of

$$P(\mu_j,\mathrm{i}k_1,\ldots,\mathrm{i}k_n)\zeta^{(j)}(k) = 0$$

being linearly independent. With $\mu =: -\mathrm{i}\omega$, we obtain, once again, the plane-wave elementary solutions (1.16). In the general system (1.12) the frequency ω thus has to satisfy

$$\det\left(P(-\mathrm{i}\omega,\mathrm{i}k_1,\ldots,\mathrm{i}k_n)\right) = 0. \tag{1.17}$$

Definition 1.4. *Equation* (1.17) *is called the* ***dispersion relation*** *of equation* (1.12).

Being an algebraic equation of m-th degree in ω, the dispersion relation (1.17) has m solutions $W_1(k),\ldots W_m(k)$. This produces m elementary solutions of (1.12).

1.2.1.2 Phase Velocity and Group Velocity

Let now

$$\theta := k\cdot x - W_j(k)t$$

be the phase of the $j-$th solution mode. The wavefronts $\theta=$const. travel with the so called phase velocity.

Definition 1.5. *The* ***phase velocity*** *of the $j-$th solution mode of* (1.13) *is*

$$v_p^{(j)}(k) = \frac{W_j(k)}{|k|^2}k.$$

In 1D ($n=1$) this, of course, reduces to $v_p^{(j)}(k) = W_j(k)/k$.

Dispersion roughly means is that waves with different wave numbers propagate with different velocities. The first idea could be to define dispersive problems as those, for which $\nabla v_p^{(j)}(k) \not\equiv 0$. A simple example, however, shows that this is not a suitable definition. Consider the dispersion relation $\omega - ak - b = 0$ with the solution $W(k) = ak+b$. The elementary solution then has the form

$$e^{-\mathrm{i}bt}e^{\mathrm{i}k(x-at)}.$$

Although $v_p(k) = a + b/k$ is not constant in k, the corresponding equation $\partial_t u + a\partial_x u + \mathrm{i}bu = 0$ is not dispersive because the solution of the corresponding PDE with initial data $u_0(x)$ is

$$u(x,t) = e^{-\mathrm{i}bt}(2\pi)^{-1/2}\int_{\mathbb{R}} e^{\mathrm{i}k(x-at)}\widehat{u}_0(k)\,\mathrm{d}k = e^{-\mathrm{i}bt}u_0(x-t).$$

The equation is a mere transport problem with a simple phase rotation.

Dispersion is mathematically defined via the group velocity instead.

Definition 1.6. *The **group velocity** of the $j-th$ solution mode of* (1.13) *is*

$$v_g^{(j)}(k) = \nabla W_j(k).$$

Definition 1.7. *The $j-th$ solution mode of* (1.13) *is dispersive if*

$$\det(D^2 W_j) \not\equiv 0.$$

Here $\det D^2 W_j$ denotes the Hessian matrix of W_j.

Definition 1.8. *Equation* (1.13) *is **dispersive** if $W_j(k) \in \mathbb{R}$ for all $j = 1, \dots, m$, if the eigenspace of the generalized eigenvalue problem*

$$\omega A(\mathrm{i}k)\xi = \mathrm{i}Q(\mathrm{i}k)\xi$$

is m-dimensional for each $k \in \mathbb{R}^n$ and if the eigenvalues $\omega = W_j(k)$ satisfy $\det(D^2 W_j) \not\equiv 0$ *for at least one $j \in \{1, \dots, m\}$.*

In 1D the definition of a dispersive mode and equation is equivalent to the condition that W_j are nonlinear functions of k. In nD this is not the case since e.g. $W(k) = k_1^2 + k_2$ satisfies $\det(D^2 W) \equiv 0$. In nD the definition is thus slightly more restrictive but it is the definition that yields certain asymptotic calculations meaningful, see Section 1.2.2.

Also note that for the special case (1.14) the condition of m linearly independent solutions $\zeta^{(j)}(k), j = 1, \dots, m$ is equivalent to the diagonalizability condition of $Q(\mathrm{i}k)$.

Clearly, the form of equation (1.12) allows only polynomial dispersion relations. As we will see in Section 1.5, for water waves a more general dependence $\omega(k)$ holds and satisfies $\det D^2\omega \not\equiv 0$. The definition of a dispersive problem is therefore often extended to any problem for which a dispersion relation makes sense, has real solutions $W_j(k)$ and where $\det D^2 W_j \not\equiv 0$ for at least one j.

1.2.2 Asymptotic Role of the Group Velocity

We consider here the scalar 1D case $m = n = 1$ and study the asymptotics of the solution of a dispersive equation (1.14) for

$$t \to \infty, \qquad \frac{x}{t} = c$$

with a constant $c \in \mathbb{R}$.

The solution is

$$u(x,t) = (2\pi)^{-1/2} \int_{\mathbb{R}} \widehat{u}_0(k) e^{-\mathrm{i}\chi(k)t}\, \mathrm{d}k, \qquad \chi(k) = W(k) - k\frac{x}{t}. \tag{1.18}$$

For large values of t the exponential $e^{-\mathrm{i}\chi(k)t}$ oscillates (for generic functions $\chi(k)$) fast in k and causes a cancellation effect in the integral. The oscillatory behavior appears everywhere except near points $k = \tilde{k}$ where $\chi(k)$ is close to constant. The largest contribution to the integral thus comes from the neighborhood of stationary points of χ, i.e. from points $k = \tilde{k}$, such that

$$\chi'(\tilde{k}) = W'(\tilde{k}) - c = 0.$$

This is the idea of the method of stationary phase. A rigorous discussion of this method and the asymptotic notation used below is in Appendix B. Here we provide only the main ideas of the calculation of the asymptotics of u. Assuming

$$W''(\tilde{k}) \neq 0 \quad \text{for all } \tilde{k} \quad \text{with } W'(\tilde{k}) = c,$$

we have (using $\chi''(k) = W''(k)$)

$$\begin{aligned} \chi(k) &\sim \chi(\tilde{k}) + \tfrac{1}{2} W''(\tilde{k})(k - \tilde{k})^2, \\ \widehat{u}_0(k) &\sim \widehat{u}_0(\tilde{k}) \end{aligned}$$

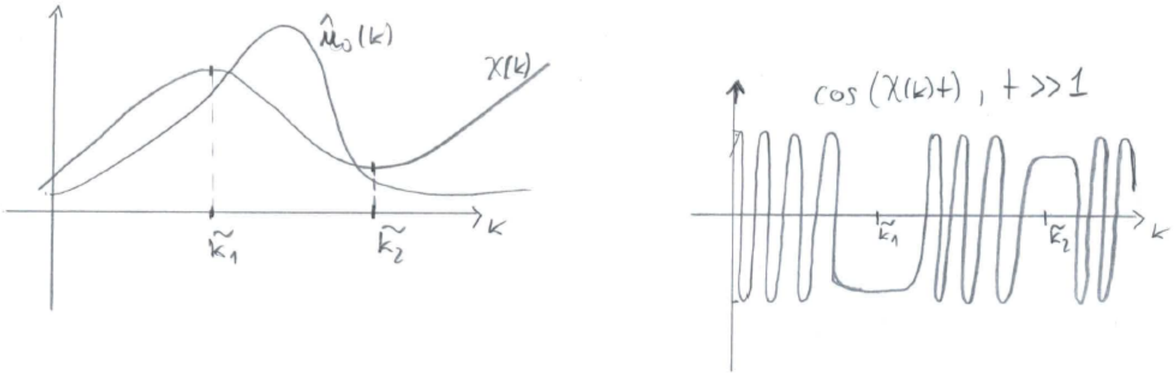

Figure 1.4: Left: example of $\widehat{u}_0(k)$ and the phase $\chi(k)$. Right: the function $e^{\mathrm{i}\chi(k)t}$ (the real part is plotted) for $t \gg 1$ is highly oscillatory except near critical points of χ.

as $k \to \tilde{k}$. With the help of Lemmas B.5 and B.6 like in the proof of Theorem B.7 (and with Remark B.6), we get

$$u(x,t) \sim (2\pi)^{-1/2} \sum_{W'(\tilde{k})=c} \widehat{u}_0(\tilde{k}) e^{-\mathrm{i}\chi(\tilde{k})t} \int_{\mathbb{R}} e^{-\frac{\mathrm{i}}{2}W''(\tilde{k})t(k-\tilde{k})^2}\,\mathrm{d}k \tag{1.19}$$

as $t \to \infty$, $\frac{x}{t} = c$. As W (as well as W') is polynomial, we have $W'(k) = c$ at finitely many points $\tilde{k} \in \{\tilde{k}_1, \ldots, \tilde{k}_m\}$ (with $\tilde{k}_j < \tilde{k}_{j+1}$). In the application of the above lemmas we split the integral in (1.18) into integrals over $(-\infty, \tilde{k}_1)$, $(\tilde{k}_1, \tilde{k}_2)$, ..., and $(\tilde{k}_m, \infty)$. For the assumptions in Theorem B.7 and Remark B.6 we need $\widehat{u}_0 \in L^1(\mathbb{R}) \cap C^1(\mathbb{R})$, $(\widehat{u}_0/\chi')' \in L^1(\mathbb{R}\backslash(\tilde{k}_1, \tilde{k}_m))$, and $(\widehat{u}_0/\chi')(k) \to 0$ for $|k| \to \infty$. As χ' is a polynomial, all these are satisfied if we choose $u_0 \in S(\mathbb{R})$ (such that also $\widehat{u}_0 \in S(\mathbb{R})$).

The integral on the right hand side of (1.19) can be evaluated using the result of Example B.1 and a suitable substitution. Let us, however, perform the simple calculation using Cauchy's integral theorem for our case separately. The integral in question has the form

$$\int_{\mathbb{R}} e^{-\mathrm{i}\alpha x^2}\,\mathrm{d}x = 2\int_0^\infty e^{-\mathrm{i}\alpha x^2}\,\mathrm{d}x.$$

The idea is to extend the integrand to the complex plane and integrate this holomorphic function along a closed contour which includes the real positive axis.

For $\alpha > 0$ we choose the contour $C := C_1 \cup C_2 \cup C_3$ as in Fig. 1.2.2 with $R > 0$ and let $R \to \infty$. Cauchy's theorem gives $\int_C e^{-\mathrm{i}\alpha z^2}\,\mathrm{d}z = 0$ so that

$$\int_0^R e^{-\mathrm{i}\alpha z^2}\,\mathrm{d}z = -\int_{C_2} e^{-\mathrm{i}\alpha z^2}\,\mathrm{d}z - \int_{C_3} e^{-\mathrm{i}\alpha z^2}\,\mathrm{d}z$$

By a direct calculation or by the mean value theorem of complex analysis we get

$$\int_{C_2} \to 0 \quad \text{as} \quad R \to \infty.$$

For $\int_{C_3}$ we get with the substitution $r = ze^{\mathrm{i}\pi/4}$

$$-\int_{C_3} e^{-\mathrm{i}\alpha z^2}\,\mathrm{d}z \to e^{-\mathrm{i}\pi/4}\int_0^\infty e^{-\alpha r^2}\,\mathrm{d}r = e^{-\mathrm{i}\pi/4}\frac{1}{2}\left(\frac{\pi}{\alpha}\right)^{1/2}$$

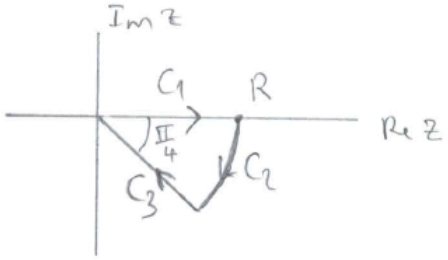

Figure 1.5: Contour for the integral $\int_{\mathbb{R}} e^{-\mathrm{i}\alpha x^2}\,\mathrm{d}x$ with $\alpha > 0$.

as $R \to \infty$.

The case $\alpha < 0$ is treated analogously with the contour being the complex conjugate of the contour in Fig. 1.2.2 and the substitution in $\int_{C_3}$ being $r = ze^{-\mathrm{i}\pi/4}$.

In summary we get

$$\int_{\mathbb{R}} e^{-\mathrm{i}\alpha x^2}\,\mathrm{d}x = e^{-\mathrm{i}\,\mathrm{sign}(\alpha)\frac{\pi}{4}} \left(\frac{\pi}{|\alpha|}\right)^{1/2}, \quad \alpha \in \mathbb{R}\backslash\{0\}\}. \tag{1.20}$$

Setting now $\alpha = W''(\tilde{k})t/2$ and using $\chi(\tilde{k}) = W(\tilde{k}) - \tilde{k}x/t$ in (1.19), we have

Theorem 1.9. *Consider* (1.14) *with* $m = n = 1$ *and with initial data* $u(\cdot, 0) = u_0 \in S(\mathbb{R})$. *Choose the velocity* $c \in \mathbb{R}$, *assume that the equation is dispersive and that* $W''(\tilde{k}) \neq 0, \widehat{u}_0(\tilde{k}) \neq 0$ *for all points* $\tilde{k}$ *such that* $W'(\tilde{k}) = c$. *The solution then satisfies*

$$u(x,t) \sim \sum_{W'(\tilde{k})=c} \left(\frac{1}{t|W''(\tilde{k})|}\right)^{1/2} \widehat{u}_0(\tilde{k}) e^{-\mathrm{i}\,\mathrm{sign}(W''(\tilde{k}))\frac{\pi}{4}} e^{\mathrm{i}(\tilde{k}x - W(\tilde{k})t)} \tag{1.21}$$

as $t \to \infty, \frac{x}{t} = c$.

Note that the L^1 condition on u_0 ensures (via Riemann-Lebesgue Lemma A.8) that $\widehat{u}_0$ is continuous so that the point values $\widehat{u}_0(\tilde{k})$ make sense. If $W''(\tilde{k}) = 0, W'''(\tilde{k}) \neq 0$, then a similar calculation leads to an expansion in which $u(x,t)$ behaves like $t^{-1/3}$ as $t \to \infty$.

There is also an analogous result in $\mathbb{R}^n$. Using the stationary phase approximation from Theorem B.8, we get

Theorem 1.10. *Consider* (1.13) *with* $m = 1, n \in \mathbb{N}$ *and with initial data* $u(\cdot, 0) = u_0 \in S(\mathbb{R}^n)$. *Choose the velocity* $\vec{c} \in \mathbb{R}^n$, *assume that the equation is dispersive and that* $\det(D^2W(\tilde{k})) \neq 0, \widehat{u}_0(\tilde{k}) \neq 0$ *for all points* $\tilde{k} \in \mathbb{R}^n$ *such that* $\nabla W(\tilde{k}) = \vec{c}$. *The solution then satisfies*

$$u(x,t) \sim \frac{1}{t^{n/2}} \sum_{\nabla W(\tilde{k})=\vec{c}} \left(\frac{1}{\left|\det\left(D^2W(\tilde{k})\right)\right|}\right)^{1/2} \widehat{u}_0(\tilde{k}) e^{-\mathrm{i}\sigma(\tilde{k})\frac{\pi}{4}} e^{\mathrm{i}(\tilde{k}\cdot x - W(\tilde{k})t)} \tag{1.22}$$

as $t \to \infty, \frac{x}{t} = \vec{c}$, *where* $\sigma(\tilde{k})$ *is the signature of* $D^2W(\tilde{k})$, *i.e. the number of positive eigenvalues minus the number of negative eigenvalues.*

These results can be interpreted as follows. Asymptotically for large times the largest contribution to the solution at $x = \vec{c}t$ comes from those wavenumbers k, for which $v_g(k) = \vec{c}$. In other words, an observer moving at the group velocity $v_g(k)$ sees waves with the wavenumber k and frequency $W(k)$ but peaks generally keep passing by. On the other hand, an observer moving at the phase velocity, i.e. an observer at $x = v_p(k)t$ sees the same crest but the local wavenumber, and hence the distance to the the next crest, as well as the local frequency generally keep changing.

Let us now study the dispersion relations and group velocities for several simple examples including the three examples at the beginning of Section 1.2.

Example 1.3. For the Klein-Gordon equation (1.11) we have the dispersion relation

$$\omega^2 = k^2 + a.$$

Its solutions are $\omega = \pm W(k)$ with $W(k) = \sqrt{a + k^2}$. Clearly $W'' \not\equiv 0$ so that the problem is dispersive. The absolute value of the group velocity

$$v_g(k) = \frac{k}{\sqrt{a + k^2}}$$

grows in $|k|$ so that shorter waves propagate faster than longer ones. This means that initially localized pulses will have disintegrated in time in such a fashion that the more oscillatory waves will have traveled further, see Fig. 1.3. Also, the group velocity is bounded, $|v_g(k)| < 1$ for all $k \in \mathbb{R}$. We will see in Theorem 1.14

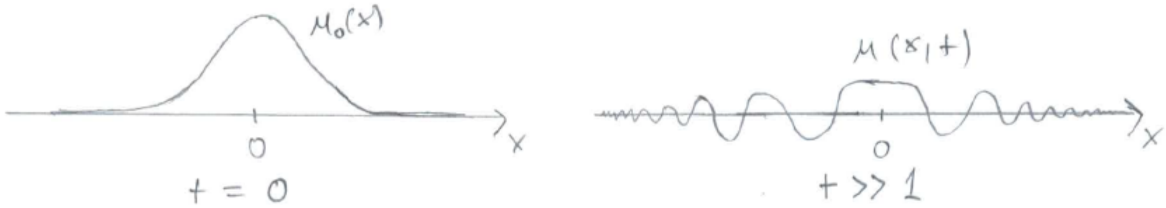

Figure 1.6: Schematic of the dispersion effect on a localized pulse in the Klein-Gordon equation.

that the Klein-Gordon equation enjoys finite speed of propagation. This is caused by the boundedness of the group velocity.

Example 1.4. For the Schrödinger equation (1.9) in $\mathbb{R}^n$ the dispersion relation reads

$$\omega = |k|^2,$$

so that $W(k) = |k|^2$. Here $\det D^2 W(k) = 2^n \neq 0$ and the problem is dispersive. The group velocity

$$v_g(k) = 2k$$

is unbounded (in modulus), and again shorter waves propagate faster than longer ones. The unboundedness of v_g causes an infinite speed of propagation as shown in Theorem 1.13.

Example 1.5. The linear KdV equation (1.10) has the dispersion relation

$$\omega = ak - k^3.$$

Hence $W''(k) = -6k \not\equiv 0$. The group velocity

$$v_g(k) = a - 3k^2$$

is bounded from above by a but unbounded from below. Also for the KdV Theorem 1.13 guarantees infinite speed of propagation. The maximal positive velocity is attained at $k = 0$ and

$$v_g(k) > 0 \quad \text{for} \quad |k| < (a/3)^{1/2}, \qquad v_g(k) < 0 \quad \text{for} \quad |k| > (a/3)^{1/2}.$$

If the initial data is a pulse with $\widehat{u}_0(k)$ localized near $k = 0$, then the main part of the pulse propagates to the right at velocity $v_g(0) = a$ and shorter waves trail behind or even propagate backwards.

Example 1.6. Let us show that the standard wave equation

$$\partial_t^2 u = c^2 \Delta u, \qquad x \in \mathbb{R}^n$$

is not dispersive. For $n = 1$ this is obvious as the dispersion relation $\omega^2 = c^2 k^2$ has the linear solutions $W(k) = \pm ck$. In general we have

$$\omega^2 = c^2 |k|^2$$

with the Euclidean norm $|k^2| = k_1^2 + \cdots + k_n^2$ and solutions $\omega = \pm W(k) = \pm c|k|$. Hence the Hessian entries are $(D^2W(k))_{i,j} = \delta_{i,j}\frac{1}{|k|} - \frac{k_i k_j}{|k|^3}$, i.e.

$$D^2W(k) = \frac{1}{|k|}\left(I - \frac{1}{|k|^2} k k^T\right).$$

Clearly $D^2W(k)k = 0$ and thus $\det(D^2W) \equiv 0$. In fact, as one easily checks, $P(k) := I - \frac{1}{|k|^2} k k^T$ is a projection onto the $(n-1)$-dimensional subspace orthogonal to k.

1.2.3 Local Wavenumber and Local Frequency

As we are about to show, the concepts of wavenumber, frequency and phase can be generalized from truly periodic functions, where they describe the function globally, to functions that are locally close to periodic. We will refer to such functions as *local wave trains.* One such example is the solution of a dispersive problem after a long time. The main ideas of this discussion come from Section 11.4 in [26].

Let us first consider the long time asymptotics of a dispersive scalar equation in 1D (i.e. $m = n = 1$ in Sec. 1.2.1) and assume

$$W \in C^2((0,\infty)),\ W'(k) > 0,\ W''(k) \neq 0 \quad \text{for all} \quad k > 0. \tag{1.23}$$

An example is $W(k) = k^2$. In Theorem 1.9 the equation

$$W'(k) = \frac{x}{t} \tag{1.24}$$

determines the dominant wavenumber k. Let us assume that $\frac{x}{t} \in W'((0,\infty))$. Then (1.24) defines for each point (x,t) one solution $k(x,t)$, which we call the **local wavenumber**. The dispersion relation $\omega = W(k)$ then defines the **local frequency** $\omega(x,t)$. We also define the **local phase**

$$\theta(x,t) = xk(x,t) - t\omega(x,t). \tag{1.25}$$

With these definitions the asymptotic form (1.21) for $t \to \infty$ becomes

$$u(x,t) \sim A(k(x,t),t)e^{\mathrm{i}\theta(x,t)}, \quad \text{where} \quad A(k,t) = \left(\frac{1}{t|W''(k)|}\right)^{1/2} \widehat{u}_0(k) e^{-\mathrm{i}\,\mathrm{sign}(W''(k))\frac{\pi}{4}} \tag{1.26}$$

as $t \to \infty$.

On the other hand, if we are given an expression

$$B(x,t)e^{\mathrm{i}\theta(x,t)} \tag{1.27}$$

with some functions B and θ, we can define the local wave number and local frequency as

$$k(x,t) := \partial_x\theta(x,t), \qquad \omega(x,t) = -\partial_t\theta(x,t). \tag{1.28}$$

These formulas, of course, determine what we intuitively understand under wave number and frequency only if $B, \partial_x\theta$ and $\partial_t\theta$ vary slowly in x and t. Under these conditions we call (1.27) a **local wave train**. Roughly speaking, a local wave train is a function which locally looks periodic. Note that because no asymptotic parameter has been specified in the above slowness condition for $B, \partial_x\theta$ and $\partial_t\theta$, a local wave train is not a rigorous mathematical object.

Let us first check that the above two definitions of local wavenumber and frequency coincide. Indeed, from (1.25) and $W'(k) = x/t, \omega = W(k)$ we get

$$\partial_x\theta = k + [x - W'(k)t]\partial_x k = k \tag{1.29}$$
$$\partial_t\theta = -\omega + [x - W'(k)t]\partial_t k = -\omega. \tag{1.30}$$

Next we show that the asymptotic approximation (1.26) of the solution of a dispersive problem is a local wave train. Under the assumptions (1.23) this holds, for instance, for x in compact sets $\frac{x}{t} \in W'([\delta, R))$ with arbitrary fixed $0 < \delta < R$. From $W'(k) = x/t$ we get

$$W''(k)\partial_x k = \tfrac{1}{t}$$

so that

$$\partial_x k = \tfrac{1}{tW''(k)} = O(t^{-1}) \quad (t \to \infty).$$

Similarly

$$W''(k)\partial_t k = -\tfrac{x}{t^2} = -W'(k)\tfrac{1}{t}$$

so that

$$\partial_t k = -\frac{W'(k)}{tW''(k)} = O(t^{-1}) \quad (t \to \infty).$$

Here we have used the boundedness of W' and the boundedness of W'' away from zero. These follow because $W \in C^2, W''(k) \neq 0$ for $k > 0$ and because $\frac{x}{t} \in W'([\delta, R))$ implies that the solution of (1.24) satisfies $k \in [\delta, R]$.

From the dispersion relation we get $\omega(x,t) = W(k(x,t))$ and again, the boundedness of $W'(k)$ for $k \in [\delta, R]$ gives

$$\partial_x\omega = O(t^{-1}) \quad \text{and} \quad \partial_t\omega = O(t^{-1}).$$

Finally, for the amplitude A we first write $A(k,t) = t^{-1/2}\tilde{A}(k)$. Hence

$$\frac{d}{dx}A = O(t^{-1/2})\tilde{A}'\partial_x k = O(t^{-3/2}) \tag{1.31}$$
$$\frac{d}{dt}A = O(t^{-3/2})\tilde{A} + O(t^{-1/2})\tilde{A}'\partial_t k = O(t^{-3/2}). \tag{1.32}$$

1.2.3.1 Phase Velocity and Group Velocity for Local Wave Trains

Let us consider a local wave train (1.27) with the local wave number and frequency in (1.28). First, by cross differentiation we get

$$\partial_t k + \partial_x\omega = 0.$$

If the local wave train describes a solution of a PDE with the dispersion relation $\omega = W(k)$, then $\partial_x \omega = W'(k)\partial_x k$ and k satisfies

$$\partial_t k + W'(k)\partial_x k = 0. \tag{1.33}$$

Equation (1.33) is a nonlinear hyperbolic equation if the original PDE is dispersive, i.e. if $W'' \not\equiv 0$. Interestingly enough, nonlinear hyperbolic theory thus plays a role in linear dispersive PDEs. Equation (1.33) reveals the fact which we observed in Theorem 1.9 for the asymptotics of dispersive PDEs, i.e. that k propagates with the group velocity $W'(k)$. The meaning of k has now been, however, generalized compared to Theorem 1.9 and we conclude that for local wave trains the local wavenumber $k(x,t)$ propagates at the velocity $W'(k(x,t))$.

To look at the phase velocity, we choose a phase value $\theta_0 \in \mathbb{R}$. The wave front $\mathcal{F} = \{x \in \mathbb{R} : \theta(x,t) = \theta_0\}$ satisfies

$$\partial_x \theta \frac{dx}{dt} + \partial_t \theta = 0$$

and so the front propagates at the velocity

$$\frac{dx}{dt} = -\frac{\partial_t \theta}{\partial_x \theta} = \frac{\omega(x,t)}{k(x,t)},$$

which agrees with our previous definition of phase velocity except that k and ω have been generalized.

The nature of the dynamics of local wave trains can be very well visualized with the help of so called group lines and phase lines. A **group line** is a level set (in the (x,t) plane) of the local wave number $k(x,t)$. A **phase line** is the level set of the local phase $\theta(x,t)$.

Let us consider the case of a smooth $W(k)$ and

$$\begin{aligned} &v_g(k) > v_p(k) > 0, v_g'(k) > 0, v_p'(k) > 0 \quad \text{for all } k > 0 \\ &v_g(0) = v_p(0) = 0,\ v_g(k), v_p(k) \to \infty \quad \text{for } k \to \infty. \end{aligned} \tag{1.34}$$

Then the group lines $x = v_g(k)t$ and and the phase lines $\frac{dx}{dt} = v_p(k)$ look qualitatively like in Fig. 1.2.3.1. The group lines are always straight. The phase lines cannot be straight because the condition $v_p(0) = 0$ would

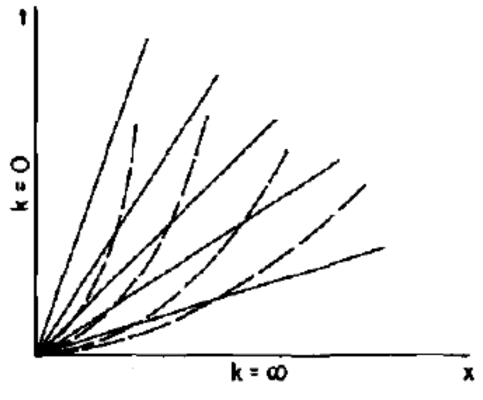

Figure 1.7: (from [26]) Group lines (full) and phase lines (dashed) for the case (1.34).

mean that group and phase lines coincide which is impossible by $v_g(k) > v_p(k)$. For the case $v_g(k) > v_p(k)$ a phase line thus has at each (x,t) a larger slope than the group line going through this point. An observer moving with the group velocity $v_g(k_0)$ sees the local wavenumber k_0 and keeps overtaking crests.

Example 1.7. An example for the case (1.34) is $W(k) = k^2$ such that $v_g(k) = 2k > k = v_p(k)$ for $k > 0$ and

$$k(x,t) = \frac{x}{2t}, \qquad \omega(x,t) = W(k(x,t)) = \left(\frac{x}{2t}\right)^2, \qquad \theta(x,t) = xk(x,t) - t\omega(x,t) = \frac{x^2}{4t}.$$

The group lines are thus $\frac{x}{t}$ =const. and the phase lines are $\frac{x^2}{t}$ =const.

1.2.4 Energy Propagation, Infinite Speed of Propagation

In this section we show that asymptotically for large times the group velocity in dispersive problems is also the velocity of propagation of the L^2-energy. First of all let us note that for dispersive problems the L^2-norm is conserved in the time evolution.

Theorem 1.11 (Conservation of the L^2-norm)**.** *For dispersive equations* (1.13) *with* $m = 1$ *and with initial data* $u(\cdot, 0) = u_0 \in L^2(\mathbb{R}^n)$ *one has*

$$\|u(\cdot,t)\|_{L^2(\mathbb{R}^n)} = \|u_0\|_{L^2(\mathbb{R}^n)} \qquad \textit{for all} \quad t > 0.$$

Proof. Using Plancherel's identity and the formulation of the solution in Fourier space $\widehat{u}(k,t) = \widehat{u}_0(k)e^{-\mathrm{i}W(k)t}$, we have

$$\|u(\cdot,t)\|_{L^2(\mathbb{R}^n)} = \|\widehat{u}(\cdot,t)\|_{L^2(\mathbb{R}^n)} = \|\widehat{u}_0 e^{-\mathrm{i}W(\cdot)t}\|_{L^2(\mathbb{R}^n)} = \|\widehat{u}_0\|_{L^2(\mathbb{R}^n)} = \|u_0\|_{L^2(\mathbb{R}^n)}.$$

□

The case $m > 1$ is left as an exercise. Here, in general, a different energy than the square of the L^2-norm is conserved. Of course, in Theorem 1.11 it is not necessary that the problem is dispersive. Namely, the condition $W'' \not\equiv 0$ can be dropped.

A question which we now pose is how the solution spreads in terms of the L^2-norm for large times. Let us consider the 1D scalar case $m = n = 1$ with a smooth W let $x_1 < x_2 \in \mathbb{R}$ and define

$$Q(t) := \int_{x_1}^{x_2} |u(x,t)|^2 \, \mathrm{d}x.$$

Assuming that for each $x \in [x_1, x_2]$ the equation $W'(\tilde{k}) = \frac{x}{t}$ has a unique solution $\tilde{k}$, let

$$\tilde{k}_{1,2} = (W')^{-1}(\tfrac{x_{1,2}}{t}).$$

If, in addition $W''(k) \neq 0$ for all $k \in [\tilde{k}_1, \tilde{k}_2]$, then the asymptotics (1.21) produce

$$Q(t) \sim \frac{1}{t} \int_{x_1}^{x_2} \frac{|\widehat{u}_0(\tilde{k}(x))|^2}{|W''(\tilde{k}(x))|} \, \mathrm{d}x \qquad (t \to \infty).$$

With the substitution $\tilde{k} = \tilde{k}(x)$ we get $\mathrm{d}x = W''(\tilde{k})t \, \mathrm{d}\tilde{k}$ and

$$Q(t) \sim \int_{\tilde{k}_1}^{\tilde{k}_2} \mathrm{sign}(W''(\tilde{k}))|\widehat{u}_0(\tilde{k})|^2 \, \mathrm{d}\tilde{k} = \int_{\min\{\tilde{k}_1,\tilde{k}_2\}}^{\max\{\tilde{k}_1,\tilde{k}_2\}} |\widehat{u}_0(k)|^2 \, \mathrm{d}k \qquad (t \to \infty).$$

If $\tilde{k}_1, \tilde{k}_2$ are $t-$independent, then $Q'(t) \sim 0$. This happens if we let $x_1 = x_1(t) = W'(\tilde{k}_1)t$ and $x_2 = x_2(t) = W'(\tilde{k}_2)t$. In summary we have that the L^2-norm stays asymptotically constant between any two group lines.

Theorem 1.12. *Consider* (1.14) *with* $m = n = 1$ *under the assumption of dispersivity and with initial data* $u_0 \in S(\mathbb{R})$*. Let* $k_1 < k_2 \in \mathbb{R}$ *and* $W''(k) > 0$ *for all* $k \in \mathbb{R}$*. Asymptotically for* $t \to \infty$ *the* L^2-*norm stays constant between the two group lines* $x = W'(k_1)t$ *and* $x = W'(k_2)t$ *and*

$$\int_{W'(k_1)t}^{W'(k_2)t} |u(x,t)|^2 \, \mathrm{d}x \sim \int_{k_1}^{k_2} |\widehat{u}_0(k)|^2 \, \mathrm{d}k \qquad (t \to \infty).$$

Note that this result holds also for $m > 1$.

The above result suggests that if in 1D the group velocity $W'(k)$ is unbounded, then the energy can be carried arbitrarily far within any finite time t. Let us assume that $W'(k)$ is unbounded from above. Then this is because for a given $t > 0$ and an interval $[x_1, x_2]$ (with arbitrarily large x_1, x_2) we can find k_1 and k_2 such that $x_1 = W'(k_1)t$ and $x_2 = W'(k_2)t$. Choosing then u_0 such that $\int_{k_1}^{k_2} |\widehat{u}_0(k)|^2 \, \mathrm{d}k > 0$, we have $\int_{x_1}^{x_2} |u(x,t)|^2 \, \mathrm{d}x > 0$.

This result is, however, only asymptotic for large times. Hence, we cannot immediately conclude an infinite speed of propagation. This will be proved next for arbitrary dimension n using the Paley-Wiener theorem.

Theorem 1.13. *(Infinite propagation speed in dispersive problems) Consider* (1.13) *with* $m = 1$ *and with an entire solution* W *of the dispersion relation. If there are* $c > 0, \theta \in (-\pi, 0) \cup (0, \pi), a > 1, j \in \{1, \dots, n\}$ *and* $\tilde{\alpha} = (\alpha_1, \dots, \alpha_{j-1}, \alpha_{j+1}, \dots, \alpha_n) \in \mathbb{R}^{n-1}$ *so that the dispersion relation satisfies*

$$\operatorname{Im}(W)(\alpha_1, \dots, \alpha_{j-1}, k_j, \alpha_{j+1}, \dots, \alpha_n) > c|\operatorname{Im}(k_j)|^a \quad \textit{for all} \quad k_j \in \mathbb{C}\backslash\{0\} \textit{ with } \quad \arg(k_j) = \theta,$$

then there exist initial data $u_0 \in L^\infty(\mathbb{R}^n)$ *with* $\operatorname{supp}(u_0)$ *compact so that the solution* u *satisfies*

$$\operatorname{supp}(u(\cdot, t)) \quad \textit{is unbounded in} \quad \mathbb{R}^n \textit{ for all } t > 0.$$

Remark 1.2. *As we will show in the proof,* u_0 *can be chosen, for instance, as the characteristic function of an* $n-$*dimensional interval centered at zero.*

Proof. The function u_0 constructed below satisfies $u_0 \in L^\infty(\mathbb{R}^n)$ with $\operatorname{supp}(u_0)$ compact so that $u_0 \in L^2(\mathbb{R}^n)$ and the solution can be written in Fourier space as

$$\widehat{u}(k,t) = e^{-\mathrm{i}W(k)t}\widehat{u}_0(k) \qquad \text{for all } k \in \mathbb{R}^n.$$

Due to the compact support of u_0 the Paley-Wiener theorem A.11 guarantees that the Fourier transform $\widehat{u}_0$ has an (analytic) extension to the complex $k \in \mathbb{C}^n$ and we can write

$$\widehat{u}(k,t) = e^{-\mathrm{i}\operatorname{Re}(W)(k_R + \mathrm{i}k_I)t} e^{\operatorname{Im}(W)(k_R + \mathrm{i}k_I)t} \widehat{u}_0(k_R + \mathrm{i}k_I) \qquad \text{for all } k = k_R + \mathrm{i}k_I \in \mathbb{C}^n.$$

For all $k = (\alpha_1, \dots, \alpha_{j-1}, k_j, \alpha_{j+1}, \dots, \alpha_n)$ with $\arg(k_j) = \theta$ we have

$$|\widehat{u}(k,t)| > e^{c_1|\operatorname{Im}(k)|^a t}|\widehat{u}_0(k)| \tag{1.35}$$

with some constant $c_1 > 0$, where $|\operatorname{Im}(k)|^2 = \sum_{j=1}^n \operatorname{Im}(k_j)^2$ and where we have used the inequality $c_2|\operatorname{Im}(k_j)| < |\operatorname{Im}(k)| < c_3|\operatorname{Im}(k_j)|$ for all $k_j \in \mathbb{C}$ and some $c_2, c_3 > 0$, which holds since all entries in k other than k_j are fixed.

Due to $a > 1$ in the estimate (1.35) the Paley-Wiener theorem produces the unbounded support of $u(\cdot, t)$. It remains, however, to show the existence of u_0 such that $\widehat{u}_0$ does not decay fast in the direction θ so that the growth $e^{c_1|\operatorname{Im}(k)|^a t}$ in (1.35) is not canceled out. We choose

$$u_0(x) = \chi_{[-1,1]}(x_j) f(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n)$$

so that

$$\widehat{u}_0(\alpha_1, \dots, \alpha_{j-1}, k_j, \alpha_{j+1}, \dots, \alpha_n) = \widehat{f}(\tilde{\alpha}) \left(\frac{2}{\pi}\right)^{1/2} \frac{\sin(k_j)}{k_j}.$$

Because $\sin(k_j)$ grows exponentially in $|k_j|$ in all directions $\arg(k_j)$ except $\arg(k_j) \in \{0, \pi\}$, the growth of $e^{c_1|\operatorname{Im}(k)|^a t}$ in (1.35) is not canceled out if $\widehat{f}(\tilde{\alpha}) \neq 0$. It remains to find $f \in L^\infty(\mathbb{R}^{n-1})$ with $\operatorname{supp}(f)$ compact and $\widehat{f}(\tilde{\alpha}) \neq 0$.

If for all $i \in I := \{1, \dots, j-1, j+1, \dots, n\}$ it is $\alpha_i \notin \pi\mathbb{Z}$, then we can choose

$$f(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n) = \Pi_{i \in I} \chi_{[-1,1]}(x_i)$$

so that

$$\widehat{f}(\tilde{\alpha}) = \left(\frac{2}{\pi}\right)^{\frac{n-1}{2}} \Pi_{i\in I} \frac{\sin(\alpha_i)}{\alpha_i} \neq 0.$$

If $\alpha_i \in \pi\mathbb{Z}$ for all $i \in I_0 \subset I$ and $\alpha_i \notin \pi\mathbb{Z}$ for all $i \in I\backslash I_0$, then we can set

$$f(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n) = \Pi_{i\in I_0} \chi_{[-1/2,1/2]}(x_i) \Pi_{i\in I\backslash I_0} \chi_{[-1,1]}(x_i)$$

so that

$$\widehat{f}(\tilde{\alpha}) = \left(\frac{2}{\pi}\right)^{\frac{n-1}{2}} \Pi_{i\in I_0} \frac{\sin(\frac{1}{2}\alpha_i)}{\alpha_i} \Pi_{i\in I\backslash I_0} \frac{\sin(\alpha_i)}{\alpha_i} \neq 0,$$

where the value of $\sin(ax)/x$ for $x = 0$ is defined as a via the limit. We have thus constructed u_0 such that

$$|\widehat{u}_0(\alpha_1, \dots, \alpha_{j-1}, k_j, \alpha_{j+1}, \dots, \alpha_n)| = |\widehat{f}(\tilde{\alpha})| \left(\frac{2}{\pi}\right)^{1/2} \frac{\sin(k_j)}{k_j} \to \infty$$

for $|k_j| \to \infty, \arg(k_j) = \theta$. The solution satisfies for $k = (\alpha_1, \dots, \alpha_{j-1}, k_j, \alpha_{j+1}, \dots, \alpha_n)$ with $\arg(k_j) = \theta$

$$|\widehat{u}(k,t)| > e^{c_1|\operatorname{Im}(k)|^a t} \quad \text{with } a > 1$$

and, as mentioned above, the Paley-Wiener theorem, implies an unbounded support of $u(\cdot, t)$. □

Remark 1.3. *The growth condition on* $\operatorname{Im}(W)$ *in Theorem 1.13 is satisfied by generic dispersion relations with an unbounded group velocity. For instance, in one dimension, $n = 1$, every polynomial $W(k) = \sum_{j=0}^{p} a_j k^j$ with $p \geqslant 2, (a_p \in \mathbb{C}\backslash\{0\})$ satisfies this condition. To see this note that $a_p k^p = |a_p||k|^p e^{\mathrm{i}(\arg(a_p)+p\theta)}$ if $\theta := \arg(k)$. For $\theta \neq 0, \pi$ it is $k_R = \cot(\theta) k_I$ and $a_p k^p = |a_p|(1+\cot^2(\theta))^{p/2} e^{\mathrm{i}(\arg(a_p)+p\theta)}|k_I|^p$. The growth condition holds if* $\sin(\arg(a_p) + p\theta) > 0$ *and $\theta \in (-\pi, 0) \cup (0, \pi)$. One can choose, e.g.*

$$\arg(a_p) + p\theta = \tfrac{\pi}{2}, \ \textit{i.e.}\ \theta = \tfrac{\pi}{2p} - \tfrac{\arg a_p}{p} \ \textit{if}\ \arg(a_p) \notin \{\tfrac{\pi}{2}, \pi(\tfrac{1}{2} \pm p)\},$$
$$\arg(a_p) + p\theta = \tfrac{\pi}{4}, \ \textit{i.e.}\ \theta = \tfrac{\pi}{4p} - \tfrac{\arg a_p}{p} \ \textit{if}\ \arg(a_p) \in \{\tfrac{\pi}{2}, \pi(\tfrac{1}{2} \pm p)\}.$$

Example 1.8. For the 1D Schrödinger equation $\mathrm{i}\partial_t u + \partial_x^2 u = 0$ the dispersion relation is $W(k) = k^2$. For $\arg(k) = \theta = \pi/4$ we get $W(k) = (k_I + \mathrm{i}k_I)^2$, such that $\operatorname{Im}(W)(k) = 2k_I^2$ and Theorem 1.13 applies.

For equations with bounded group velocities one expects, on the other hand, a finite speed of propagation. For the Klein-Gordon equation this can be proved quite easily using the Paley-Wiener theorem.

Theorem 1.14. *(Finite speed of propagation for the Klein-Gordon equation) Let $f, g \in L^2(\mathbb{R})$,* $\operatorname{supp}(f) \subset B_R$ *and* $\operatorname{supp}(g) \subset B_R$ *with some $R > 0$. Then the solution of the Klein-Gordon equation $\partial_t^2 u - \partial_x^2 u + au = 0, a > 0$ with $u(x,0) = f(x), \partial_t u(x,0) = g(x)$ satisfies*

$$\operatorname{supp}(u(\cdot,t)) \subset B_{R+t} \quad \textit{for all } t \geqslant 0.$$

Proof. In the Fourier variables the problem reads $\frac{d^2}{dt^2}\widehat{u} = -(a+k^2)\widehat{u}$, $\widehat{u}(k,0) = \widehat{f}(k)$, $\frac{d}{dt}\widehat{u}(k,0) = \widehat{g}(k)$. The solution is

$$\widehat{u}(k,t) = \widehat{f}(k)\cos(t\sqrt{a+k^2}) + \widehat{g}(k)\frac{\sin(t\sqrt{a+k^2})}{\sqrt{a+k^2}}.$$

First we show the analyticity of $\widehat{u}(\cdot,t)$. The functions $\widehat{f}(k)$ and $\widehat{g}(k)$ are entire due to the assumption of compact support of f and g and the Paley-Wiener theorem. The inner function $k \mapsto a + k^2$ is clearly entire.

For $z \mapsto \cos(\sqrt{z})$ and $z \mapsto \frac{\sin\sqrt{z}}{\sqrt{z}}$ note that $\sqrt{z} := \sqrt{|z|}e^{\mathrm{i}\arg z/2}$ for $\arg(z) \in (-\pi, \pi]$ is analytic everywhere except along the ray $\arg(z) = \pi$ where it is discontinuous. For z_0 with $\arg(z_0) = \pi$ the two distinct limits of $\sqrt{z}$ for $z \to z_0$ are $\sqrt{z_0}$ and $-\sqrt{z_0}$ and because the functions $z \mapsto \cos(z)$ and $z \mapsto \frac{\sin(z)}{(z)}$ are even, the compositions $z \mapsto \cos(\sqrt{z})$ and $z \mapsto \frac{\sin\sqrt{z}}{\sqrt{z}}$ are entire.

Next we check the growth condition on $|\widehat{u}(k,t)|$, namely we need to show

$$|\widehat{u}(k,t)| \leqslant c(t)(1+|k|)^N e^{(R+t)|\operatorname{Im}(k)|} \quad \text{for some } N \in \mathbb{N}, c(t) > 0 \text{ and all } k \in \mathbb{C}. \tag{1.36}$$

Let $z = u + \mathrm{i}v$. For $|\cos(u+\mathrm{i}v)|$ we have

$$|\cos(u+\mathrm{i}v)| = \frac{1}{2}|e^{\mathrm{i}(u+\mathrm{i}v)} + e^{-\mathrm{i}(u+\mathrm{i}v)}| \leqslant \frac{1}{2}(e^{-v} + e^{v}) \leqslant e^{|v|}.$$

Let us now look at $\frac{|\sin(u+\mathrm{i}v)|}{|u+\mathrm{i}v|}$. The case $|u+\mathrm{i}v| \geqslant 1$ is analogous since $\frac{|\sin(u+\mathrm{i}v)|}{|u+\mathrm{i}v|} \leqslant |\sin(u+\mathrm{i}v)| \leqslant e^{|v|}$. The case $|u+\mathrm{i}v| < 1$ needs a little more care. Using the identity $\sin(u+\mathrm{i}v) = \sin(u)\cosh(v) + \mathrm{i}\cos(u)\sinh(v)$ we get

$$\left|\frac{\sin(u+\mathrm{i}v)}{u+\mathrm{i}v}\right|^2 \leqslant \left|\frac{\sin(u)}{u}\right|^2 \cosh^2(v) + \cos^2(u)\left|\frac{\sinh(v)}{v}\right|^2 \leqslant 2\cosh^2(v) \leqslant 2e^{2|v|},$$

where in the second inequality one uses $|\sinh(v)| \leqslant |v|\cosh(v)$. Setting

$$u + \mathrm{i}v := t\sqrt{a+k^2}, \tag{1.37}$$

we have

$$|\cos(t\sqrt{a+k^2})| \leqslant e^{|v|}, \qquad \left|\frac{\sin(t\sqrt{a+k^2})}{\sqrt{a+k^2}}\right| \leqslant t\sqrt{2}e^{|v|}.$$

Writing $k = k_R + \mathrm{i}k_I$, it remains to show that with the definition 1.37

$$|v| \leqslant t|k_I|.$$

From $t^2(a+k^2) = (u+\mathrm{i}v)^2$ we have $t^2(a+k_R^2-k_I^2) = u^2 - v^2$ and $t^2 k_R k_I = uv$, which can be combined into $v^2(t^2(a+k_R^2-k_I^2)+v^2) = t^4k_R^2k_I^2$ and rewritten as

$$\left(v^2 + t^2\frac{a+k_R^2-k_I^2}{2}\right)^2 = t^4\left(\frac{a+k_R^2+k_I^2}{2}\right)^2 - t^4ak_I^2 \leqslant t^4\left(\frac{a+k_R^2+k_I^2}{2}\right)^2,$$

where the inequality holds due to $a > 0$. This implies $v^2 \leqslant t^2k_I^2$ and thus $|v| \leqslant t|k_I|$.

For $\widehat{f}(k)$ and $\widehat{g}(k)$ The Paley-Wiener theorem implies

$$|\widehat{f}(k)|, |\widehat{g}(k)| \leqslant C(1+|k|)^N e^{R|k_I|} \quad \text{for some } C > 0, N \in \mathbb{N} \text{ and all } k \in \mathbb{C}.$$

In summary we get (1.36) with $c(t) = C\sqrt{2}(1+t)$. □

1.3 Smoothing Effects of Dispersion: Schrödinger Equation

Similarly to parabolic equations certain dispersive equations possess a smoothing property. A singularity can be qualitatively described by the presence of large local wavenumbers at the singularity location. While in parabolic equations smoothing happens via a strong damping of large wavenumbers, in dispersive problems energy is conserved and smoothing occurs due to the different velocities (group velocities) of distinct wavenumbers. The heuristic idea is that a singularity is characterized by a wide range of local wavenumbers - in particular by the presence of many large wavenumbers. Due to their different velocities in dispersive problems the singularity vanishes. For problems where the group velocity diverges to infinity for $|k| \to \infty$ the smoothing is instantaneous.

Here we restrict our attention to the Schrödinger equation

$$\begin{aligned} \mathrm{i}\partial_t u + \Delta u =&0, \qquad x \in \mathbb{R}^n, t > 0 \\ u(\cdot, 0) =&u_0 \in L^2(\mathbb{R}^n) \end{aligned} \tag{1.38}$$

and present some classical rigorous results on its smoothing properties. In particular, we show below the *local smoothing* result

$$u_0 \in L^2(\mathbb{R}^n) \;\Rightarrow\; u(\cdot, t) \in H^{1/2}_{\mathrm{loc}}(\mathbb{R}^n) \text{ for almost all } t > 0$$

as well as the *global smoothing* result

$$u_0 \in L^{p'}(\mathbb{R}^n), p' \in [1, 2], \frac{1}{p} + \frac{1}{p'} = 1 \;\Rightarrow\; u(\cdot, t) \in L^p(\mathbb{R}^n) \text{ for all } t > 0.$$

Let us denote the solution operator of the Cauchy problem for the Schrödinger equation by

$$e^{\mathrm{i}t\Delta} : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n), \quad u(\cdot, t) = e^{\mathrm{i}t\Delta} u_0.$$

It is the operator which produces the solution $u(\cdot, t)$ for given initial data u_0.

As we have seen in Sec. 1.2.1, the solution of (1.38) can be written very succinctly using the Fourier transform

$$\widehat{u}(k, t) = e^{-\mathrm{i}|k|^2 t} \widehat{u}_0(k). \tag{1.39}$$

Therefore, if $u_0 \in S(\mathbb{R}^n)$, Lemma A.7 can be used to get the solution formula in physical space: $u(x, t) = (2\pi)^{-n/2} \left(\left(e^{-\mathrm{i}|\cdot|^2 t}\right)^{\vee} * u_0 \right)(x)$. A direct calculation using (1.20) produces for $u_0 \in S(\mathbb{R}^n)$

$$u(x, t) = e^{\mathrm{i}t\Delta} u_0(x) = \left(\frac{1}{4\pi \mathrm{i}t}\right)^{n/2} \int_{\mathbb{R}^n} e^{\mathrm{i}\frac{|x-y|^2}{4t}} u_0(y) dy. \tag{1.40}$$

This can be compared with the solution of the Cauchy problem for the heat equation in $\mathbb{R}^n$, where $u(x, t) = \left(\frac{1}{4\pi t}\right)^{n/2} \int_{\mathbb{R}^n} e^{-\frac{|x-y|^2}{4t}} u_0(y) dy$. Hence, formally, the solution of the Schrödinger equation can be obtained from the solution of the heat equation by replacing $t \rightsquigarrow it$. Note that the expression on the right hand side of (1.40) makes sense also for $u_0 \in L^1(\mathbb{R}^n)$ but then it may generally not be the classical solution of (1.38). We, however, have

Theorem 1.15. *If $u_0 \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ and $\int_{\mathbb{R}^n} |x|^m |u_0(x)| dx < \infty$ for some $m \geqslant 2$, then*

$$u(x, t) = \left(\frac{1}{4\pi \mathrm{i}t}\right)^{n/2} \int_{\mathbb{R}^n} e^{\mathrm{i}\frac{|x-y|^2}{4t}} u_0(y) dy$$

solves $\mathrm{i}\partial_t u + \Delta u = 0$ in the classical sense and $u(\cdot, t) \in C^m(\mathbb{R}^n)$ for all $t \neq 0$. If, in addition, $u_0 \in C^1_c(\mathbb{R}^n)$, then $u(x, t) \to u_0(x)$ as $t \to 0$ for each $x \in \mathbb{R}^n$.

Proof. The condition $u_0 \in L^1(\mathbb{R}^n)$ ensures that the integral formula (1.40) makes sense. Since

$$\int_{\mathbb{R}^n} |x-y|^p |u_0(y)| dy < \infty$$

for all $p \leqslant m$, the Lebesgue dominated convergence theorem implies that (1.40) can be differentiated m times in x under the integral sign. By the same argument the integral can also be differentiated in t under the integral sign. A direct calculation then produces $\mathrm{i}\partial_t u + \Delta u = 0$. The Lebesgue dominated convergence theorem also guarantees continuity of all the derivatives up to order m due to the continuity of polynomials in $x - y$ and of $e^{\mathrm{i}\frac{|x-y|^2}{4t}}$.

The initial condition $u(x,t) \to u_0(x)$ follows directly from the $n-$dimensional version of the stationary phase approximation in Theorem B.8. □

Theorem 1.15 shows that decay of initial data translates into smoothness of the solution. This smoothing can be heuristically understood based on dispersion. Because planewaves with large k travel faster for the Schrödinger equation, the large local wavenumbers at singularities travel fast away to infinity. At the same time, large wavenumbers may come from infinity and spoil the smoothness of u. When, however, the solution is initially well localized, less can come from infinity.

Distributional Solution of (1.38) We call $u \in S'(\mathbb{R}^{n+1})$ a distributional solution of the differential equation in (1.38) if

$$-\mathrm{i}u(\partial_t \varphi) + u(\Delta\varphi) = 0 \quad \text{for all } \varphi \in S(\mathbb{R}^{n+1}).$$

It is left as an exercise to show that for $u_0 \in L^2(\mathbb{R}^n)$ the Fourier representation (1.39) generates via

$$u(x,t) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{\mathrm{i}k\cdot x} e^{-\mathrm{i}|k|^2 t} \hat{u_0}(k) dk \tag{1.41}$$

a distributional solution and that the initial data are satisfied in the L^2-sense, i.e. $\|u(\cdot,0) - u_0(\cdot)\|_{L^2(\mathbb{R}^n)} = 0$.

1.3.0.1 Properties of the Solution Operator $e^{\mathrm{i}t\Delta}$

We investigate now some properties of the solution operator $e^{\mathrm{i}t\Delta} : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n), u_0 \mapsto u$ defined by (1.41). The proof of the following properties is left as an exercise.

1. Isometry
$$\|e^{\mathrm{i}t\Delta} f\|_{L^2(\mathbb{R}^n)} = \|f\|_{L^2(\mathbb{R}^n)} \quad \text{for all } f \in L^2(\mathbb{R}^n),$$

2. $$e^{\mathrm{i}t\Delta} e^{\mathrm{i}s\Delta} = e^{\mathrm{i}(t+s)\Delta} \quad \text{and} \quad \left(e^{\mathrm{i}t\Delta}\right)^{-1} = e^{-\mathrm{i}t\Delta} \quad \text{for all } t,s \in \mathbb{R},$$

3. $$e^{\mathrm{i}0\Delta} = \mathrm{Id},$$

4. for a fixed $f \in L^2(\mathbb{R}^n)$ the map $\Phi_f : \mathbb{R} \to L^2(\mathbb{R}^n)$, defined by
$$\Phi_f : t \mapsto \Phi_f(t) = e^{\mathrm{i}t\Delta} f,$$
is continuous, i.e. it describes a curve in $L^2(\mathbb{R}^n)$.

Remark 1.4. *Any operator family $\{T_t\}_{t\in\mathbb{R}}$ which satisfies the properties (2-4) is called a one parameter operator group. If $\{T_t\}_{t\in\mathbb{R}}$ satisfies only $T_{t+s} = T_t \circ T_s$ for all $t,s \geqslant 0$, $T_0 = Id$, and $T_t f : \mathbb{R} \to L^2(\mathbb{R}^n)$ continuous for any $f \in L^2(\mathbb{R}^n)$, then it is called a one parameter operator semigroup. For example, the solution operator of the heat equation forms a semigroup.*

As the next lemma shows, $e^{\mathrm{i}t\Delta}$ is an isometry not only in $L^2(\mathbb{R}^n)$ but also in $H^s(\mathbb{R}^n)$.

Lemma 1.16. *The operator $e^{\mathrm{i}t\Delta}$ is an isometry in $H^s(\mathbb{R}^n), s > 0$, i.e. $\|e^{\mathrm{i}t\Delta}f\|_{H^s(\mathbb{R}^n)} = \|f\|_{H^s(\mathbb{R}^n)}$ for all $f \in H^s(\mathbb{R}^n)$.*

Proof.

$$\|e^{\mathrm{i}t\Delta}f\|_{H^s} = \|(1+|k|^s)e^{-\mathrm{i}|k|^2t}\widehat{f}\|_{L^2} = \|(1+|k|^s)\widehat{f}\|_{L^2} = \|f\|_{H^s}.$$

□

This simple lemma has the important consequence that the Schrödinger equation flow *does not possess global smoothing in the sense of* $H^s(\mathbb{R}^n)$, i.e. if $u_0 \in H^s(\mathbb{R}^n)$ and $u_0 \notin H^r(\mathbb{R}^n)$ for some $r > s$, then $e^{\mathrm{i}t\Delta}u_0 \in H^s(\mathbb{R}^n)$ and $e^{\mathrm{i}t\Delta}u_0 \notin H^r(\mathbb{R}^n)$.

1.3.0.2 Local Smoothing of the Schrödinger Equation

Here we prove the already advertised result

$$u_0 \in L^2(\mathbb{R}^n) \;\Rightarrow\; u(\cdot,t) \in H^{1/2}_{\mathrm{loc}}(\mathbb{R}^n) \text{ for almost all } t > 0.$$

The proof comes essentially from [14]. The first step is the following

Theorem 1.17. *There is a constant $c > 0$ such that for $n = 1$*

$$\sup_{x\in\mathbb{R}} \int_{-\infty}^{\infty} |D_x^{1/2} e^{\mathrm{i}t\Delta} f(x)|^2 dt \leqslant c\|f\|^2_{L^2(\mathbb{R}^n)}, \tag{1.42}$$

and for $n \geqslant 2$ and every $j \in \{1,\dots,n\}$

$$\sup_{x_j\in\mathbb{R}} \int_{\mathbb{R}^n} |D_{x_j}^{1/2} e^{\mathrm{i}t\Delta} f(x)|^2 d_{x_1}\dots d_{x_{j-1}} d_{x_{j+1}}\dots d_{x_n} dt \leqslant c\|f\|^2_{L^2(\mathbb{R}^n)}, \tag{1.43}$$

where $D_{x_j}^{1/2} g(x,t) := \big(|k_j|^{1/2}\widehat{g}(k,t)\big)^{\vee}(x,t)$.

Proof. For $n = 1$ we write

$$D_x^{1/2} e^{\mathrm{i}t\Delta} f = \Big(|k|^{1/2} e^{-\mathrm{i}k^2t}\widehat{f}\Big)^{\vee} = \underbrace{\Big(|k|^{1/2} e^{-\mathrm{i}k^2t}\widehat{f}_+\Big)^{\vee}}_{=:h_+} + \underbrace{\Big(|k|^{1/2} e^{-\mathrm{i}k^2t}\widehat{f}_-\Big)^{\vee}}_{=:h_-},$$

where $\widehat{f}_\pm(k) := \chi_{\mathbb{R}_\pm}(k)\widehat{f}(k)$. Because

$$|D_x^{1/2} e^{\mathrm{i}t\Delta} f|^2 \leqslant |h_+|^2 + |h_-|^2 + 2|h_+||h_-| \leqslant 2(|h_+|^2 + |h_-|^2)$$

and

$$\|f\|^2_{L^2} = \|\widehat{f}\|^2_{L^2} = \|\widehat{f}_+\|^2_{L^2} + \|\widehat{f}_-\|^2_{L^2},$$

it suffices to show $\int_{\mathbb{R}} |h_\pm(x,t)|^2 dt \leqslant c\|\widehat{f}_\pm\|^2_{L^2}$. We present in detail the estimate for h_+. For h_- the calculation is analogous.

$$\int_{\mathbb{R}} |h_+(x,t)|^2 dt = \int_{-\infty}^{\infty} |D_x^{1/2} e^{\mathrm{i}t\Delta} f_+|^2 dt = (2\pi)^{-1} \int_{-\infty}^{\infty} \left| \int_{\mathbb{R}} |k|^{1/2} e^{\mathrm{i}kx} e^{-\mathrm{i}k^2t} \widehat{f}_+(k) dk \right|^2 dt.$$

Using the substitution $r = k^2$ (such that $dk = \frac{1}{2}r^{-1/2}dr$), we get

$$\int_{-\infty}^{\infty} |D_x^{1/2} e^{\mathrm{i}t\Delta} f_+|^2 dt = c \int_{-\infty}^{\infty} \left| \int_0^{\infty} r^{-1/4} e^{\mathrm{i}x\sqrt{r}} e^{-\mathrm{i}rt} \widehat{f}_+(\sqrt{r}) dr \right|^2 dt.$$

Because $\widehat{f}_+(k) = 0$ for $k < 0$, the inner integral equals $\sqrt{2\pi}\left(r^{-1/4}e^{\mathrm{i}x\sqrt{r}}\widehat{f}_+(\sqrt{r})\right)\widehat{\ }(t)$. The Plancherel identity thus yields

$$\int_{-\infty}^{\infty} |D_x^{1/2} e^{\mathrm{i}t\Delta} f_+|^2 dt = c \int_0^{\infty} |r^{-1/4} e^{\mathrm{i}x\sqrt{r}} \widehat{f}_+(\sqrt{r})|^2 dr = c \int_0^{\infty} |\widehat{f}_+(\sqrt{r})|^2 r^{-1/2} dr.$$

Undoing the substitution, we get

$$\int_{-\infty}^{\infty} |D_x^{1/2} e^{\mathrm{i}t\Delta} f_+|^2 dt = c \int_0^{\infty} |\widehat{f}_+(k)|^2 dk = c\|f_+\|^2_{L^2(\mathbb{R})}.$$

For h_- the calculation is analogous and one uses the substitution $r = -k^2$.

For $n \geqslant 2$ the same idea applies. Without loss of generality let us set $j = 1$. We define

$$\widehat{f}_\pm(k) := \chi_{\mathbb{R}_\pm}(k_1)\widehat{f}(k), \quad \overline{x} := (x_2, \ldots, x_n)^T, \quad \overline{k} := (k_2, \ldots, k_n)^T.$$

Once again, we perform the proof only for f_+. We use the transformation $(r, \overline{k}) := (k_1^2 + \cdots + k_n^2, k_2, \ldots, k_n) =: \Phi(k)$ so that

$$\det \Phi = \begin{vmatrix} 2k_1 & 2k_2 & \ldots & 2k_n \\ 0 & 1 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{vmatrix} = 2k_1$$

and $dk = dk_1 d\overline{k} = |\det \Phi|^{-1} dr d\overline{k} = (2|k_1|)^{-1} dr d\overline{k}$. We obtain

$$\begin{aligned} \int_{\mathbb{R}^n} |D_{x_1}^{1/2} e^{\mathrm{i}t\Delta} f_+(x)|^2 d\overline{x} dt =& c \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} e^{\mathrm{i}k\cdot x} |k_1|^{1/2} e^{-\mathrm{i}|k|^2 t} \widehat{f}_+(k) dk \right|^2 d\overline{x} dt \\ =& c \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} e^{\mathrm{i}x_1\sqrt{r - |\overline{k}|^2}} (r - |\overline{k}|^2)^{-1/4} e^{\mathrm{i}(\overline{k}\cdot\overline{x} - rt)} \widehat{f}_+(\Phi^{-1}(r, \overline{k})) dr d\overline{k} \right|^2 d\overline{x} dt, \end{aligned}$$

where only the branch $k_1 = \sqrt{r - |\overline{k}|^2}$ was used since $\widehat{f}_+(k) = 0$ for $k_1 < 0$. The inner integral equals

$$(2\pi)^{n/2} \left((r - |\overline{k}|^2)^{-1/4} e^{\mathrm{i}x_1\sqrt{r - |\overline{k}|^2}} \widehat{f}_+(\Phi^{-1}(r, \overline{k})) \right)\widehat{\ } (t, -\overline{x}).$$

By the Plancherel identity

$$\begin{aligned} \int_{\mathbb{R}^n} |D_{x_1}^{1/2} e^{\mathrm{i}t\Delta} f_+(x)|^2 d\overline{x} dt &= c \int_{\mathbb{R}^n} \left| (r - |\overline{k}|^2)^{-1/4} e^{\mathrm{i}x_1\sqrt{r - |\overline{k}|^2}} \widehat{f}_+(\Phi^{-1}(r, \overline{k})) \right|^2 dr d\overline{k} \\ &= c \int_{\mathbb{R}^n} \left| \widehat{f}_+(\Phi^{-1}(r, \overline{k})) \right|^2 (r - |\overline{k}|^2)^{-1/2} dr d\overline{k}, \end{aligned}$$

and finally undoing the change of variables,

$$\int_{\mathbb{R}^n} |D_{x_1}^{1/2} e^{\mathrm{i}t\Delta} f_+(x)|^2 d\overline{x} dt = c \int_{\mathbb{R}^n} \left| \widehat{f}_+(k) \right|^2 dk = c\|f_+\|^2_{L^2(\mathbb{R}^n)}$$

with c independent of x_1. For f_- the calculation is analogous and one uses the substitution $\Phi(k) = (-|k|^2, k_2, \ldots, k_n)$. $\square$

Corollary 1.18. *For all $f \in L^2(\mathbb{R}^n)$ and $R > 0$*

$$\int_{-\infty}^{\infty}\int_{|x|\leqslant R}\left|D_x^{1/2}e^{\mathrm{i}t\Delta}f\right|^2 dxdt \leqslant cR\|f\|^2_{L^2(\mathbb{R}^n)},$$

where $D_x^{1/2}g(x,t) := \left(|k|^{1/2}\widehat{g}(k,t)\right)^{\vee}(x,t)$.

Proof. (of Corollary 1.18) For $n = 1$ the statement follows directly from (1.42).

For $n > 1$ let us first define the sectors $D_j := \{k \in \mathbb{R}^n : |k_j| > \frac{1}{\sqrt{2n}}|k|\}, j = 1,\ldots,n$. It is easy to see that $\cup_{j=1}^n D_j = \mathbb{R}^n\backslash\{0\}$. We choose next a partition of unity $\{\varphi_j\}_{j=1}^n$ on the open set $\mathbb{R}^n\backslash\{0\}$ such that

$$\varphi_j \in C^\infty, \quad \mathrm{supp}\varphi_j \subset D_j, \quad 0 \leqslant \varphi_j \leqslant 1, \quad \sum_{j=1}^n \varphi_j = 1.$$

To prove the statement we need to somehow carry the estimates on each partial half-derivative $D_{x_j}^{1/2}$ in (1.43) over to an estimate on the $D_x^{1/2}$ derivative. The main idea is that this can be done in each sector D_j due to the estimate $|k| \leqslant c|k_j|$.

For $f \in L^2(\mathbb{R}^n)$ define

$$\widehat{f}_j := \widehat{f}\varphi_j, \quad \widehat{g} := |k|^{1/2}\widehat{f}, \quad \widehat{g}_j := \varphi_j\widehat{g}.$$

Clearly, $\widehat{f} = \sum_{j=1}^n \widehat{f}_j$. We also have the estimate

$$\||k_j|^{-1/2}\widehat{g}_j\|_{L^2(\mathbb{R}^n)} \leqslant c\|\widehat{f}\|_{L^2(\mathbb{R}^n)} \tag{1.44}$$

because $\||k_j|^{-1/2}\widehat{g}_j\|_{L^2(\mathbb{R}^n)} = \||k_j|^{-1/2}\widehat{g}_j\|_{L^2(D_j)} \leqslant c\||k|^{-1/2}\widehat{g}_j\|_{L^2(\mathbb{R}^n)} \leqslant c\||k|^{-1/2}\widehat{g}\|_{L^2(\mathbb{R}^n)} = c\|\widehat{f}\|_{L^2(\mathbb{R}^n)}$, where the last step follows from $|\widehat{g}_j| = |\varphi_j\widehat{g}| \leqslant |\widehat{g}|$.

We can now estimate

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{|x|\leqslant R}\left|e^{\mathrm{i}t\Delta}g\right|^2 dxdt &\leqslant c\sum_{j=1}^n\int_{-\infty}^{\infty}\int_{|x|\leqslant R}\left|e^{\mathrm{i}t\Delta}g_j\right|^2 dxdt = c\sum_{j=1}^n\int_{-\infty}^{\infty}\int_{|x|\leqslant R}\left|(|k_j|^{1/2}e^{-\mathrm{i}t|k|^2}|k_j|^{-1/2}\widehat{g}_j)^{\vee}\right|^2 dxdt\\ &\leqslant c\sum_{j=1}^n\int_{|x_j|\leqslant R}\int_{\mathbb{R}^n}\left|(|k_j|^{1/2}e^{-\mathrm{i}t|k|^2}|k_j|^{-1/2}\widehat{g}_j)^{\vee}\right|^2 dxdt\\ &\leqslant Rc\sum_{j=1}^n\||k_j|^{-1/2}\widehat{g}_j\|^2_{L^2(\mathbb{R}^n)},\end{aligned}$$

where the last step follows from the 1D estimate (1.42) and the Plancherel identity. Using now (1.44), we obtain

$$\int_{-\infty}^{\infty}\int_{|x|\leqslant R}\left|e^{\mathrm{i}t\Delta}g\right|^2 dxdt \leqslant Rcn\|\widehat{f}\|^2_{L^2(\mathbb{R}^n)}.$$

This is the estimate in the corollary because $e^{\mathrm{i}t\Delta}g = \left(|k|^{1/2}e^{-\mathrm{i}t|k|^2}\widehat{f}\right)^{\vee} = D_x^{1/2}e^{\mathrm{i}t\Delta}f$.

□

The shift invariance $x \mapsto x + x_0, x_0 \in \mathbb{R}^n$ of the Schrödinger equation implies that $e^{\mathrm{i}t\Delta}f(\cdot + x_0) = (e^{\mathrm{i}t\Delta}f)(\cdot + x_0)$ such that

$$\sup_{x_0\in\mathbb{R}^n}\int_{-\infty}^{\infty}\int_{x\in B_R(x_0)}\left|D_x^{1/2}e^{\mathrm{i}t\Delta}f\right|^2 dxdt \leqslant cR\|f\|^2_{L^2(\mathbb{R}^n)}$$

and hence $\|D_x^{1/2}e^{\mathrm{i}t\Delta}f\|_{L^2_{\mathrm{loc}}} < \infty$ for almost all $t \in \mathbb{R}$. From the L^2-isometry of $e^{\mathrm{i}t\Delta}$ we also have $\|e^{\mathrm{i}t\Delta}f\|_{L^2} = \|f\|_{L^2} < \infty$ for all $t \in \mathbb{R}$. In summary

$$e^{\mathrm{i}t\Delta}f \in H^{1/2}_{\mathrm{loc}} \text{ for almost all } t \in \mathbb{R},$$

where

$$H^s_{\mathrm{loc}}(\mathbb{R}^n) = \{u : \mathbb{R}^n \to \mathbb{C} : \psi u \in H^s(\mathbb{R}^n) \text{ for all } \psi \in C_0^\infty(\mathbb{R}^n)\}.$$

Clearly, functions $f \in H^s_{\mathrm{loc}}(\mathbb{R}^n)$ are those, for which $f|_\Omega = g|_\Omega$ with some $g \in H^s(\mathbb{R}^n)$ for every compact set $\Omega \subset \mathbb{R}^n$.

1.3.0.3 Global Smoothing of the Schrödinger Equation

As we mentioned already after Lemma 1.16, the Schrödinger group $e^{\mathrm{i}t\Delta}$ does not have global smoothing in the sense of $H^s(\mathbb{R}^n)$, i.e. in general $f \in H^s(\mathbb{R}^n)$ does not imply $e^{\mathrm{i}t\Delta}f \in H^r(\mathbb{R}^n)$ with $r > s$. There is, however, a global smoothing in L^p spaces.

Theorem 1.19. *Let $\frac{1}{p} + \frac{1}{p'} = 1, p' \in [1,2]$ and $t \neq 0$. Then the operator $e^{\mathrm{i}t\Delta} : L^{p'}(\mathbb{R}^n) \to L^p(\mathbb{R}^n)$ is continuous and*

$$\|e^{\mathrm{i}t\Delta}f\|_{L^p(\mathbb{R}^n)} \leqslant c|t|^{-\frac{n}{2}(\frac{1}{p'}-\frac{1}{p})}\|f\|_{L^{p'}(\mathbb{R}^n)}.$$

Note that for $p' < 2$ is $p \geqslant p'$ so that $e^{\mathrm{i}t\Delta}f$ is, indeed, more regular than f in the sense that the blowup at singularities must be milder. If $f \in L^1$, then Theorem 1.19 implies $e^{\mathrm{i}t\Delta}f \in L^\infty$. For $p' = 2$ the theorem provides nothing else than what we already know from the isometry property of $e^{\mathrm{i}t\Delta}$ in L^2.

One can easily show that the result is sharp, i.e. that for a given $p \geqslant 2$ and $r > p$ there are $f \in L^p$ such that $e^{\mathrm{i}t\Delta}f \notin L^r$. Take namely $g \in L^{p'}\backslash L^r$ with $1/p + 1/p' = 1$ and let $f := e^{-\mathrm{i}t\Delta}g$. Then $f \in L^p$ due to Theorem 1.19 and $e^{\mathrm{i}t\Delta}f = g \notin L^r$.

To prove Theorem 1.19, we need the Riesz-Thorin interpolation theorem and a simple form of the Young's inequality for convolutions. Here we prove only the latter one.

Theorem 1.20. *(Riesz-Thorin interpolation theorem) Let T be bounded as an operator $T : L^{p_0}(\mathbb{R}^n) \to L^{q_0}(\mathbb{R}^n)$ as well as $T : L^{p_1}(\mathbb{R}^n) \to L^{q_1}(\mathbb{R}^n)$ with $1 \leqslant p_0, p_1, q_0, q_1 \leqslant \infty$. Define*

$$M_0 := \|T\|_{L^{p_0}\to L^{q_0}} \quad \text{and} \quad M_1 := \|T\|_{L^{p_1}\to L^{q_1}}.$$

Then is $T : L^{p_\theta}(\mathbb{R}^n) \to L^{q_\theta}(\mathbb{R}^n)$ bounded and

$$M_\theta \leqslant M_0^{1-\theta}M_1^\theta,$$

where

$$\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}, \quad \frac{1}{q_\theta} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1}, \ \theta \in (0,1), \tag{1.45}$$

and $M_\theta := \|T\|_{L^{p_\theta}\to L^{q_\theta}}$.

For the proof see Theorem 2.1 in [14].

Lemma 1.21. *(Young's inequality for convolutions) Let $f \in L^1(\mathbb{R}^n), g \in L^p(\mathbb{R}^n)$ and $1 \leqslant p \leqslant \infty$. Then*

$$\|f * g\|_{L^p(\mathbb{R}^n)} \leqslant \|f\|_{L^1(\mathbb{R}^n)}\|g\|_{L^p(\mathbb{R}^n)}.$$

Proof. The case $p = \infty$ is simple

$$|(f*g)(x)| = \left|\int_{\mathbb{R}^n} f(x-y)g(y)dy\right| \leqslant \|g\|_{L^\infty(\mathbb{R}^n)}\int_{\mathbb{R}^n}|f(x-y)|dy = \|g\|_{L^\infty(\mathbb{R}^n)}\|f\|_{L^1(\mathbb{R}^n)}.$$

For $1 \leqslant p < \infty$ one argues using the Hölder inequality. With $\frac{1}{p} + \frac{1}{p'} = 1$ we get

$$\begin{aligned}\int_{\mathbb{R}^n} |f(x-y)g(y)|dy &= \int_{\mathbb{R}^n} |f(x-y)|^{\frac{1}{p'}} |f(x-y)|^{\frac{1}{p}} |g(y)|dy \\ &\leqslant \left(\int_{\mathbb{R}^n} |f(x-y)|dy\right)^{\frac{1}{p'}} \left(\int_{\mathbb{R}^n} |f(x-y)||g(y)|^p dy\right)^{\frac{1}{p}} \\ &= \|f\|_{L^1(\mathbb{R}^n)}^{\frac{1}{p'}} \left(\int_{\mathbb{R}^n} |f(x-y)||g(y)|^p dy\right)^{\frac{1}{p}}\end{aligned}$$

such that

$$\begin{aligned}\int_{\mathbb{R}^n} \left|\int_{\mathbb{R}^n} |f(x-y)g(y)|dy\right|^p dx &\leqslant \|f\|_{L^1(\mathbb{R}^n)}^{\frac{p}{p'}} \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} |f(x-y)||g(y)|^p dydx \\ &= \|f\|_{L^1(\mathbb{R}^n)}^{\frac{p}{p'}} \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} |f(x-y)|dx|g(y)|^p dy = \|f\|^p_{L^1(\mathbb{R}^n)} \|g\|^p_{L^p(\mathbb{R}^n)},\end{aligned}$$

where Tonelli's theorem for non-negative functions has been used to interchange the order of integration. □

Let us now prove Theorem 1.19.
Proof. (of Theorem 1.19) The case $p = p' = 2$ follows by the isometry of $e^{\mathrm{i}t\Delta}$. For $p' = 1, p = \infty$ we have

$$\|e^{\mathrm{i}t\Delta} f\|_{L^\infty} = \|(4\pi \mathrm{i}t)^{-n/2} e^{\mathrm{i}\frac{|\cdot|^2}{4t}} * f\|_{L^\infty} \leqslant \|(4\pi \mathrm{i}t)^{-n/2} e^{\mathrm{i}\frac{|\cdot|^2}{4t}}\|_{L^\infty} \|f\|_{L^1} \leqslant c|t|^{-\frac{n}{2}} \|f\|_{L^1}.$$

Hence, we have that $e^{\mathrm{i}t\Delta}$ is bounded as an operator $e^{\mathrm{i}t\Delta} : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ with $M_1 := \|e^{\mathrm{i}t\Delta}\|_{L^2 \to L^2} = 1$ and as an operator $e^{\mathrm{i}t\Delta} : L^1(\mathbb{R}^n) \to L^\infty(\mathbb{R}^n)$ with $M_0 := \|e^{\mathrm{i}t\Delta}\|_{L^1 \to L^\infty} = c|t|^{-\frac{n}{2}}$.

For $p' \in (1, 2)$ we can thus use the Riesz-Thorin Theorem 1.20 with $p_1 = q_1 = 2$, $p_0 = 1, q_0 = \infty$ and $p_\theta = p', q_\theta = p$. Indeed, we show that there is $\theta \in (0, 1)$ such that (1.45) holds. The second equation in (1.45) reads $1/p = \theta/2$ which holds if we set $\theta := 2/p$. Clearly, $\theta \in (0, 1)$ as $p \in (2, \infty)$. The first equation reads $1 - \theta/2 = 1/p'$, which hold due to the choice of θ and $1/p + 1/p' = 1$.

In conclusion $e^{\mathrm{i}t\Delta}$ is bounded as $e^{\mathrm{i}t\Delta} : L^{p'} \to L^p$ with

$$\|e^{\mathrm{i}t\Delta} f\|_{L^p} \leqslant M_0^{1-\theta} M_1^\theta \|f\|_{L^{p'}} = c|t|^{-\frac{n}{2}(\frac{1}{p'} - \frac{1}{p})} \|f\|_{L^{p'}}.$$

□

Much more can be said about the smoothing properties of the Schrödinger equation and of other linear dispersive equations with unbounded group velocities. For example, the linear Korteweg-de Vries equation $\partial_t u + \partial_x^3 u = 0, x \in \mathbb{R}$ satisfies a property similar to the above local smoothing of the Schrödinger equation; namely for $s \in [0, 1]$ and $u_0 \in L^2(\mathbb{R})$ one has

$$\sup_{x \in R} \int_{\mathbb{R}} |D^s u(x,t)|^2 dt \leqslant c|t|^{\frac{1-s}{3}} \|u_0\|_{L^2(\mathbb{R}^n)},$$

see [5].

Many results go under the name of Strichartz estimates, see e.g. [14] and [23]. For example, for the Schrödinger equation one has

$$\|e^{\mathrm{i}t\Delta} f\|_{L^q_t L^r_x(\mathbb{R} \times \mathbb{R}^n)} \leqslant c\|f\|_{L^2(\mathbb{R}^n)}$$

if $2 \leqslant q, r \leqslant \infty, 2/q + n/r = n/2$ and $(q, r, n) \neq (2, \infty, 2)$, see Theorem 2.3 in [23].

1.4 Waves in Periodic Structures

As a natural generalization of wave problems in homogenous media we consider those in periodic media, i.e. those modeled by PDEs with spatially periodic coefficients. This setting is also physically highly relevant with some classical examples being light in photonic crystals, electron waves in atomic crystalline structures, elastic waves in periodically arranged mechanic constructions or Bose-Einstein condensates in optical lattices.

For the sake of simplicity of the presentation we restrict here to waves in one spatial dimension.

1.4.1 Bloch Transformation

A basic tool in the analysis of waves in periodic media is the Bloch transformation. It is a generalization of the Fourier transformation. Its definition is motivated by the following calculation. For $f \in S(\mathbb{R})$ and an arbitrary constant $L > 0$ we get by a simple calculation

$$f(x) = (2\pi)^{-1/2} \int_{\mathbb{R}} e^{\mathrm{i}kx} \hat{f}(k) dk = (2\pi)^{-1/2} \sum_{m \in \mathbb{Z}} \int_{-\frac{\pi}{L}}^{\frac{\pi}{L}} e^{\mathrm{i}(k+m\frac{2\pi}{L})x} \hat{f}(k + m\tfrac{2\pi}{L}) dk = (2\pi)^{-1/2} \int_{-\frac{\pi}{L}}^{\frac{\pi}{L}} e^{\mathrm{i}kx} \tilde{f}(x,k) dk, \tag{1.46}$$

where

$$\tilde{f}(x,k) := \sum_{m \in \mathbb{Z}} \widehat{f}\left(k + m\tfrac{2\pi}{L}\right) e^{\mathrm{i}m\frac{2\pi}{L}x} \tag{1.47}$$

and where in the last step of the calculation the order of integration and summation was exchanged using Fubini's theorem with the counting mass.

The *Bloch transformation* is the operator $\mathcal{T} : f \mapsto \mathcal{T}(f) := \tilde{f}$.

Next we prove some simple properties of $\mathcal{T}$ which are useful for the analysis of PDEs with periodic coefficients.

Lemma 1.22. *For all $f \in S(\mathbb{R})$, $x, k \in \mathbb{R}$, and $p \in \mathbb{N}$ is*

(i) $\tilde{f}(x + L, k) = \tilde{f}(x,k)$,

(ii) $\tilde{f}\left(x, k + \frac{2\pi}{L}\right) = e^{-\mathrm{i}\frac{2\pi}{L}x} \tilde{f}(x,k)$,

(iii) $\mathcal{T}(\partial_x^p f)(x,k) = (\partial_x + \mathrm{i}k)^p \tilde{f}(x,k)$,

(iv) $\mathcal{T}(Vf)(x,k) = V(x)\tilde{f}(x,k)$ *if* $V \in C(\mathbb{R})$, *piecewise* $C^1(\mathbb{R})$ *and* $V(x + L) = V(x)$ *for all* $x \in \mathbb{R}$.

Proof. Properties (i) and (ii) follow directly from the definition (1.47). For (iii) we use property (A.3) for the Fourier transform of the derivative:

$$\begin{aligned} \mathcal{T}(\partial_x^p f)(x,k) &= \sum_{m \in \mathbb{Z}} \mathrm{i}^p \left(k + \tfrac{2m\pi}{L}\right)^p \widehat{f}\left(k + \tfrac{2m\pi}{L}\right) e^{\mathrm{i}\frac{2m\pi}{L}x} \\ &= \sum_{m \in \mathbb{Z}} (\partial_x + \mathrm{i}k)^p \left(\widehat{f}\left(k + \tfrac{2m\pi}{L}\right) e^{\mathrm{i}\frac{2m\pi}{L}x} \right) = (\partial_x + \mathrm{i}k)^p \tilde{f}(x,k). \end{aligned}$$

For (iv) we write first V as a Fourier series (which converges pointwise due to the assumptions on V):

$$V(x) = \sum_{j \in \mathbb{Z}} V_j e^{\mathrm{i}j\frac{2\pi}{L}x}, \ V_j \in \mathbb{C} \text{ for all } j \in \mathbb{Z}.$$

Using the relation $\widehat{\left(e^{\mathrm{i}j\frac{2\pi}{L}\cdot}f(\cdot)\right)}(k) = \hat{f}\left(k - j\frac{2\pi}{L}\right)$, we obtain

$$\begin{aligned}\mathcal{T}(Vf)(x,k) &= \sum_{m\in\mathbb{Z}}\sum_{j\in\mathbb{Z}} V_j\hat{f}\left(k+(m-j)\tfrac{2\pi}{L}\right)e^{\mathrm{i}m\frac{2\pi}{L}x}\\ &= \sum_{j\in\mathbb{Z}} V_j e^{\mathrm{i}j\frac{2\pi}{L}x}\sum_{m\in\mathbb{Z}}\hat{f}\left(k+(m-j)\tfrac{2\pi}{L}\right)e^{\mathrm{i}(m-j)\frac{2\pi}{L}x} = V(x)\tilde{f}(x,k).\end{aligned}$$

□

In other words, the Bloch transformation produces a family of L-periodic functions parametrized by $k \in (-\pi/L, \pi/L]$. We denote this interval by

$$\mathbb{B} := (-\pi/L, \pi/L].$$

In Section 1.4.2 we will study the application of this transform to problems with $L-$periodic coefficients. In that setting $\mathbb{B}$ is called the *first Brillouin zone.* The restriction to $k \in \mathbb{B}$ is possible due to the periodicity relation in Lemma 1.22 (ii). The Bloch transform commutes with $L-$periodic functions and its action on the derivative produces the shifted derivative $\partial_x + \mathrm{i}k$.

In order to put the Bloch transform in a proper functional analytic setting, we define the spaces

$$\begin{aligned}H^s_m(\mathbb{R}) :=& \{u \in H^s(\mathbb{R}) : \|u\|_{H^s_m(\mathbb{R})} := \|u\rho\|_{H^s(\mathbb{R})} < \infty,\ \text{where } \rho(x) = (1+x^2)^{m/2}\},\\ Z^s_m :=& H^m_{bp}(\mathbb{B}, H^s(0,L)) = \text{closure}_{H^m(\mathbb{B},H^s(0,L))} C^\infty_{bp}(\mathbb{B}, H^s(0,L)),\\ C^\infty_{bp}(\mathbb{B}, H^s(0,L)) :=& \{u : \mathbb{B} \to H^s(0,L), k \mapsto u(\cdot,k) : \exists v \in C^\infty(\mathbb{R}, H^s(0,L))\\ & \left.\text{with } v\left(x, k+\tfrac{2\pi}{L}\right) = e^{-\mathrm{i}\frac{2\pi}{L}x}v(x,k), u = v|_{k\in\mathbb{B}}\right\},\end{aligned}$$

and prove the following theorem (cf. Lemma 5.4 in [21]).

Theorem 1.23. *For each $s, m \in \mathbb{N}_0$ the Bloch transform $\mathcal{T}$ is an isomorphism from $H^s_m(\mathbb{R})$ to Z^s_m.*

Remark 1.5. *This should be compared with the Fourier transform, which is an isomorphism from $H^s_m(\mathbb{R})$ to $H^m_s(\mathbb{R})$. Also note that for $m = 0$ the isomorphism is from $H^s(\mathbb{R})$ to $L^2_{bp}(\mathbb{B}, H^s(0,L))$.*

Proof. Let us first denote the operator given by (1.46) as $\mathcal{T}_1 : \tilde{f} \mapsto f$. It is a direct calculation to show that for $u \in S(\mathbb{R})$ is $\mathcal{T}_1\mathcal{T}u = u$ and for $\tilde{u} \in C^\infty_{bp}(\mathbb{B}, H^s(0,L))$ is $\mathcal{T}\mathcal{T}_1\tilde{u} = \tilde{u}$.

The idea of the proof is to first show that $\mathcal{T}$ and $\mathcal{T}_1$ are uniformly continuous maps defined on S and C^∞_{bp} respectively and because S is dense in H^s_m and C^∞_{bp} is dense in Z^s_m, Lemma 1.24 provides continuous extensions on the larger spaces. The fact that $\mathcal{T}_1 = \mathcal{T}^{-1}$ on the dense subspace then guarantees that the extensions are also the inverse of each other.

We first show the L^2-isometry (up to the factor L).

$$\begin{aligned}\|u\|^2_{L^2(\mathbb{R})} =& \|\hat{u}\|_{L^2(\mathbb{R})} = \sum_{m\in\mathbb{Z}}\int_{-\frac{\pi}{L}}^{\frac{\pi}{L}} \left|\hat{u}\left(k+m\tfrac{2\pi}{L}\right)\right|^2 dk = \int_{-\frac{\pi}{L}}^{\frac{\pi}{L}}\sum_{m\in\mathbb{Z}}\left|\hat{u}\left(k+m\tfrac{2\pi}{L}\right)\right|^2 dk\\ =& \frac{1}{L}\int_{-\frac{\pi}{L}}^{\frac{\pi}{L}}\int_0^L |\tilde{u}(x,k)|^2 dx dk = \frac{1}{L}\|\tilde{u}\|^2_{L^2(\mathbb{B},L^2(0,L))},\end{aligned}$$

where the sum and the integral were exchanged using the Fubini theorem with the counting mass and where in the one but last step $\int_0^L |\tilde{u}(x,k)|^2 dx = L\sum_{m\in\mathbb{Z}}\left|\hat{u}\left(k+m\frac{2\pi}{L}\right)\right|^2$ follows from $\int_0^L e^{\mathrm{i}(m-n)\frac{2\pi}{L}x} = 0$ for all $m, n \in \mathbb{Z}, m \neq n$.

For the proof of the uniform continuity of $\mathcal{T}_1 : C_{bp}^\infty \to S$ we use the following relations for the Fourier transform $\mathcal{F}$. For all $p, j \in \mathbb{N}_0$

$$\mathcal{F}^{-1}(\partial_k^p \widehat{f})(x) = (-\mathrm{i}x)^p f(x), \quad \mathcal{F}^{-1}(\partial_k^p \widehat{\partial_x^j f})(x) = (-\mathrm{i}x)^p \partial_x^j f(x).$$

Hence, for some constants $c_p, \tilde{c}_p > 0, p \in \{0, \dots, m\}$ and all $\tilde{u} \in C_{bp}^\infty$

$$\begin{aligned}
\|u\|_{H_m^s(\mathbb{R})}^2 &= \sum_{j=0}^s \int_{\mathbb{R}} |\partial_x^j u(x)|^2 (1+x^2)^m dx = \sum_{p=0}^m c_p \sum_{j=0}^s \int_{\mathbb{R}} |\partial_k^p \widehat{\partial_x^j u}|^2 dk = \sum_{p=0}^m c_p \sum_{j=0}^s \int_{\mathbb{R}} |\partial_k^p ((\mathrm{i}k)^j \widehat{u})|^2 dk \\
&\leqslant \sum_{p=0}^m \tilde{c}_p \sum_{j=0}^s \int_{\mathbb{R}} k^{2j} |\partial_k^p \widehat{u}|^2 dk = \sum_{p=0}^m \tilde{c}_p \sum_{j=0}^s \sum_{n\in\mathbb{Z}} \int_{-\frac{\pi}{L}}^{\frac{\pi}{L}} \left(k + n\tfrac{2\pi}{L}\right)^{2j} \left|\partial_k^p \widehat{u}\left(k + n\tfrac{2\pi}{L}\right)\right|^2 dk \\
&\leqslant c \sum_{p=0}^m \sum_{j=0}^s \int_{-\frac{\pi}{L}}^{\frac{\pi}{L}} \sum_{n\in\mathbb{Z}} \left(n\tfrac{2\pi}{L}\right)^{2j} \left|\partial_k^p \widehat{u}\left(k + n\tfrac{2\pi}{L}\right)\right|^2 dk = \frac{c}{L} \sum_{p=0}^m \sum_{j=0}^s \int_{-\frac{\pi}{L}}^{\frac{\pi}{L}} \int_0^L |\partial_x^j \partial_k^p \tilde{u}(x,k)|^2 dx dk = \frac{c}{L} \|\tilde{u}\|_{Z_m^s}^2.
\end{aligned}$$

Analogously, one shows that $\|\tilde{u}\|_{Z_m^s} \leqslant c_2 \|u\|_{H_m^s(\mathbb{R})}$ for all $u \in S$ and some $c_2 > 0$. Clearly, $\mathcal{T} : S \to C_{bp}^\infty$ is uniformly continuous since $\|\mathcal{T}(f-g)\|_{Z_m^s} \leqslant \sqrt{c_2}\|f-g\|_{Z_m^s}$ and similarly $\mathcal{T}_1 : C_{bp}^\infty \to S$ is uniformly continuous. Using Lemma 1.24 there is a unique continuous extension of $\mathcal{T}^1$ to $\mathbb{Z}_m^s$ and of $\mathcal{T}$ to H_m^s. We denote these extensions again by $\mathcal{T}_1$ and $\mathcal{T}$ respectively.

It remains to show that $\mathcal{T}_1 = \mathcal{T}^{-1}$. Note that $\mathcal{T}_1\mathcal{T} = I$ and $\mathcal{T}\mathcal{T}_1 = I$ on dense subspaces. Let $u \in H_m^s$ and choose an approximating sequence $(u_n)_{n\in\mathbb{N}} \subset S$ such that $\|u_n - u\|_{H_m^s} \to 0$ as $n \to \infty$. Then

$$\|\mathcal{T}_1\mathcal{T}u - u\|_{H_m^s} \leqslant \lim_{n\to\infty} \left(\|\mathcal{T}_1\mathcal{T}u_n - \mathcal{T}_1\mathcal{T}u\|_{H_m^s} + \|\mathcal{T}_1\mathcal{T}u_n - u_n\|_{H_m^s} + \|u_n - u\|_{H_m^s}\right) = 0$$

because $\mathcal{T}_1\mathcal{T}u_n = u_n$ and $\mathcal{T}_1\mathcal{T}$ is continuous. Analogously one shows $\mathcal{T}\mathcal{T}_1 = I$. □

Lemma 1.24. *Let X and Y be two metric spaces, where Y is complete. Let $A \subset X$ be a dense subset and $f : A \to Y$ be uniformly continuous. Then f has a unique continuous extension $g : X \to Y$.*

For the proof see e.g. Theorem I.6.17 in [8].

1.4.2 Application of the Bloch Transformation to the Analysis of PDEs with Periodic Coefficients

Let us consider for the sake of simplicity only scalar equations in one spatial dimension of the form

$$\partial_t u + \sum_{j=0}^p \alpha_j(x)\partial_x^j u = 0, \quad x \in \mathbb{R}, t > 0, \qquad u(x,0) = u_0(x) \tag{1.48}$$

with

$$\alpha_j(x+L) = \alpha_j(x) \quad \text{for all } x \in \mathbb{R}.$$

After the application of $\mathcal{T}$ to (1.48) we get the following periodic system parametrized by $k \in \mathbb{B}$

$$\partial_t \tilde{u} + \sum_{j=0}^p \alpha_j(x)(\partial_x + \mathrm{i}k)^j \tilde{u} = 0, x \in [0, L], \qquad \tilde{u}(x,k,0) = \tilde{u}_0(x,k). \tag{1.49}$$

Assuming that the operator

$$\mathcal{L}(\cdot, k) : H_{\text{per}}^s(0,L) \to H_{\text{per}}^{s-p}(0,L), \ \tilde{u} \mapsto \sum_{j=0}^p \alpha_j(\cdot)(\partial_x + \mathrm{i}k)^j \tilde{u}(\cdot, k)$$

is skew symmetric and has a compact resolvent, i.e.

$$(\mathcal{L}(\cdot,k)-\lambda I)^{-1}: H^s_{\text{per}}(0,L)\to H^s_{\text{per}}(0,L)$$

is compact for some λ in the resolvent set of $\mathcal{L}(\cdot,k)$ (and hence automatically for all λ in the resolvent set), then a standard result of functional analysis is that

- the spectrum $\sigma(\mathcal{L}(\cdot,k))$ is imaginary, at most countable, has no accumulation points and is given by
$$\sigma(\mathcal{L}(\cdot,k)) = \cup_{n\in\mathbb{N}}\{\mathrm{i}\omega_n(k)\},\quad \text{where } \omega_n(k)\in\mathbb{R} \text{ for all } n\in\mathbb{N},$$
- for each $k\in\mathbb{B}$ the eigenfunctions $(p_n(\cdot,k))_{n\in\mathbb{N}}$ of $\mathcal{L}(\cdot,k)$ form (after a proper normalization) an orthonormal basis of $L^2(0,L)$.

Therefore, we have for any $f\in L^2(0,L)$ the expansion $f(x)=\sum_{n\in\mathbb{N}}F_n(k)p_n(x,k)$ with some $(F_n(k))_{n\in\mathbb{N}}\subset\mathbb{C}$. Applying this expansion to $\tilde{u}(\cdot,k,t)$ and $\tilde{u}_0(\cdot,k)$ such that

$$\tilde{u}(x,k,t)=\sum_{n\in\mathbb{N}}U_n(k,t)p_n(x,k),\qquad \tilde{u}_0(x,k)=\sum_{n\in\mathbb{N}}U_{0,n}(k)p_n(x,k),$$

problem (1.49) becomes

$$\partial_t U_n(k,t)+\mathrm{i}\omega_n(k)U_n(k,t)=0,\quad k\in\mathbb{B},n\in\mathbb{N},t>0,\qquad U_n(k,0)=U_{0,n}(k).$$

This is a (infinite dimensional) system of decoupled ODEs parametrized by $k\in\mathbb{B}$. The solution is clearly

$$U_n(k,t)=e^{-\mathrm{i}\omega_n(k)t}U_{0,n}(k).$$

For the transformed variable $\tilde{u}$ we get $\tilde{u}(x,k,t)=\sum_{n\in\mathbb{N}}p_n(x,k)U_{0,n}(k)e^{-\mathrm{i}\omega_n(k)t}$ and hence the solution of (1.48) is given by

$$u(x,t)=(2\pi)^{-1/2}\int_{\mathbb{B}}\sum_{n\in\mathbb{N}}p_n(x,k)U_{0,n}(k)e^{\mathrm{i}(kx-\omega_n(k)t)}dk.$$

In analogy with constant coefficient problems discussed in Section 1.2.1, the sequence $(\omega_n(k))_{n\in\mathbb{N}}$ plays the role of $W(k)$, i.e. of the solution of the dispersion relation. The image $\cup_{n\in\mathbb{N}}\omega_n(\mathbb{B})$ is called the *band structure*. The pair $(\omega_n(k),p_n(x,k))$ is for each $k\in\mathbb{B},n\in\mathbb{N}$ an eigenpair of the Bloch eigenvalue problem

$$\begin{aligned}\mathcal{L}(x,k)p_n(x,k)&=\mathrm{i}\omega_n(k)p_n(x,k),x\in(0,L)\\ p_n(L,k)&=p_n(0,k).\end{aligned}\tag{1.50}$$

The functions $\Psi_n^{(k)}(x,t):=p_n(x,k)e^{\mathrm{i}(kx-\omega_n(k)t)}$ are called *Bloch waves* and are a generalization of the plane waves $e^{\mathrm{i}(kx-\omega t)}$ in the constant coefficient case. The *group velocity* of the Bloch wave is given by $\omega_n'(k)$.

1.4.2.1 Asymptotics for $t\to\infty$, $\frac{x}{t}=c\in\mathbb{R}$

Just like in Section 1.2.2 for constant coefficient problems, we show here that an observer moving at the velocity c sees after a long time basically only Bloch waves with the group velocity equal to c. Indeed, writing the solution as

$$u(x,t)=(2\pi)^{-1/2}\int_{\mathbb{B}}\sum_{n\in\mathbb{N}}p_n(x,k)U_{0,n}(k)e^{-\mathrm{i}\chi_n(k)t}dk,\ \chi_n(k)=\omega_n(k)-k\tfrac{x}{t},$$

we use the method of stationary phase (Theorem B.7) for $t\to\infty, x/t=c$ to get

$$u(x,t)\sim\sum_{\{(n,\tilde{k})\in\mathbb{N}\times\mathbb{B}:\omega_n'(\tilde{k})=c\}}\left(\frac{1}{t|\omega_n''(\tilde{k})|}\right)^{1/2}U_{0,n}(\tilde{k})p_n(x,\tilde{k})e^{-\mathrm{i}\,\mathrm{sign}(\omega_n''(\tilde{k}))\frac{\pi}{4}}e^{\mathrm{i}(\tilde{k}x-\omega_n(\tilde{k})t)}$$

if $\omega_n''(k)\neq 0$ for all $(n,k)\in\mathbb{N}\times\mathbb{B}$ such that $\omega_n'(\tilde{k})=c$.

1.5 Water Waves

In this section we derive equations for waves at a free interface of air and a fluid. We make several simplifying assumptions. Firstly, we consider only *incompressible inviscid* fluids in a *constant gravitational field*. Water in a standard situation satisfies these conditions to a high extent.

The starting point of our derivation are the Euler equations

$$\partial_t \vec{u} + (\vec{u} \cdot \nabla)\vec{u} = -\frac{1}{\rho}\nabla p - g\vec{j}, \tag{1.51a}$$

$$\nabla \cdot \vec{u} = 0, \tag{1.51b}$$

where $\vec{u} = (u_1, u_2, u_3)^T$ is the fluid velocity vector, p is the pressure, g is the acceleration due to gravity ($\approx 9.8m/s^2$), ρ is the density, and $\vec{j} = (0,0,1)^T$. Equation (1.51a) describes the conservation of momentum while (1.51b) is the incompressibility condition.

Note that $(\vec{u} \cdot \nabla)\vec{u} = (\sum_{j=1}^3 u_j\partial_j u_1, \sum_{j=1}^3 u_j\partial_j u_2, \sum_{j=1}^3 u_j\partial_j u_3)^T$ coincides with $(\nabla\vec{u})\vec{u}$, where $\nabla\vec{u}$ is the Jacobian matrix of $\vec{u}$.

In the field of fluid mechanics it is common to use the notation of the *material derivative* $D_t\vec{w} := \partial_t\vec{w} + (\vec{u} \cdot \nabla)\vec{w}$. The following calculation motivates this concept by showing that the rate of change of a quantity at a material point (i.e. a fluid paricle) flowing with the fluid is given by the material derivative. Let $x(t;x_0)$ be the position of a particle (material point) moving with the fluid and being at the point x_0 at $t = 0$. Then

$$\partial_t x(t;x_0) = \vec{u}(x(t;x_0), x), \quad x(0;x_0) = x_0.$$

Let now $\vec{w}(x,t)$ be a physical quantity at (x,t). $\vec{w}(x(t;x_0),t)$ is then this quantity at the above material point and the chain rule yields

$$\begin{aligned}\frac{d}{dt}\vec{w}(x(t;x_0),t) =& \partial_t\vec{w}(x(t;x_0),t) + \partial_t x_1(t;x_0)\partial_{x_1}\vec{w}(x(t;x_0),t) + \cdots + \partial_t x_3(t;x_0)\partial_{x_3}\vec{w}(x(t;x_0),t)\\ =& \partial_t\vec{w}(x(t;x_0),t) + (\vec{u}(x(t;x_0),t)\cdot\nabla)\,\vec{w}(x(t;x_0),t) = (D_t\vec{w})(x(t;x_0),t).\end{aligned}$$

The material derivative of a quantity $\vec{w}$ thus describes the temporal rate of change of $\vec{w}$ along the trajectory of particles flowing with the fluid.

Applying D_t to the velocity field $\vec{u}$, the Euler equation (1.51a) reads

$$D_t\vec{u} = -\frac{1}{\rho}\nabla p + g\vec{j}. \tag{1.52}$$

We restrict our attention to *irrotational* flows, i.e. we have zero vorticity

$$\nabla \times \vec{u} \equiv 0.$$

This is justified by the following discussion in which we show that if a flow is irrotational at the initial time, then it remains so for all times. First, a direct calculation, see e.g. [16], produces for the vorticity $\vec{\omega} := \nabla \times \vec{u}$

$$D_t\vec{\omega} = (\vec{\omega} \times \nabla)\vec{u}.$$

We will apply the following general result from [16].

Lemma 1.25. *Let $x(t;x_0)$ be the trajectory of a material point in a smooth velocity field $\vec{u}(x,t)$, such that*

$$\partial_t x(t;x_0) = \vec{u}(x(t;x_0);t), \quad x(0;x_0) = x_0, \tag{1.53}$$

and let $\vec{h}(x,t)$ be a smooth vector field. Then

$$D_t\vec{h} = (\vec{h} \cdot \nabla)\vec{u}, \quad \vec{h}(x_0,0) = \vec{h}_0(x_0) \tag{1.54}$$

if and only if

$$\vec{h}(x(t;x_0),t) = \nabla_{x_0}x(t;x_0)\vec{h}_0(x_0).$$

Proof. Differentiating equation (1.53) with respect to x_0, we have

$$\frac{d}{dt}\nabla_{x_0}x(t;x_0) = \nabla\vec{u}(x(t;x_0),t)\nabla_{x_0}x(t;x_0),$$

where $\nabla\vec{u}$ and $\nabla_{x_0}x$ are 3×3 Jacobian matrices. Hence

$$\frac{d}{dt}\nabla_{x_0}x(t;x_0)\vec{h_0}(x_0) = \nabla\vec{u}(x(t;x_0),t)\nabla_{x_0}x(t;x_0)\vec{h_0}(x_0).$$

Equation (1.54) is equivalent to

$$D_t\vec{h} = \frac{d}{dt}\vec{h}(x(t;x_0),t) = \nabla\vec{u}(x(t;x_0),t)\vec{h}(x(t;x_0),t)$$

because $(\vec{h}\cdot\nabla)\vec{u} = (\nabla\vec{u})\vec{h}$.

Hence, $\vec{h}(x(t;x_0),t)$ and $\nabla_{x_0}x(t;x_0)\vec{h_0}(x_0)$ satisfy the same initial value ODE problem with the initial data

$$\vec{h}_0(x_0).$$

Due to the uniqueness of the ODE solutions, we have the equivalence. □

Applying this lemma to $\vec{h} = \vec{\omega}$ with $\vec{\omega}(\cdot,0)\equiv 0$, we get $\vec{\omega}(\cdot,t)\equiv 0$ for all $t>0$.

Using now the Helmholtz decomposition

$$\vec{u} = \nabla\times\vec{\psi} + \nabla\varphi$$

with $\vec{\psi}(x,t)\in\mathbb{R}^3, \varphi(x,t)\in\mathbb{R}$, the zero vorticity condition allows us to set $\vec{u}$ equal to a gradient

$$\vec{u} = \nabla\varphi.$$

In other words, we can consider the so called ***potential flow***. Equation (1.51a) now becomes

$$\nabla\left(\partial_t\varphi + \frac{1}{2}|\nabla\varphi|^2 + \frac{p-p_0}{\rho} + gx_3\right) = 0$$

for an arbitrary constant $p_0\in\mathbb{R}$. Thus

$$\partial_t\varphi + \frac{1}{2}|\nabla\varphi|^2 + gx_3 = \frac{p_0-p}{\rho} + B(t)$$

for arbitrary scalar functions $B(t)$. Because adding an $x-$independent function to φ does not alter $\vec{u}$, we can replace $\varphi(x,t)$ by $\varphi(x,t)+\int_0^t B(s)ds$ and obtain the final form

$$\partial_t\varphi + \frac{1}{2}|\nabla\varphi|^2 + gx_3 = \frac{p_0-p}{\rho}, \tag{1.55a}$$

$$\Delta\varphi = 0. \tag{1.55b}$$

1.5.0.1 Free Interface of Water and Air

We consider water on a perfectly flat bottom. The depth of the water at rest is $h_0>0$, measured along the vertical direction x_3. The surface of the water at rest is the plane $x_3=0$ and in general it is described by

$$x_3 = \eta(x_1,x_2,t).$$

We make now three standard assumptions. Firstly it is the kinematic condition that water particles on the surface stay there for all time. Secondly, the dynamic condition says that the interface has no mass and lastly we assume no flow through the bottom at $x_3=-h_0$.

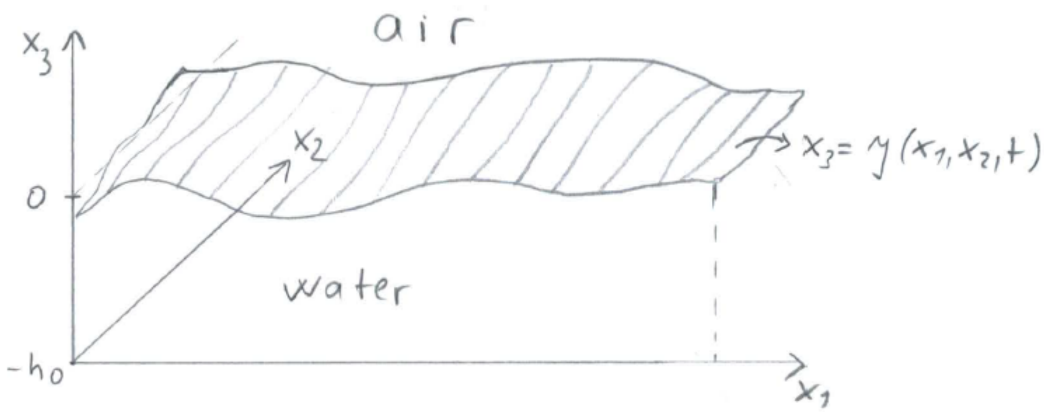

Figure 1.8: A schematic of a free interface between air and water.

1. The kinematic condition, i.e. that water particles on the surface stay on the surface for all time can be formulated as
$$v_F = v_I \quad \text{on } x_3 = \eta(x_1, x_2, t), \tag{1.56}$$
where v_F is the speed of the water in the normal direction to the surface and v_I is the speed of the surface itself in the normal direction. The interface is defined as
$$F(x_1, x_2, x_3, t) := x_3 - \eta(x_1, x_2, t) = 0,$$
i.e. as the level set $F = 0$. The normal n of the interface is thus the gradient of F
$$n = \nabla F = \frac{1}{\sqrt{(\partial_{x_1}\eta)^2 + (\partial_{x_2}\eta)^2 + 1}} \begin{pmatrix} -\partial_{x_1}\eta \\ -\partial_{x_2}\eta \\ 1 \end{pmatrix}.$$
Hence we have
$$v_I = n \cdot \dot{\vec{x}} = n \cdot \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_1\partial_{x_1}\eta + \dot{x}_2\partial_{x_2}\eta + \partial_t\eta \end{pmatrix} = \frac{\partial_t\eta}{\sqrt{(\partial_{x_1}\eta)^2 + (\partial_{x_2}\eta)^2 + 1}}$$
and
$$v_F = n \cdot \vec{u} = \frac{u_3 - u_1\partial_{x_1}\eta - u_2\partial_{x_2}\eta}{\sqrt{(\partial_{x_1}\eta)^2 + (\partial_{x_2}\eta)^2 + 1}}.$$
Condition (1.56) is thus
$$u_3 = \partial_t\eta + u_1\partial_{x_1}\eta + u_2\partial_{x_2}\eta \quad \text{on } x_3 = \eta(x_1, x_2, t). \tag{1.57}$$
Note that (1.57) can be obtained much faster using the material derivative notation since the condition that particles at the interface stay at the interface means $D_tF = 0$, which is precisely (1.57).

2. The dynamic condition states that the interface has no mass so that the forces on either side of the interface are equal, i.e.
$$p_0 - p = 2TH \quad \text{on } x_3 = \eta(x_1, x_2, t), \tag{1.58}$$
where p_0 and p are the pressure of air and water respectively, $2TH$ is the surface tension, T is the coefficient of surface tension and H is the mean curvature of the interface. The mean curvature is given by
$$\begin{aligned} 2H =& |\nabla F|^{-3} \left(\nabla F D^2F(\nabla F)^T - |\nabla F|^2 \text{trace}(D^2F)\right) \\ =& (1 + (\partial_{x_1}\eta)^2 + (\partial_{x_2}\eta)^2)^{-3/2} \left[(1 + (\partial_{x_2}\eta)^2)\partial_{x_1}^2\eta + (1 + (\partial_{x_1}\eta)^2)\partial_{x_2}^2\eta - 2\partial_{x_1}\eta\partial_{x_2}\eta\partial_{x_1}\partial_{x_2}\eta\right]. \end{aligned} \tag{1.59}$$

3. Finally, the condition of no flow through the bottom $x_3 = -h_0$ is simply

$$u_3 = \partial_{x_3}\varphi = 0 \quad \text{on } x_3 = -h_0. \tag{1.60}$$

In summary, combining (1.55), (1.57), (1.58), and (1.60), we have the following system for the free interface problem between an inviscid incompressible irrotational fluid and air

$$\Delta\varphi = 0, \quad -h_0 < x_3 < \eta(x_1, x_2, t), \tag{1.61a}$$

$$\partial_{x_3}\varphi = 0, \quad x_3 = -h_0, \tag{1.61b}$$

$$\partial_t\varphi + \frac{1}{2}|\nabla\varphi|^2 + g\eta - \frac{T}{\rho}\frac{(1+(\partial_{x_2}\eta)^2)\partial_{x_1}^2\eta + (1+(\partial_{x_1}\eta)^2)\partial_{x_2}^2\eta - 2\partial_{x_1}\eta\partial_{x_2}\eta\partial_{x_1}\partial_{x_2}\eta}{(1+(\partial_{x_1}\eta)^2+(\partial_{x_2}\eta)^2)^{3/2}} = 0, \quad x_3 = \eta(x_1, x_2, t), \tag{1.61c}$$

$$\partial_t\eta + \partial_{x_1}\varphi\partial_{x_1}\eta + \partial_{x_2}\varphi\partial_{x_2}\eta = \partial_{x_3}\varphi, \quad x_3 = \eta(x_1, x_2, t). \tag{1.61d}$$

Equation (1.61c) is called the Bernoulli equation. In many case the surface tension can be neglected, as, for instance, for long waves, where the mean curvature is very small. Then equation (1.61c) reduces to

$$\partial_t\varphi + \frac{1}{2}|\nabla\varphi|^2 + g\eta = 0, \quad x_3 = \eta(x_1, x_2, t). \tag{1.62}$$

1.5.1 Linear Theory

We study here small solutions of (1.61) so that only the linearization of the system is important. For the linearized system we derive the dispersion relation and explore it in several interesting asymptotic regimes.

1.5.1.1 Without Surface Tension

Let us first consider the case without surface tension. Water waves are then often called *gravity waves* since the only restoring force is gravity. It can be shown that the linearization of the system (1.61a), (1.61b), (1.61d) and (1.62) is

$$\Delta\varphi = 0, \quad -h_0 < x_3 < 0, \tag{1.63a}$$

$$\partial_{x_3}\varphi = 0, \quad x_3 = -h_0, \tag{1.63b}$$

$$\partial_t^2\varphi + g\partial_{x_3}\varphi = 0, \quad x_3 = 0, \tag{1.63c}$$

$$\eta + \frac{1}{g}\partial_t\varphi = 0, \quad x_3 = 0. \tag{1.63d}$$

We consider plane waves propagating in the horizontal directions $(x_1, x_2, 0)$, i.e. we set

$$\eta = Ae^{\mathrm{i}(k_1x_1+k_2x_2-\omega t)} + \text{c.c.}, \quad \varphi = Y(x_3)e^{\mathrm{i}(k_1x_1+k_2x_2-\omega t)} + \text{c.c.} \tag{1.64}$$

with a constant $A \in \mathbb{C}$ and a scalar function $Y : [-h_0, \infty) \to \mathbb{C}$ and where c.c. stands for the complex conjugate of the expression preceeding it.

From (1.63a), (1.63b) we get

$$Y''(x_3) - |k|^2 Y(x_3) = 0 \quad \text{for } -h_0 < x_3 < 0, \quad Y'(-h_0) = 0$$

such that $Y(x_3) = C\cosh(|k|(x_3+h_0))$, $C \in \mathbb{C}$. Here $|k|$ denotes the Euclidean norm of the vector $k = (k_1, k_2)$. Equation (1.63d) implies $A = \frac{\mathrm{i}\omega}{g}Y(0)$ and thus

$$Y(x_3) = \frac{-\mathrm{i}g}{\omega}A\frac{\cosh(|k|(x_3+h_0))}{\cosh(|k|h_0)}.$$

With the expression for Y we can recover from equation (1.63c) the *dispersion relation for water waves without surface tension*

$$\omega^2 = g|k|\tanh(|k|h_0). \tag{1.65}$$

The solutions of (1.65) have the form

$$\omega = \pm W(k) := \pm U(|k|), \text{ where } U(s) = \sqrt{gs\tanh(sh_0)}.$$

The group velocity $v_g(k) = \pm\nabla W(k)$ has the form

$$v_g(k) = \pm U'(|k|)\frac{k}{|k|}$$

such that $|v_g(k)| = |U'(|k|)|$. One can see that $|v_g(k)|$ is decreasing in $|k|$, i.e. shorter waves are slower.

Let us now look at two asymptotic cases of the horizontal waves.

(a) **Deep water waves**: $|k|h_0 \gg 1$

In this asymptotic regime the wavelength is much smaller than the depth (but still large enough such that surface tension can be neglected). We obtain

$$W(k) \sim \sqrt{g|k|} \ (|k|h_0 \to \infty)$$

such that

$$|v_p(k)| \sim \left(\frac{g}{|k|}\right)^{1/2}, \quad |v_g(k)| \sim \frac{1}{2}\left(\frac{g}{|k|}\right)^{1/2}.$$

This regime is clearly dispersive and, as mentioned above, longer waves propagate faster.

A physical example of deep water gravity waves are *wind generated waves* in the ocean. These have the typical wavelength λ of 60 to 150 m. With a typical depth of $h_0 = 4000\text{m}$ we get $|k|h_0 \sim \frac{2\pi h_0}{\lambda}$ approximately between 65 ad 168, i.e. values much larger than 1. Another example is waves generated by explosions, as shown in Fig. 1.9(a). In the figure one clearly observes that longer waves have traveled further than shorter ones.

(b) **Shallow water waves**: $|k|h_0 \ll 1$

Here

$$W(k) \sim \left(g|k|(h_0|k| - \tfrac{1}{3}(|k|h_0)^3)\right)^{1/2} \sim \sqrt{gh_0}|k| \ (|k|h_0 \to 0).$$

Hence, in the asymptotics the dispersion relation is equivalent to that of the wave equation

$$\partial_t^2\eta - gh_0(\partial_{x_1}^2\eta + \partial_{x_2}^2\eta) = 0.$$

The phase and group velocities are

$$|v_p(k)| \sim \sqrt{gh_0}, \quad |v_g(k)| \sim \sqrt{gh_0} \ (|k|h_0 \to 0)$$

such that the problem is not dispersive in this asymptotic regime.

Though it may sound surprising, an example of shallow water waves are *tsunamis* in the free ocean. This is not due a small depth but rather due to the large wavelength, which is typically about 200 km. By a depth of 4km we get $|k|h_0 \approx 8\pi/200 \ll 1$. One can very easily approximate the propagation speed of such a tsunami wave

$$|v_g(k)| \sim \sqrt{gh_0} \approx \sqrt{40000}\text{m/s} = 200\text{m/s} = 720\text{km/h!}$$

1.5.1.2 With Surface Tension

In the presence of surface tension the full system (1.61) has to be linearized. The linearization of the mean curvature in (1.59) is $2H = \partial_{x_1}^2\eta + \partial_{x_2}^2\eta$ and we obtain

$$\Delta\varphi = 0, \quad -h_0 < x_3 < 0, \tag{1.66a}$$

$$\partial_{x_3}\varphi = 0, \quad x_3 = -h_0, \tag{1.66b}$$

$$\partial_t\varphi + g\eta - \frac{T}{\rho}(\partial_{x_1}^2\eta + \partial_{x_2}^2\eta) = 0, \quad x_3 = 0, \tag{1.66c}$$

$$\partial_t\eta - \partial_{x3}\varphi = 0, \quad x_3 = 0. \tag{1.66d}$$

Considering again horizontal waves (1.64), we get from (1.63a), (1.63b) once again $Y(x_3) = C\cosh(|k|(x_3 + h_0))$. Equation (1.66d) now implies $-\mathrm{i}\omega A = Y'(0) = c|k|\sinh(|k|(x_3 + h_0))$ such that

$$Y(x_3) = -\mathrm{i}\omega A\frac{\cosh(|k|(x_3 + h_0))}{|k|\sinh(|k|h_0)}.$$

Finally, from (1.66c) we get the *dispersion relation for water waves with surface tension*

$$\omega^2 = g|k|\left(1 + \frac{T}{\rho g}|k|^2\right)\tanh(|k|h_0). \tag{1.67}$$

Note that for $|k| \ll 1$ this dispersion relation reduces to (1.65) as expected since long waves have small mean curvature. Surface tension is thus relevant for $|k| \gg 1$ and for intermediate values of $|k|$. First we take a look at the case of **capillary waves** $|k| \gg 1$. Here surface tension is the dominant restoring force. A typical physical example are waves generated by rain drops.

One obtains

$$\omega^2(k) \sim \frac{T}{g}|k|^3\tanh(|k|h_0) \quad (|k| \to \infty) \tag{1.68}$$

such that

$$|v_g(k)| \sim \frac{3}{2}\left(\frac{T}{\rho}|k|\tanh(h_0|k|)\right)^{1/2}\left(1 + \frac{2h_0|k|}{3\sinh(2h_0|k|)}\right).$$

Clearly, for $|k|$ large the norm of $v_g(k)$ grows with $|k|$ such that shorter waves are faster, see Fig. 1.9 (b) for an example of a capillary wave generated by a water drop. The faster propagation of shorter waves is well visible.

Also for capillary waves we can study the two asymptotic scenarios of deep and shallow water, i.e. $|k|h_0 \gg 1$ and $|k|h_0 \ll 1$ respectively.

(a) **Deep water capillary waves**: $|k|h_0 \gg 1$

When $|k|h_0 \gg 1$ (and $|k| \gg 1$), we get from (1.68) the asymptotics $W(|k|) \sim \left(\frac{T}{\rho}\right)^{1/2}|k|^{3/2}$ for ($|k|h_0 \to \infty$) and

$$|v_p(k)| \sim \left(\frac{T}{\rho}\right)^{1/2}|k|^{1/2}, \quad |v_g(k)| \sim \frac{3}{2}\left(\frac{T}{\rho}\right)^{1/2}|k|^{1/2} \qquad (|k|h_0 \to \infty).$$

(b) **Shallow water capillary waves**: $|k|h_0 \ll 1$

When $|k|h_0 \ll 1$ (and $|k| \gg 1$), we get from (1.68) the asymptotics $W(|k|) \sim \left(\frac{Th_0}{\rho}\right)^{1/2}|k|^2$ for ($|k|h_0 \to 0$)

$$|v_p(k)| \sim 2\left(\frac{Th_0}{\rho}\right)^{1/2}|k|, \quad |v_g(k)| \sim 2\left(\frac{Th_0}{\rho}\right)^{1/2}|k| \qquad (|k|h_0 \to 0).$$

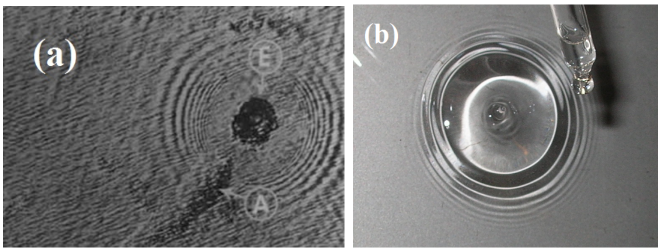

Figure 1.9: (a) deep water gravity wave from an underwater explosion at point E; (b) capillary waves generated by a drop of water. Picture in (a) is from [22]. It shows the effect of an Italian airplane trying to bomb the United, a British submarine at A, creating an explosion at E; (b) is a photo of a class room experiment.

Finally, for the intermediate values of $|k|$ in (1.65), one needs to analyze this full dispersion relation. An interesting observation is that in the deep water asymptotics $|k|h_0 \to \infty$ there is a wavenumber, namely

$$|k| = \left(\frac{2-\sqrt{3}}{\sqrt{3}}\right)^{1/2} \left(\frac{\rho}{T} g\right)^{1/2},$$

at which the group velocity is minimized. For water at 15 °C the surface tension coefficient is $T \approx 0.073 N/m$ and density $\rho \approx 1000 kg/m^3$. The resulting wavelength $\lambda = \frac{2\pi}{|k|}$ is about 4.36 cm.

Chapter 2

Nonlinear Waves

Under nonlinear waves we understand simply waves described by nonlinear PDEs. We will restrict our attention to those PDEs whose linear part is dispersive. One of the important properties of nonlinear dispersive probems which distinguishes it from linear dispersive ones is the possibility of coherent propagation of localized waves, i.e. waves whose spatially localized shape is invariant under the time evolution. In other words, in nonlinear problems the dispersion induced destruction of pulses can be arrested. This occurs when the shape of the solution achieves a perfect balance between dispersion and nonlinearity. There are a number of possible effects of nonlinearity. The most common ones are focusing, e.g in the nonlinear Schrödinger equation, and steepening, e.g. in the Korteweg-de Vries equation.

Let us list a few classical but also physically highly relevant nonlinear dispersive models.

1. The Korteweg-de Vries (KdV) equation

$$\partial_t u + 6u\partial_x u \pm \partial_x^3 u = 0, \quad x \in \mathbb{R}, t > 0$$

describes, e.g., one dimensional shallow water waves of relatively small amplitude as well as long waves in the Fermi-Pasta-Ulam problem. A formal derivation of KdV in these contexts is carried out in Sections 2.1.1 and 2.1.2.

2. The nonlinear Schrödinger equation

$$\mathrm{i}\partial_z u - \alpha\partial_t^2 u + \gamma|u|^2 u = 0, \quad t \in \mathbb{R}, z > 0, \alpha, \gamma \in \mathbb{R}$$

is a description of optical pulses in Kerr nonlinear photonic fibers with the longitudinal direction denoted by z. It is, however, also a universal asymptotic model for slowly varying wave packets of small amplitude in rather general nonlinear dispersive problems, which we explain in Section 2.2.1.

3. The Gross-Pitaevskii equation

$$\mathrm{i}\partial_t u + \frac{\hbar^2}{2m}\Delta u - V(x)u + \gamma|u|^2 u = 0, \quad x \in \mathbb{R}^n, t > 0$$

for $n \in \{1, 2, 3\}$ models Bose-Einstein condensates (BECs) in an external potential described by V. BECs are systems of identical bosons in the same quantum mechanical state. $\hbar$ is the Planck's constant and m is the mass of the boson. For $n = 1, 2$ the same mathematical form of the equation is also an asymptotic model for slowly modulated optical beams in photonic crystals.

4. The Sine-Gordon equation

$$\partial_t^2 u - \partial_x^2 u + \sin(u) = 0, \quad x \in R, t > 0$$

is a model of the Josephson effect, i.e. the tunneling electric current between two superconductors.

5. The Coupled Mode Equations

$$\mathrm{i}(\partial_t u + \partial_z u) + \kappa v + \Gamma(|u|^2 + 2|v|^2)u = 0$$
$$\mathrm{i}(\partial_t v - \partial_z v) + \kappa u + \Gamma(|v|^2 + 2|u|^2)v = 0$$

for $z \in \mathbb{R}, t > 0, \kappa, \Gamma \in \mathbb{R}$ are an asymptotic description of pulses in optical fibers with a Bragg grating, i.e. a specific periodic structure in the longitudinal direction z.

2.1 Korteweg-de Vries Equation

2.1.1 Korteweg-de Vries Equation for Shallow Water Waves

Here we carry out a formal asymptotic derivation of KdV for one dimensional shallow water waves along the lines of Chapter 5 in [1].

Shallow water waves have an especially important place in the history of understanding nonlinear dispersive waves. This is due to the first scientific record of a solitary wave done by Sir John Scott Russell in 1834 on a canal near Edinburgh, where he observed a wave generated by a ship which came to a sudden halt. The wave continued traveling in the direction of the ship's movement and did not change its shape over a large distance [20]. Here are his own words

I was observing the motion of a boat which was rapidly drawn along a narrow channel by a pair of horses, when the boat suddenly stopped-not so the mass of water in the channel which it had put in motion; it accumulated round the prow of the vessel in a state of violent agitation, then suddenly leaving it behind, rolled forward with great velocity, assuming the form of a large solitary elevation, a rounded, smooth and well-defined heap of water, which continued its course along the channel apparently without change of form or diminution of speed. I followed it on horseback, and overtook it still rolling on at a rate of some eight or nine miles an hour [14 km/h], preserving its original figure some thirty feet [9 m] long and a foot to a foot and a half [300-450 mm] in height. Its height gradually diminished, and after a chase of one or two miles [2-3 km] I lost it in the windings of the channel. Such, in the month of August 1834, was my first chance interview with that singular and beautiful phenomenon which I have called the Wave of Translation.

Let us recall system (1.61a), (1.61b), (1.61d) and (1.62) for the free interface of water and air under the assumptions of incompressibility, no viscosity, no surface tension and an irrotational flow. We consider only waves propagating solely in the x_1-direction with a profile that is independent of x_2. Such waves are possible, for instance, in a narrow channel of uniform width. Hence, we set

$$\varphi = \varphi(x_1, x_3, t) \quad \text{and} \quad \eta = \eta(x_1, t).$$

The first step in the derivation is to make the equations dimensionless. This step is important for the subsequent asymptotic analysis where orders of magnitude of terms have to be identified. A choice of units can, of course, artificially change the size of a given term.

Let us consider waves with a typical wavelength λ and a typical amplitude a. We define the two numbers

$$\varepsilon_1 := \frac{h_0}{\lambda}, \quad \varepsilon_2 := \frac{a}{h_0}.$$

Later we will assume smallness of $\varepsilon_{1,2}$ but at the moment these are simply new names.

Next, we define the following dimensionless (primed) variables

$$x_3' := \tfrac{x_3}{h_0}, \qquad x_1' := \tfrac{x_1}{\lambda} = \tfrac{\varepsilon_1}{h_0}x_1, \qquad t' := \tfrac{\sqrt{gh_0}}{\lambda}t = \varepsilon_1\sqrt{\tfrac{g}{h_0}}t$$
$$\eta' := \tfrac{\eta}{a} = \tfrac{\eta}{\varepsilon_2 h_0}, \quad \varphi' := \tfrac{\sqrt{gh_0}}{\lambda g a}\varphi = \tfrac{\varepsilon_1}{\varepsilon_2 h_0\sqrt{gh_0}}\varphi.$$

After a simple calculation one finds that in dimensionless form system (1.61a), (1.61b), (1.61d) and (1.62) for one dimensional waves reads

$$\begin{aligned}\varepsilon_1^2\partial_{x_1'}^2\varphi' + \partial_{x_3'}^2\varphi' &=0, \quad -1 < x_3' < \varepsilon_2\eta',\\ \partial_{x_3'}\varphi' &=0, \quad x_3' = -1,\\ \partial_{t'}\varphi' + \frac{\varepsilon_2}{2}\left[(\partial_{x_1'}\varphi')^2 + \frac{1}{\varepsilon_1^2}(\partial_{x_3'}\varphi')^2\right] + \eta' &=0, \quad x_3' = \varepsilon_2\eta',\\ \varepsilon_1^2\left(\partial_{t'}\eta' + \varepsilon_2\partial_{x_1'}\varphi'\partial_{x_1'}\eta'\right) &=\partial_{x_3'}\varphi', \quad x_3' = \varepsilon_2\eta'.\end{aligned}$$

Now we make our asymptotic assumptions. Namely, we consider *shallow water* such that $\varepsilon_1 \ll 1$ and waves with *small amplitude*: $\varepsilon_2 \ll 1$. In order to balance the effects of as many terms as possible ("Kruskal's principle of maximal balance"), we choose the relation

$$\varepsilon_2 = \varepsilon_1^2 =: \varepsilon.$$

With this choice the leading order parts of all the $o(1)$ terms (as $\varepsilon_1, \varepsilon_2 \to 0$) are of the same order.

With this definition of ε and dropping the primes, we obtain

$$\varepsilon\partial_{x_1}^2\varphi + \partial_{x_3}^2\varphi =0, \quad -1 < x_3 < \varepsilon\eta, \tag{2.1a}$$

$$\partial_{x_3}\varphi =0, \quad x_3 = -1, \tag{2.1b}$$

$$\partial_t\varphi + \frac{\varepsilon}{2}(\partial_{x_1}\varphi)^2 + \frac{1}{2}(\partial_{x_3}\varphi)^2 + \eta =0, \quad x_3 = \varepsilon\eta, \tag{2.1c}$$

$$\varepsilon\left(\partial_t\eta + \varepsilon\partial_{x_1}\varphi\partial_{x_1}\eta\right) =\partial_{x_3}\varphi, \quad x_3 = \varepsilon\eta. \tag{2.1d}$$

Let us attempt to find an approximation of φ via the regular perturbation ansatz

$$\varphi = \varphi_0 + \varepsilon\varphi_1 + \varepsilon^2\varphi_2 + \dots.$$

At $O(1)$ we have from equation (2.1a)

$$\partial_{x_3}^2\varphi_0 = 0, \quad -1 < x_3 < \varepsilon\eta$$

resulting in

$$\varphi_0 = A + B(x_3+1) \quad \text{with arbitrary functions} \quad (A,B) = (A,B)(x_1,t).$$

From the boundary condition (2.1b) we get $B = 0$ such that $\varphi_0(x_1,x_3,t) = A(x_1,t)$.

At $O(\varepsilon)$ we obtain

$$\partial_{x_1}^2\varphi_0 + \partial_{x_3}^2\varphi_1 = 0, \quad -1 < x_3 < \varepsilon\eta,$$

i.e. $\partial_{x_3}^2\varphi_1 = -\partial_{x_1}^2 A$. Using (2.1b), the solution reads $\varphi_1 = -\frac{(x_3+1)^2}{2}\partial_{x_1}^2 A + D$ with an arbitrary $D = D(x_1,t)$. The "constant" D can be absorbed in A because for $A = O(1)$ is also $A + \varepsilon D = O(1)$. Hence

$$\varphi_1 = -\frac{(x_3+1)^2}{2}\partial_{x_1}^2 A.$$

The last order, which we consider, is $O(\varepsilon^2)$. It consists of

$$\partial_{x_1}^2\varphi_1 + \partial_{x_3}^2\varphi_2 = 0, \quad -1 < x_3 < \varepsilon\eta,$$

which leads to

$$\varphi_2 = \frac{(x_3+1)^4}{4!}\partial_{x_1}^4 A$$

after, once again, applying (2.1b) and absorbing the integration constant in A.

In conclusion, we formally have

$$\varphi = A - \varepsilon\frac{(x_3+1)^2}{2}\partial_{x_1}^2 A + \varepsilon^2\frac{(x_3+1)^4}{4!}\partial_{x_1}^4 A + O(\varepsilon^3), \quad -1 < x_3 < \varepsilon\eta$$

with $A = A(x_1, t)$.

Next we use equations (2.1c) and (2.1d) to provide an expansion for η and an equation for the currently free function A.

From (2.1c) we get

$$\partial_t A - \frac{\varepsilon}{2}(1+\varepsilon\eta)^2\partial_{x_1}^2\partial_t A + \frac{\varepsilon}{2}\left(\partial_{x_1} A - \frac{\varepsilon}{2}(1+\varepsilon\eta)^2\partial_{x_1}^3 A\right)^2 + \frac{1}{2}\left(\varepsilon(1+\varepsilon\eta)\partial_{x_1}^2 A\right)^2 + \eta = O(\varepsilon^2)$$

and hence

$$\eta = -\partial_t A + \frac{\varepsilon}{2}\left(\partial_{x_1}^2\partial_t A - (\partial_{x_1} A)^2\right) + O(\varepsilon^2). \tag{2.2}$$

Finally, (2.1d) produces

$$\partial_t^2 A - \partial_{x_1}^2 A = \varepsilon\left[\frac{1}{2}\partial_{x_1}^2\partial_t^2 A - 2\partial_{x_1} A\partial_{x_1}\partial_t A - \frac{1}{6}\partial_{x_1}^4 A - \partial_{x_1}^2\partial_t A\right] + O(\varepsilon^2). \tag{2.3}$$

Remark 2.1. *It is interesting to note that the linear part of equation* (2.3) *is up to* $O(\varepsilon^2)$ *equivalent to the Boussinesq equation* $\partial_t^2 A - \partial_{x_1}^2 A = \frac{\varepsilon}{3}\partial_{x_1}^2\partial_t^2 A$*. This follows by applying the equation itself, i.e.* $\partial_{x_1}^2 A = \partial_t^2 A + O(\varepsilon)$*, to replace* $\partial_{x_1}^4 A = \partial_{x_1}^2(\partial_{x_1}^2 A)$ *by* $\partial_{x_1}^2\partial_t^2 A + O(\varepsilon)$*.*

The last step of the asymptotic discussion is to provide an ε-independent problem from (2.3). At the leading order equation (2.3) is the wave equation. Hence, we expect that A depends on the characteristic variables $x_1 + t$ and $x_1 - t$. We make a multiscale asymptotic ansatz for A

$$A(x_1, t) = A_0(\xi, \nu, T) + \varepsilon A_1(\xi, \nu, T) + O(\varepsilon^2), \tag{2.4}$$

where $\xi := x_1 - t, \nu := x_1 + t$, and $T := \varepsilon t$. This transformation of variables leads to

$$\begin{aligned}
\partial_{x_1} &= \partial_\xi + \partial_\nu, \quad \partial_{x_1}^2 = \partial_\xi^2 + \partial_\nu^2 + 2\partial_\xi\partial_\nu,\\
\partial_t &= \partial_\nu - \partial_\xi + \varepsilon\partial_T, \quad \partial_t^2 = \partial_\xi^2 + \partial_\nu^2 - 2\partial_\xi\partial_\nu + 2\varepsilon(\partial_\nu\partial_T - \partial_\xi\partial_T) + \varepsilon^2\partial_T^2,\\
\partial_{x_1}\partial_t &= \partial_\nu^2 - \partial_\xi^2 + \varepsilon(\partial_T\partial_\xi + \partial_T\partial_\nu).
\end{aligned}$$

Inserting ansatz (2.4) in (2.3) yields at $O(1)$

$$-4\partial_\xi\partial_\nu A_0 = 0$$

such that

$$A_0 = F(\xi, T) + G(\nu, T)$$

for arbitrary functions F and G. At $O(\varepsilon)$ we obtain

$$\begin{aligned}
-4\partial_\xi\partial_\nu A_1 + 2\left(\partial_\nu\partial_T - \partial_\xi\partial_T\right)A_0 = &\frac{1}{2}(\partial_\xi + \partial_\nu)^2(\partial_\xi - \partial_\nu)^2 A_0 - \frac{1}{6}(\partial_\xi + \partial_\nu)^4 A_0\\
&- 2(\partial_\xi + \partial_\nu)A_0(\partial_\nu^2 - \partial_\xi^2)A_0 - (\partial_\xi + \partial_\nu)^2 A_0(\partial_\nu - \partial_\xi)A_0.
\end{aligned}$$

Using $\partial_\xi\partial_\nu A_0 = 0$ this reduces to

$$-4\partial_\xi\partial_\nu A_1 = \frac{1}{3}(\partial_\xi^4 F + \partial_\nu^4 G) - \partial_\xi F\partial_\nu^2 G + \partial_\nu G\partial_\xi^2 F + 3\partial_\xi F\partial_\xi^2 F - 3\partial_\nu G\partial_\nu^2 G + 2(\partial_\xi\partial_T F - \partial_\nu\partial_T G).$$

Due to the linearity of the problem we can write $A_1 = A_1^{(a)} + A_1^{(b)} + A_1^{(c)}$, where

$$\begin{aligned}
-4\partial_\xi\partial_\nu A_1^{(a)} =& -\partial_\xi F\partial_\nu^2 G + \partial_\nu G\partial_\xi^2 F,\\
-4\partial_\xi\partial_\nu A_1^{(b)} =& 2\partial_\xi\partial_T F + \frac{1}{3}\partial_\xi^4 F + 3\partial_\xi F\partial_\xi^2 F =: P_1,\\
-4\partial_\xi\partial_\nu A_1^{(c)} =& -2\partial_\nu\partial_T G + \frac{1}{3}\partial_\nu^4 G - 3\partial_\nu G\partial_\nu^2 G =: P_2.
\end{aligned}$$

These equations can be easily integrated to give

$$A_1^{(a)} = \frac{1}{4}(F\partial_\nu G - G\partial_\xi F) + \alpha(\xi, T) + \beta(\nu, T), \tag{2.5}$$

$$A_1^{(b)} = -\frac{1}{4}\left[2\partial_T F + \frac{1}{3}\partial_\xi^3 F + \frac{3}{2}(\partial_\xi F)^2\right]\nu, \tag{2.6}$$

$$A_1^{(c)} = -\frac{1}{4}\left[-2\partial_T G + \frac{1}{3}\partial_\nu^3 G - \frac{3}{2}(\partial_\nu G)^2\right]\xi \tag{2.7}$$

with arbitrary integration constants $\alpha(\xi, T)$ and $\beta(\nu, T)$. Note that it is not necessary to include integration constants in $A_1^{(b)}$ and $A_1^{(c)}$ since only the sum $A_1^{(a)} + A_1^{(b)} + A_1^{(c)}$ is relevant for us. In fact, since the initial data for φ and η have not been prescribed, we are free to choose $\alpha(\xi, T)$ and $\beta(\nu, T)$. For a localized solution with $\varphi(x_1, t), \eta(x_1, t) \to 0$ as $|x_1| \to \infty$ we actually need $\alpha = \beta = 0$ since otherwise $A_1^{(a)}$ does not decay as $|x_1| \to \infty$.

Although the solutions φ and η are unknown, we look for bounded solutions which requires A to be bounded in x_1. Clearly, because F is independent of ν and G is independent of ξ, the above functions $A_1^{(b)}$ and $A_1^{(c)}$ are unbounded (grow linearly) unless the expressions in the squared brackets in (2.6) and (2.7) vanish. Polynomially growing terms in asymptotic expansions of bounded solutions of differential equations are called *secular terms*. The occurrence of these secular terms is no surprise since the right hand sides P_1 and P_2 lie in the kernel of the adjoint of the operator $\partial_\xi\partial_\nu$ on the left hand side. In other words P_1 and P_2 are in resonance with the homogeneous solution. While setting the squared brackets equal zero would provide for a bounded expansion (up to $O(\varepsilon)$), the same can be achieved by setting

$$P_1 = P_2 = 0,$$

which is the more traditional approach. It is also formally a weaker assumption because it is an assumption on derivatives of the squared brackets. Defining then

$$U := \partial_\xi F \text{ and } V := \partial_\nu G,$$

we get the Korteweg-de Vries (KdV) equations

$$2\partial_T U + \frac{1}{3}\partial_\xi^3 U + 3U\partial_\xi U = 0, \tag{2.8a}$$

$$2\partial_T V - \frac{1}{3}\partial_\nu^3 V + 3V\partial_\nu V = 0, \tag{2.8b}$$

where the equation for U is that for right propagating waves and that for V is for left propagating ones.

A suitable scaling

$$u(\xi, T) := \alpha U(\beta\xi, \gamma T),\ \ v(\xi, T) = \alpha V(\beta\xi, \gamma T)$$

for some $\alpha, \beta, \gamma > 0$ transforms (2.8) to the standard KdV equations

$$\partial_T u + 6u\partial_\xi u + \partial_\xi^3 u = 0, \tag{2.9a}$$

$$\partial_T v + 6v\partial_\xi v - \partial_\xi^3 v = 0. \tag{2.9b}$$

To obtain an approximation of η, we can use (2.2) to get

$$\eta(x_1,t) = -\partial_t A_0(\xi,\nu,T) + O(\varepsilon) = -\partial_t F(\xi,t) - \partial_t G(\nu,t) + O(\varepsilon).$$

Due to $\partial_t = \partial_\nu - \partial_\xi + \varepsilon\partial_T$ we have

$$\eta = \partial_\xi F - \partial_\nu G + O(\varepsilon) = U - V + O(\varepsilon). \tag{2.10}$$

The above derivation of the KdV equations is, of course, formal and does not guarantee that the truncated expansion of φ and η, with U and V given by solutions of KdV, is close (in a suitable norm) to a true solution (η,φ) of (2.1) for ε small. For a rigorous justification of KdV for shallow water waves see e.g. the book [13]. Below in Sec. 2.2.2 we address this justification question in detail for a simpler case, namely for slowly modulated wavepackets in the 1D nonlinear wave equation. In this case the envelope is shown to be effectively described by the nonlinear Schrödinger equation.

2.1.1.1 KdV in Dimensional Form

Let us now look at the KdV in the dimensional variables. Restricting to the right propagating waves, we have from (2.10) the approximation $\eta = U + O(\varepsilon)$. Going back to the dimensional variables (denoted once again by the same letters), we get then from (2.8a) the dimensional KdV equation

$$\frac{1}{\sqrt{gh_0}}\partial_t\eta + \partial_{x_1}\eta + \frac{h_0^2}{6}\partial_{x_1}^3\eta + \frac{3}{2h_0}\eta\partial_{x_1}\eta = 0. \tag{2.11}$$

The linear part of (2.11) can be also obtained directly by expanding the one dimensional version (i.e. $k = k_1$) of the dispersion relation (1.65) (which becomes $\omega^2 = gk_1\tanh(k_1h_0)$) in the small parameter $|k_1h_0| \ll 1$. We leave this as a simple exercise.

2.1.1.2 KdV for Water Waves with Surface Tension

When surface tension is taken into account, the full form of the Bernoulli equation (1.61c) needs to be used. Considering, however, only one dimensional waves, we obtain

$$\partial_t\varphi + \frac{1}{2}\left((\partial_{x_1}\varphi)^2 + (\partial_{x_3}\varphi)^2\right) + g\eta - \frac{T}{\rho}\frac{\partial_{x_1}^2\eta}{\left(1+(\partial_{x_1}\eta)^2\right)^{3/2}} = 0, \quad x_3 = \eta(x_1,t).$$

An analogous asymptotic calculation to that for the zero surface tension case produces the following dimensional KdV for right propagating waves

$$\frac{1}{\sqrt{gh_0}}\partial_t\eta + \partial_{x_1}\eta + \gamma\partial_{x_1}^3\eta + \frac{3}{2h_0}\eta\partial_{x_1}\eta = 0, \quad \gamma = \frac{h_0^2}{6} - \frac{T}{2\rho g}. \tag{2.12}$$

The qualitatively new effect compared to (2.11) is that the coefficient γ of the dispersive term becomes negative for large enough values of the surface tension coefficient T.

The dimensionless form of (2.12) is

$$\partial_T u + 6su\partial_\xi u + \partial_\xi^3 u = 0, \quad s = \text{sign}(\gamma). \tag{2.13}$$

2.1.1.3 Solitary Waves of the KdV

The KdV is one of the equations which support solitary wave solutions.

Definition 2.1. *A solitary wave is a localized wave of constant shape.*

Let us thus search for a solution $u(\xi,T) = f(y), y := \xi - cT, c \in \mathbb{R}$. For the localization we require $f(\xi), f'(\xi), f''(\xi) \to 0$ for $|\xi| \to \infty$. This ansatz produces in (2.13)

$$-cf' + 6sff' + f''' = 0.$$

After an integration (from $-\infty$ to ξ), multiplication by f' and another integration we get

$$(f')^2 = f^2(c - 2sf). \tag{2.14}$$

Note that the integration constants vanish due to the decay assumptions.

Since the left hand side in (2.14) is non-negative and since $f(\xi) \to 0$ for $|\xi| \to \infty$, we conclude that

$$c \geqslant 0.$$

The integration of (2.14) yields

$$\int \frac{df}{f(c-2sf)^{1/2}} = \pm(y - x_0)$$

with an arbitrary $x_0 \in \mathbb{R}$. With the substitution $f = s\frac{c}{2}\text{sech}^2\theta$ and using $(c-2sf)^{1/2} = \sqrt{c}|\tanh(\theta)|$, we get

$$-\frac{2}{\sqrt{c}}\int \frac{\tanh(\theta)}{|\tanh(\theta)|}d\theta = \pm(y - x_0)$$

such that

$$|\theta| = \pm\sqrt{c}(y - x_0).$$

As a result, because sech^2 is even , we have $f(y) = s\frac{c}{2}\text{sech}^2(\sqrt{c}(y - x_0))$. In conclusion

$$u(\xi,T) = s\frac{c}{2}\text{sech}^2(\sqrt{c}(\xi - cT - x_0)), \quad c \geqslant 0, x_0 \in \mathbb{R}. \tag{2.15}$$

Clearly, for $s = 1$ the solution (2.15) is positive, a so called *wave of elevation*, cf., e.g., the "great wave of translation" of J.S.Russell [20], which was first observed (and documented) in water in 1834. For $s = -1$ the solution is negative and it is called a *wave of depression*. It was first observed only in 2002 in mercury, the surface tension of which is large enough to produce $\gamma < 0$, see [10].

While (2.15) is the only solution of the KdV (2.13) which fits the ansatz $u(\xi,T) = f(y), y = \xi - cT, c \in \mathbb{R}$ with f localized, the KdV has infinitely many other solutions which consist of traveling localized components. These are the famous N-*solitons* which consist of $N \in \mathbb{N}$ pulses, each of which propagates like a solitary wave with a distinct velocity and when these pulses collide, they are not destroyed but rather continue propagating with the same shape and velocity. The only footprint of the interaction is a spatial shift and a shift of the phase of the pulses. Solution (2.15) is the special and rather boring $1-$soliton. The existence of solitons and their properties can be obtained via the technique of the inverse scattering transformation [6, 2].

2.1.2 The Fermi-Pasta-Ulam Problem and the Korteweg-de Vries Equation

2.1.2.1 The Fermi-Pasta-Ulam Problem

The Fermi-Pasta-Ulam (FPU) Problem is of historical importance in the area of nonlinear dispersive problems and soliton theory. In the years 1954-55 Enrico Fermi, John Pasta, and Stanislaw Ulam [11] at the Los Alamos Laboratories performed numerical experiments simulating a chain of particles coupled by strings with a nonlinear restoring force. Their aim was to understand the dynamics of a crystalline atomic structure towards a thermal equilibrium.

Consider a chain of $N \in \mathbb{N}$ particles (along a line) with identical mass m coupled by springs. Denoting by y_n the displacement of the $n-$th particle (ordered from left to right), Newton's second law says

$$m\frac{d^2y_n}{dt^2} = F_R(n) - F_L(n), \; n \in \{1, \ldots, N\}$$

where $F_{R,L}(n)$ is the force acting on the $n-$th particle from the right/left respectively.

Before we discuss the FPU problem, where $F_{R,N}$ are nonlinear functions of $\vec{y}$, let us recall the linear problem. A classical model of *linear springs* is *Hooke's law*, which states

$$F_R(n) = -K(y_n - y_{n+1}),\ F_L(n) = -K(y_{n-1} - y_n) \quad \text{with some } K > 0.$$

This leads to the system

$$m\frac{d^2 y_n}{dt^2} = K(y_{n+1} - 2y_n + y_{n-1}),\ n \in \{1, \dots, N\}, \tag{2.16}$$

i.e.

$$\frac{d^2\vec{y}}{dt^2} = A\vec{y}, \quad A = \frac{K}{m}\begin{pmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -2 \end{pmatrix} \in \mathbb{R}^{N\times N},$$

where $\vec{y} = (y_1, y_2, \dots, y_N)^T$. System (2.16) has $2N$ linearly independent solutions (modes) $e^{\pm i\omega_j t}\xi^{(j)}, j \in \{1, \dots, N\}$, where $(\omega_j, \xi^{(j)})$ are eigenpairs of A

$$A\xi^{(j)} = \omega_j \xi^{(j)}.$$

In fact, the eigenvalues can be calculated explicitly: $\omega_j = \sqrt{\frac{2K}{m}}\left(1 - \cos\left(\frac{j\pi}{N+1}\right)\right)^{1/2}$.

An important feature of any linear dynamics is that there is no exchange of energy between modes. Let us restrict to real solutions. When the initial data are composed of $M < N$ complex conjugate pairs

$$\begin{pmatrix} \vec{y}(0) \\ \frac{d\vec{y}}{dt}(0) \end{pmatrix} = \sum_{j\in\{j_1,\dots,j_M\}\subset\{1,\dots,N\}} c_j \begin{pmatrix} \xi^{(j)} \\ i\omega_j \xi^{(j)} \end{pmatrix} + \overline{c_j}\begin{pmatrix} \xi^{(j)} \\ -i\omega_j \xi^{(j)} \end{pmatrix},$$

then clearly

$$\vec{y}(t) = \sum_{j\in\{j_1,\dots,j_M\}} (c_j e^{i\omega_j t} + e^{-i\omega_j t}\overline{c_j})\xi^{(j)},$$

such that only the modes $j_1, \dots, j_M$ come into play.

In a nonlinear case this changes dramatically. When the system is nonlinear, no principle of superposition holds, of course. There is no concept of solution modes either. When initial data are composed of certain linear modes, then one expects that due to the nonlinearity all other linear modes will be excited in the time evolution.

Fermi, Pasta and Ulam considered a quadratic nonlinearity as a correction of the linear Hooke's law:

$$F_R(n) = K(y_{n+1} - y_n) + \alpha K(y_{n+1} - y_n)^2,\ F_L(n) = K(y_n - y_{n-1}) + \alpha K(y_n - y_{n-1})^2 \quad \text{with } \alpha > 0$$

leading to

$$m\frac{d^2 y_n}{dt^2} = K(y_{n+1} - 2y_n + y_{n-1}) + K\alpha\left[(y_{n+1} - y_n)^2 - (y_n - y_{n-1})^2\right],\ n \in \{1, \dots, N\} \tag{2.17}$$

and they used periodic boundary conditions, i.e. $y_0 = y_N, y_{N+1} = y_1$. Arguing based on the ergodicity of Brownian motion, i.e. that every state has the same probability, they even expected that after a long time there will be an equipartition of energy between all the linear modes. What they observed in their numerical experiments (with $N = 32, \alpha = 1/4$) was that practically only 5 linear modes got excited and the energy oscillated among these in a nearly periodic fashion. The amount of energy in the remaining linear modes was negligible.

This surprising result initiated a flurry of research on nonlinear PDEs and is sometimes considered the birth of nonlinear physics. A mathematical explanation of this phenomenon in the case of the FPU-problem is based on an asymptotic approximation of the FPU system (2.17) by the Korteweg-de Vries equation. This continuum approximation is valid for waves that are long compared to the distance of the particles. Since all initial data in the KdV lead to an $m-$soliton ($m \in \mathbb{N}$) plus a radiation wave of amplitude $O(t^{-1/3})(t \to \infty)$, with periodic boundary conditions the form of a soliton repeats periodically on the finite spatial interval. Hence a nearly periodic solution is observed in the simulation.

2.1.2.2 KdV Approximation of long waves in the FPU Problem

Here we formally derive the KdV equation as a continuum approximation of the FPU problem (2.17). Let us first assume that the spacing between articles at rest is $h \ll 1$ and denote the locations at rest by $x_n := nh$. In a continuum approximation one assumes the existence of a smooth function $y(x,t)$ such that $y(x_n,t) = y_n(t)$. Taylor-expanding then y in x around x_n, we have

$$y_{n+1}(t) = y(x_{n+1},t) = y(x_n,t) + h\partial_x y(x_n,t) + \tfrac{h^2}{2}\partial_x^2 y(x_n,t) + \tfrac{h^3}{6}\partial_x^3 y(x_n,t) + \tfrac{h^4}{24}\partial_x^4 y(x_n,t) + \tfrac{h^5}{5!}\partial_x^5 y(x_n,t) + O(h^6),$$
$$y_{n-1}(t) = y(x_{n-1},t) = y(x_n,t) - h\partial_x y(x_n,t) + \tfrac{h^2}{2}\partial_x^2 y(x_n,t) - \tfrac{h^3}{6}\partial_x^3 y(x_n,t) + \tfrac{h^4}{24}\partial_x^4 y(x_n,t) - \tfrac{h^5}{5!}\partial_x^5 y(x_n,t) + O(h^6),$$

such that

$$y_{n+1} - 2y_n + y_{n-1} = h^2\partial_x^2 y(x_n,t) + \tfrac{h^4}{12}\partial_x^4 y(x_n,t) + O(h^6),$$
$$(y_{n+1} - y_n)^2 - (y_n - y_{n-1})^2 = 2h^3\partial_x y(x_n,t)\partial_x^2 y(x_n,t) + O(h^5).$$

Equation (2.17) then implies

$$\frac{m}{Kh^2}\partial_t^2 y - \partial_x^2 y = \tfrac{h^2}{12}\partial_x^4 y + 2\alpha h\partial_x y\partial_x^2 y + O(h^3).$$

The KdV equation is obtained if one assumes that the nonlinearity coefficient α scales like h, i.e.

$$\alpha = O(h), \alpha \neq o(h) \quad (h \to 0).$$

Denoting then $\varepsilon := 2\alpha h$ and $\kappa := \frac{h^2}{2\varepsilon}$, and rescaling time by defining $\tau := \sqrt{\frac{K}{m}}ht$, $\tilde{y}(x,\tau) := y(x, (\frac{K}{m})^{-\frac{1}{2}}h^{-1}\tau)$ and dropping the tilde, we arrive at

$$\partial_\tau^2 y - \partial_x^2 y = \varepsilon(\partial_x y\partial_x^2 y + \kappa\partial_x^4 y) + O(\varepsilon^{3/2}). \tag{2.18}$$

It is now left as an exercise to formally derive the KdV using a perturbation expansion for right traveling waves in (2.18).

2.2 The Nonlinear Schrödinger Equation (NLS)

The KdV holds a very prominent place in the history of the science of nonlinear dispersive waves. Of course, it remains to be physically highly relevant even today. Another fundamental equation describing nonlinear dispersive waves is the nonlinear Schrödinger equation (NLS)

$$\mathrm{i}\partial_t u + \Delta u + \gamma|u|^2 u = 0, \; x \in \mathbb{R}^n. \tag{2.19}$$

As we show in Section 2.2.1, the NLS is extremely universal being an effective approximative model for slowly modulated wavepackets of small amplitude in a large class of nonlinear dispersive equations. As such, the NLS is often used as a prototype model for studying nonlinear wave phenomena.

Let us mention two classical physical applications of the NLS.

1. wavepackets on deep water

   Consider the water/air interface problem (1.61) without surface tension, i.e. $T = 0$, in the deep water limit $h_0 = -\infty$ and restrict to one dimensional waves. With the wavepacket ansatz

$$\varphi \sim \varepsilon A(\varepsilon(x_1 - v_g t), \varepsilon^2 t) e^{\mathrm{i}(kx_1 - \omega(k)t) + |k|x_3} + \text{c.c.},$$
$$\eta \sim \varepsilon B(\varepsilon(x_1 - v_g t), \varepsilon^2 t) e^{\mathrm{i}(kx_1 - \omega(k)t)} + \text{c.c.}$$

   for $\varepsilon \to 0$, where $\omega(k)$ is the dispersion relation for waves propagating in the x_1 direction, one obtains after some calculation the following effective NLS equations for the envelopes A and B

$$\mathrm{i}\partial_T A + \frac{\omega''(k)}{2}\partial_\xi^2 A - \frac{2k^4}{\omega(k)}|A|^2 A = 0$$
$$\mathrm{i}\partial_T B + \frac{\omega''(k)}{2}\partial_\xi^2 B - 2k^2\omega(k)|B|^2 B = 0$$

   with $T := \varepsilon^2 t, \xi := \varepsilon(x_1 - v_g t)$, see Sec. 6.4, 6.5 in [1].

2. wavepacket-pulses in cubically nonlinear optical fibers

   As optical pulses are electromagnetic phenomena, the starting model are typically the Maxwell's equations. In the second order formulation for the electric field we have

$$\nabla \times \nabla \times \vec{E} + \frac{1}{c^2}\partial_t^2 \vec{E} = -\frac{1}{c^2\epsilon_0}\partial_t^2 \vec{P}, \quad c = (\epsilon_0\mu_0)^{-1/2},$$

   where ϵ_0, μ_0 and c are the electric permittivity, magnetic permeability and speed of light in vacuum respectively while $\vec{P}$ is the polarisation which describes the response of the medium to the field. For media with a so called inversion symmetry and with further isotropic symmetries on the atomic level (valid, for instance, for glass used in photonic fibers) the polarisation can be written in the form

$$\vec{P}(x,t) = \epsilon_0 \int_{\mathbb{R}} \chi_1(t-s)\vec{E}(x,s)ds + \epsilon_0 \int_{\mathbb{R}} \chi_3(t-s_1, t-s_2, t-s_3)(\vec{E}(x,s_1)\cdot\vec{E}(x,s_2))\vec{E}(x,s_3)ds_1 ds_2 ds_3,$$

   where the scalars $\chi_{1,3}$ are the linear and the cubic susceptibilities of the medium, see [17]. In more general media (without isotropy) the cubic susceptibility χ_3 is a tensor. For a fiber with x_3 being its longitudinal direction we can assume the cylindrical symmetry such that $\chi_{1,3} = \chi_{1,3}(r,t), r := (x_1^2 + x_2^2)^{1/2}$.

   For the wavepacket ansatz

$$\vec{E} \sim \vec{U}(r)A(\varepsilon(t - v_g x_3), \varepsilon^2 x_3)e^{\mathrm{i}(k(\omega)x_3 - \omega t)} + \text{c.c.}, \quad v_g(k) = k'(\omega),$$

   where $k(\omega)$ is the (inverse of) the dispersion relation for the Maxwell's equation, a lengthy calculation produces the following NLS as an effective model for the envelope A

$$\mathrm{i}\partial_Z A - \frac{1}{2}k''(\omega)\partial_\tau^2 A + \gamma|A|^2 A = 0,$$

   where $Z = \varepsilon^2 x_3, \tau = \varepsilon(t - v_g x_3)$. The full derivation can be found in [17].

2.2.1 Universality of the NLS for Slowly Modulated Wavepackets of Small Amplitude

As we show below, the NLS is not limited to any set of specific physical situations. It describes the asymptotics of wavepackets in a large class of nonlinear dispersive problems by being the effective equation for the corresponding envelopes.

To keep the calculations reasonably simple, let us restrict to a scalar PDE

$$L(\partial_t, \nabla)u + f(u) = 0, \quad x \in \mathbb{R}^n \tag{2.20}$$

with a smooth function $L : \mathbb{R}^{n+1} \to \mathbb{R}$ and with the following behavior of $f(u)$ near $u = 0$

$$f(u) \sim \alpha u^3 \text{ as } u \to 0 \text{ with } \alpha \in \mathbb{R}\backslash\{0\}.$$

Typically, L is a polynomial but the following calculation allows a general smooth L.

As the following analysis shows, the NLS is an effective asymptotic model for wavepacket envelopes when the wavepacket has the following form

$$u(x,t) \approx \varepsilon A(\varepsilon(x - v_g(k_0)t), \varepsilon^2 t)e^{\mathrm{i}(k_0\cdot x - \omega_0 t)} + \text{c.c.} =: \varepsilon u_{\text{app}}(x,t) + \text{c.c.} \tag{2.21}$$

for some $\varepsilon > 0$ small, where $\omega_0, k_0, v_g(k_0) \in \mathbb{R}$ come from the dispersion relation of (2.20), i.e. from

$$L(-\mathrm{i}\omega, \mathrm{i}k) = 0 \tag{2.22}$$

with the solution $\omega = \omega(k)$. We define $\omega_0 := \omega(k_0), v_g(k) := \nabla\omega(k)$. Note that while (2.21) may seem highly special, it is in fact the only ansatz for a broad wavepacket with the carrier wave $e^{\mathrm{i}(k_0\cdot x - \omega_0 t)}$ such that the nonlinear and dispersive effects of (2.20) are asymptotically balanced.

Let us now calculate the group velocity $v_g(k)$ as well as the Hessian matrix $D^2\omega(k)$, which will be needed later. By differentiating (2.22) in k we get

$$-\mathrm{i}(\partial_1 L)(-\mathrm{i}\omega(k), \mathrm{i}k)v_g(k) + \mathrm{i}(\nabla L)(-\mathrm{i}\omega(k), \mathrm{i}k) = 0. \tag{2.23}$$

Differentiating twice leads to

$$\mathrm{i}(\partial_1 L)(-\mathrm{i}\omega(k), \mathrm{i}k)(D^2\omega)(k) = 2(\nabla\partial_1 L)(-\mathrm{i}\omega(k), \mathrm{i}k)v_g^T(k) - (D^2 L)(-\mathrm{i}\omega(k), \mathrm{i}k) - (\partial_1^2 L)(-\mathrm{i}\omega(k), \mathrm{i}k)v_g(k)v_g^T(k), \tag{2.24}$$

where we are using the notation $\partial_1 L = \partial_s L, \nabla L = \nabla_{\vec{\xi}} L, D^2 L = \nabla_{\vec{\xi}}\nabla_{\vec{\xi}} L$ for $L = L(s, \vec{\xi})$ with $\vec{\xi} \in \mathbb{R}^n$.

In order for (2.20) to be a dispersive equation, we assume

$$\omega(k) \in \mathbb{R}, v_g(k) \in \mathbb{R}^n, \text{ for all } k \in \mathbb{R}^n \text{ and } \det(D^2\omega)(k_0) \neq 0.$$

Next we carry out a formal asymptotic analysis for the ansatz (2.21) and seek an effective equation for the envelope A.

Let us first work only on the linear part of (2.20). This can be done more easily in Fourier variables. Hence, we apply the Fourier transform to $L(\partial_t, \nabla)u_{\text{app}}$ and get

$$(L(\partial_t, \nabla)u_{\text{app}})\widehat{\ }(k,t) = L(\partial_t, \mathrm{i}k)\widehat{u}_{\text{app}}(k,t).$$

Next, with the substitution $y := \varepsilon(x - v_g t)$

$$\begin{aligned}\widehat{u}_{\text{app}}(k,t) &= (2\pi)^{-n/2}e^{-\mathrm{i}\omega_0 t}\varepsilon^{-n}\int_{\mathbb{R}^n} A(y, \varepsilon^2 t)e^{\mathrm{i}(k_0 - k)\cdot\left(\frac{y}{\varepsilon} + v_g t\right)}dy \\ &= (2\pi)^{-n/2}e^{-\mathrm{i}(\omega_0 + (k - k_0)\cdot v_g)t}\varepsilon^{-n}\int_{\mathbb{R}^n} A(y, \varepsilon^2 t)e^{-\mathrm{i}\frac{k - k_0}{\varepsilon}\cdot y}dy = \varepsilon^{-n}e^{-\mathrm{i}(\omega_0 + \varepsilon\kappa\cdot v_g)t'}\widehat{A}(\kappa, T),\end{aligned}$$

where $\kappa := \frac{k-k_0}{\varepsilon}, t' := t$, and $T := \varepsilon^2 t$. We see that the Fourier transform of a small wavepacket, broad in x, is a tightly localized function in k of large amplitude and the localization is near the carrier wavevector k_0. The scaling of the amplitude and the width can be best seen in the simple identity $\widehat{f(\varepsilon\cdot)}(k) = \varepsilon^{-n}\widehat{f}(\varepsilon k)$.

Using the new variables t', T, and κ, we have

$$L(\partial_t, \mathrm{i}k) = L(\partial_{t'} + \varepsilon^2\partial_T, \mathrm{i}(k_0 + \varepsilon\kappa)).$$

This operator can be expanded in a Taylor series in ε assuming that the function to which the operator is applied is smooth enough:

$$L(\partial_t, \mathrm{i}k) = L(\partial_{t'}, \mathrm{i}k_0) + \varepsilon^2(\partial_1 L)(\partial_{t'}, \mathrm{i}k_0)\partial_T + \mathrm{i}\varepsilon\kappa\cdot(\nabla L)(\partial_{t'}, \mathrm{i}k_0) - \frac{\varepsilon^2}{2}\kappa^T(D^2L)(\partial_{t'}, \mathrm{i}k_0)\kappa + O(\varepsilon^3). \quad (2.25)$$

Because the derivative $\partial_{t'}$ acts only on the exponential factor $e^{-\mathrm{i}(\omega_0+\varepsilon\kappa\cdot v_g)t'}$ in $\widehat{u}_{\text{app}}$, we get

$$\begin{aligned}L(\partial_{t'}, \mathrm{i}k_0)\widehat{u}_{\text{app}} =& \varepsilon^{-n}\left[L(-\mathrm{i}\omega_0, \mathrm{i}k_0)\widehat{A} - \mathrm{i}\varepsilon\kappa\cdot v_g(\partial_1 L)(-\mathrm{i}\omega_0, \mathrm{i}k_0)\widehat{A} - \frac{\varepsilon^2}{2}(\kappa\cdot v_g)^2(\partial_1^2 L)(-\mathrm{i}\omega_0, \mathrm{i}k_0)\widehat{A} + O(\varepsilon^3)\right]e^{-\mathrm{i}(\omega_0+\varepsilon\kappa\cdot v_g)t'}\\ =& \varepsilon^{-n}\left[\underline{-\mathrm{i}\varepsilon\kappa\cdot v_g(\partial_1 L)(-\mathrm{i}\omega_0, \mathrm{i}k_0)\widehat{A}} - \underline{\frac{\varepsilon^2}{2}\kappa\cdot(v_g v_g^T)\kappa(\partial_1^2 L)(-\mathrm{i}\omega_0, \mathrm{i}k_0)\widehat{A}} + O(\varepsilon^3)\right]e^{-\mathrm{i}(\omega_0+\varepsilon\kappa\cdot v_g)t'}.\end{aligned}$$

Remark 2.2. *Note that if L is a polynomial, one can, of course, use the Leibniz rule. For some $q \in \mathbb{N}$ we have then*

$$L(\partial_{t'}, \mathrm{i}k_0)(fg) = \sum_{j=0}^{q} a_j(\mathrm{i}k_0)\partial_{t'}^j(fg) = \sum_{j=0}^{q} a_j(\mathrm{i}k_0)\sum_{l=0}^{j}\binom{j}{l}\partial_{t'}^{j-l}f\partial_{t'}^l g.$$

Because we apply this formula to $f = e^{-\mathrm{i}\omega_0 t'}, g = e^{-\mathrm{i}\varepsilon\kappa\cdot v_g t'}$ and because $O(\varepsilon^3)$ terms will be neglected, we are interested in at most second derivatives of g. Terms with the first derivative are

$$\sum_{j=0}^{q} a_j(\mathrm{i}k_0)j\partial_{t'}^{j-1}f\partial_{t'}g = \partial_{t'}g(\partial_1 L)(\partial_{t'}, \mathrm{i}k_0)f$$

and terms with the second derivative are

$$\sum_{j=0}^{q} a_j(\mathrm{i}k_0)\tfrac{1}{2}j(j-1)\partial_{t'}^{j-2}f\partial_{t'}^2 g = \tfrac{1}{2}\partial_{t'}^2 g(\partial_1^2 L)(\partial_{t'}, \mathrm{i}k_0)f.$$

We continue with a calculation of the terms in (2.25):

$$\varepsilon^2(\partial_1 L)(\partial_{t'}, \mathrm{i}k_0)\partial_T\widehat{u}_{\text{app}} = \varepsilon^{-n}\left[\varepsilon^2(\partial_1 L)(-\mathrm{i}\omega_0, \mathrm{i}k_0)\partial_T\widehat{A} + O(\varepsilon^3)\right]e^{-\mathrm{i}(\omega_0+\varepsilon\kappa\cdot v_g)t'},$$

$$\begin{aligned}\mathrm{i}\varepsilon\kappa\cdot(\nabla L)(\partial_{t'}, \mathrm{i}k_0)\widehat{u}_{\text{app}} =& \varepsilon^{-n}\left[\mathrm{i}\varepsilon\kappa\cdot(\nabla L)(-\mathrm{i}\omega_0, \mathrm{i}k_0)\widehat{A} + \varepsilon^2(\kappa\cdot v_g)\kappa\cdot(\nabla\partial_1 L)(-\mathrm{i}\omega_0, \mathrm{i}k_0)\widehat{A} + O(\varepsilon^3)\right]e^{-\mathrm{i}(\omega_0+\varepsilon\kappa\cdot v_g)t'}\\ =& \varepsilon^{-n}\left[\underline{\mathrm{i}\varepsilon\kappa\cdot(\nabla L)(-\mathrm{i}\omega_0, \mathrm{i}k_0)\widehat{A}} + \underline{\varepsilon^2\kappa\cdot\left((\nabla\partial_1 L)(-\mathrm{i}\omega_0, \mathrm{i}k_0)v_g^T\kappa\right)}\widehat{A} + O(\varepsilon^3)\right]e^{-\mathrm{i}(\omega_0+\varepsilon\kappa\cdot v_g)t'},\end{aligned}$$

and

$$-\frac{\varepsilon^2}{2}\kappa^T(D^2L)(\partial_{t'}, \mathrm{i}k_0)\kappa\widehat{u}_{\text{app}} = -\varepsilon^{-n}\left[\underline{\frac{\varepsilon^2}{2}\kappa^T(D^2L)(-\mathrm{i}\omega_0, \mathrm{i}k_0)\kappa\widehat{A}} + O(\varepsilon^3)\right]e^{-\mathrm{i}(\omega_0+\varepsilon\kappa\cdot v_g)t'}.$$

The sum of the terms underlined by the dashed line vanishes due to (2.23) and the ones underlined by the full line add up to

$$\mathrm{i}\frac{\varepsilon^2}{2}(\partial_1 L)(-\mathrm{i}\omega_0, \mathrm{i}k_0)(\kappa^T(D^2\omega)(k_0)\kappa)\widehat{A}$$

due to (2.24). As a result

$$L(\partial_t, \mathrm{i}k)\widehat{u}_{\text{app}} = \varepsilon^{-n}\left[\varepsilon^2(\partial_1 L)(-\mathrm{i}\omega_0, \mathrm{i}k_0)\left(\partial_T\widehat{A} + \mathrm{i}\kappa^T\frac{(D^2\omega)(k_0)}{2}\kappa\widehat{A}\right) + O(\varepsilon^3)\right]e^{-\mathrm{i}(\omega_0+\varepsilon\kappa\cdot v_g)t'}.$$

With the inverse Fourier transform we get

$$L(\partial_t, \nabla)\varepsilon u_{\text{app}} = \varepsilon^3(\partial_1 L)(-\mathrm{i}\omega_0, \mathrm{i}k_0)\left[\partial_T A - \mathrm{i}\nabla\cdot\left(\frac{(D^2\omega)(k_0)}{2}\nabla A\right)\right]e^{\mathrm{i}(k_0\cdot x-\omega_0 t)} + O(\varepsilon^4).$$

The nonlinear term $f(u)$ in (2.20) is simpler. We get

$$f(u) = \varepsilon^3\alpha(u_{\text{app}} + \overline{u_{\text{app}}})^3 + O(\varepsilon^4) = \varepsilon^3\alpha\left[3|A|^2Ae^{\mathrm{i}(k_0\cdot x-\omega_0 t)} + A^3e^{3\mathrm{i}(k_0\cdot x-\omega_0 t)} + \text{c.c.}\right] + O(\varepsilon^4).$$

The aim is to make the residual $L(\partial_t, \nabla)(u_{\text{app}} + \overline{u_{\text{app}}}) + f(u_{\text{app}} + \overline{u_{\text{app}}})$ of $O(\varepsilon^4)$. The $O(\varepsilon^3)$ terms proportional to $e^{\pm\mathrm{i}(k_0\cdot x-\omega_0 t)}$ vanish if we require

$$\boxed{\mathrm{i}\partial_T A + \frac{1}{2}\nabla\cdot\big((D^2\omega)(k_0)\nabla A\big) + \gamma|A|^2A = 0,} \tag{2.26}$$

where $\gamma = \frac{3\mathrm{i}\alpha}{(\partial_1 L)(-\mathrm{i}\omega_0, \mathrm{i}k_0)}$. Typically, it is $\gamma \in \mathbb{R}$.

To eliminate also $O(\varepsilon^3)$ terms proportional to $e^{\pm 3\mathrm{i}(k_0\cdot x-\omega_0 t)}$, we modify the ansatz u_{app} by adding an $O(\varepsilon^3)$ correction. Namely, assuming the non-resonance condition

$$L(-3\mathrm{i}\omega_0, 3\mathrm{i}k_0) \neq 0, \tag{2.27}$$

we set

$$\varepsilon u_{\text{app}}(x,t) := \varepsilon A(\varepsilon(x - v_g(k_0)t), \varepsilon^2 t)e^{\mathrm{i}(k_0\cdot x-\omega_0 t)} - \varepsilon^3\frac{\alpha}{L(-3\mathrm{i}\omega_0, 3\mathrm{i}k_0)}A^3(\varepsilon(x - v_g(k_0)t), \varepsilon^2 t)e^{3\mathrm{i}(k_0\cdot x-\omega_0 t)}. \tag{2.28}$$

Equation (2.26) is a slight generalization of the *nonlinear Schrödinger equation* (2.19). In the isotropic case when $D^2\omega(k_0) = \alpha I, \alpha \in \mathbb{R}$ we get $\nabla\cdot\big((D^2\omega)(k_0)\nabla\big) = \alpha\Delta$ and recover (2.19).

2.2.2 Justification of the NLS for the Nonlinear Wave Equation

In contrast to Section 2.2.1, where the NLS was only formally derived and no proof was given, we want to be rigorous in this section. The question we want to address is whether the above $\varepsilon u_{\text{app}}$, with A a solution of the NLS, provides an approximation of a solution u of (2.20) - and in which norm. Note that we need the approximation to be valid on a time interval of length $O(\varepsilon^{-2})$ because changes in the envelope A occur on the time scale $O(\varepsilon^2)$.

To keep the analysis and the notation reasonably simple, we restrict to one specific example of (2.20), namely the following nonlinear wave equation

$$\partial_t^2 u - \Delta u + u - u^3 = 0, \quad x \in \mathbb{R}^n. \tag{2.29}$$

2.2.2.1 Justification in One Spatial Dimension in $H^1(\mathbb{R})$

In one spatial dimension the problem becomes

$$\partial_t^2 u - \partial_x^2 u + u - u^3 = 0, \quad x \in \mathbb{R}. \tag{2.30}$$

Hence, with the notation of (2.20) we have $L(\partial_t, \partial_x)u = \partial_t^2 u - \partial_x^2 u + u, f(u) = -u^3$. The dispersion relation is

$$\omega^2 = k^2 + 1$$

such that

$$\omega'(k) = \frac{k}{\omega(k)}, \ \omega''(k) = \frac{\omega(k) - k\omega'(k)}{\omega^2(k)} = \frac{1 - (\omega'(k))^2}{\omega(k)}.$$

Choosing a wavenumber $k_0 \in \mathbb{R}$, we define the corresponding frequency and group velocity

$$\omega_0 := \omega(k_0), \ v_g := \omega'(k_0) = \frac{k_0}{\omega_0}.$$

The wavepacket ansatz corresponding to (2.28) is

$$u_{\text{as}}(x,t) = \varepsilon A(X,T) e^{\mathrm{i}(k_0 x - \omega_0 t)} + \varepsilon^3 \frac{A^3(X,T)}{9k_0^2 - 9\omega_0^2 + 1} e^{3\mathrm{i}(k_0 x - \omega_0 t)} + \text{c.c.}, \tag{2.31}$$

where $X = \varepsilon(x - v_g t), T = \varepsilon^2 t$ and the effective NLS as given by (2.26) is

$$\mathrm{i}\partial_T A + \frac{1}{2}\frac{1 - v_g^2}{\omega_0}\partial_X^2 A + \frac{3}{2\omega_0}|A|^2 A = 0, \tag{2.32}$$

where we have used the fact that $(\partial_1 L)(-\mathrm{i}\omega_0, \mathrm{i}k_0) = -2\mathrm{i}\omega_0$. The non-resonance condition (2.27) now reads $9\omega_0^2 \neq 9k_0^2 + 1$, which holds because $9\omega^2(k) = 9k^2 + 9$.

Our aim is to show that if $u(x,0) = u_{\text{as}}(x,0), \partial_t u(x,0) = \partial_t u_{\text{as}}(x,0)$ and A solves (2.32) on an interval $[0, T_0]$, then $\|u(\cdot,t) - u_{\text{as}}(\cdot,t)\|_{H^1(\mathbb{R})} \leqslant c\varepsilon^\alpha$ with some $\alpha > 0$ for all $t \in [0, T_0\varepsilon^{-2}]$ and $\varepsilon > 0$ small enough. It turns out that the optimal value of α is $3/2$. Therefore, we define the error as

$$\varepsilon^{3/2} R(x,t) := u(x,t) - u_{\text{as}}(x,t)$$

and aim to show $\|R(\cdot,t)\|_{H^1(\mathbb{R})} \leqslant c$ for all $t \in [0, T_0\varepsilon^{-2}]$. The residual is

$$\mathrm{R}es(x,t) := \partial_t^2 u_{\text{as}}(x,t) - \partial_x^2 u_{\text{as}}(x,t) + u_{\text{as}}(x,t) - u_{\text{as}}^3(x,t).$$

The equation for R is easily derived: we arrive at the Cauchy problem

$$\begin{aligned} &\partial_t^2 R = \partial_x^2 R - R + f(R), \ x \in \mathbb{R}, t > 0 \quad \text{where} \ \ f(R) := 3u_{\text{as}}^2 R + 3\varepsilon^{3/2} u_{\text{as}} R^2 + \varepsilon^3 R^3 - \varepsilon^{-3/2}\mathrm{R}es \\ &R(x,0) = \partial_t R(x,0) = 0, \end{aligned} \tag{2.33}$$

where the trivial initial data follow from the assumption $u(x,0) = u_{\text{as}}(x,0), \partial_t u(x,0) = \partial_t u_{\text{as}}(x,0)$.

We proceed in two steps. First we use a fixed point argument to show the existence and uniqueness of the solution to (2.33) on a time interval of length $O(1)$ ($\varepsilon \to 0$) such that on this time interval it is bounded by a (ε-independent) constant in the H^1 norm. Afterwards a Gronwall argument shows that, in fact, the interval can be extended to $[0, T_0\varepsilon^{-2}]$ by redefining the constant and by making ε small enough.

For the Banach fixed point argument let us first consider the following inhomogeneous linear equation (posed in $\mathbb{R}^n, n \in \mathbb{N}$ because there is no additional difficulty compared to $\mathbb{R}$).

$$\begin{aligned} &\partial_t^2 u - \Delta u + u = F(x,t), \quad x \in \mathbb{R}^n, t > 0 \\ &u(x,0) = g(x), \partial_t u(x,0) = h(x). \end{aligned} \tag{2.34}$$

In Fourier variables problem (2.34) becomes

$$\partial_t^2 \widehat{u} + (|k|^2+1)\widehat{u} = \widehat{F}(k,t), \quad k \in \mathbb{R}^n, t > 0$$
$$\widehat{u}(k,0) = \widehat{g}(k), \partial_t \widehat{u}(k,0) = \widehat{h}(k),$$

which has the solution

$$\widehat{u}(k,t) = \widehat{g}(k)\cos(\sqrt{|k|^2+1}t) + \frac{\sin(\sqrt{|k|^2+1}t)}{\sqrt{|k|^2+1}}\widehat{h}(k) + \int_0^t \widehat{F}(k,\tau)\frac{\sin(\sqrt{|k|^2+1}(t-\tau))}{\sqrt{|k|^2+1}}d\tau. \tag{2.35}$$

This representation makes it easy to prove the following energy estimate.

Lemma 2.2. *If $F \in L^2(0,T_*,H^{s-1}(\mathbb{R}^n)) \cap C([0,T],L^2(\mathbb{R}^n))$, $s \geqslant 2$, then (2.35) gives the unique $C([0,T_*],H^s(\mathbb{R}^n)) \cap C^1([0,T_*],H^{s-1}(\mathbb{R}^n))$ solution of (2.34), i.e. $\partial_t^2 u - \Delta u + u = F(\cdot,t)$ in $L^2(\mathbb{R}^n)$ for all $t \in (0,T_*]$). Moreover*

$$\|u(\cdot,t)\|_{H^s(\mathbb{R}^n)} + \|\partial_t u(\cdot,t)\|_{H^{s-1}(\mathbb{R}^n)} \leqslant c\left[\|g\|_{H^s(\mathbb{R}^n)} + \|h\|_{H^{s-1}(\mathbb{R}^n)} + \sqrt{t}\left(\int_0^t \|F(\cdot,\tau)\|_{H^{s-1}(\mathbb{R}^n)}d\tau\right)^{1/2}\right]$$

for all $t \in [0,T_*]$.

Proof. The solution property is equivalent to $\partial_t^2 \hat{u} + (|k|^2+1)\hat{u} = \hat{F}(\cdot,t)$ in $L^2(\mathbb{R})$ for all $t \in (0,T_*)$. This follows by differentiating (2.35). The derivative with respect to t inside the integral is treated using the Lebesgue dominated convergence theorem. For instance, $\tau \mapsto \widehat{F}(k,\tau)\frac{\sin(\sqrt{|k|^2+1}(t-\tau))}{\sqrt{|k|^2+1}}$ is measurable for all $t \in [0,T_*]$ and almost all $k \in \mathbb{R}^n$, $t \mapsto \widehat{F}(k,\tau)\frac{\sin(\sqrt{|k|^2+1}(t-\tau))}{\sqrt{|k|^2+1}}$ and $t \mapsto \widehat{F}(k,\tau)\cos(\sqrt{|k|^2+1}(t-\tau))$ is continuous on $[0,T_*]$ for almost all $\tau \in (0,T_*)$ and $k \in \mathbb{R}^n$ and $\widehat{F}(k,\cdot) \in L^1(0,T_*)$ for almost all k by the Cauchy-Schwarz inequality since $\widehat{F}(k,\cdot) \in L^2(0,T_*)$ and $(0,T_*)$ is bounded.

Next, we show the estimate. By the definition of the H^s norm

$$\begin{aligned}\|u(\cdot,t)\|^2_{H^s(\mathbb{R}^n)} &= \int_{\mathbb{R}^n} |\widehat{u}(k,t)|^2(1+|k|^s)^2 dk \\ &\leqslant 3\left[\int_{\mathbb{R}^n} |\widehat{g}(k)|^2(1+|k|^s)^2 dk + \int_{\mathbb{R}^n} |\widehat{h}(k)|^2 \frac{(1+|k|^s)^2}{1+|k|^2}dk + \int_{\mathbb{R}^n}\left(\int_0^t |\widehat{F}(k,\tau)|d\tau\right)^2 \frac{(1+|k|^s)^2}{1+|k|^2}dk\right],\end{aligned}$$

where the estimate follows from $2ab \leqslant a^2+b^2$ for any $a,b \in \mathbb{R}$. Because there are constants $c_1, c_2 > 0$ such that $c_1(1+a^2)^s \leqslant (1+a^s)^2 \leqslant c_2(1+a^2)^s$ for all $a \in \mathbb{R}$, we have

$$\|u(\cdot,t)\|^2_{H^s(\mathbb{R}^n)} \leqslant c\left[\|g\|^2_{H^s(\mathbb{R}^n)} + \|h\|^2_{H^{s-1}(\mathbb{R}^n)} + t\int_0^t \|F(\cdot,\tau)\|^2_{H^{s-1}(\mathbb{R}^n)}d\tau\right],$$

where we have also used the Cauchy-Schwarz inequality $(\int_0^t |F(k,\tau)|d\tau)^2 \leqslant t\int_0^t |F(k,\tau)|^2 d\tau$. Similarly, because

$$\partial_t \widehat{u}(k,t) = -\sqrt{|k|^2+1}\widehat{g}(k)\sin(\sqrt{|k|^2+1}t) + \widehat{h}(k)\cos(\sqrt{|k|^2+1}t) + \int_0^t \widehat{F}(k,\tau)\cos(\sqrt{|k|^2+1}(t-\tau))d\tau,$$

we get also

$$\|\partial_t u(\cdot,t)\|^2_{H^{s-1}(\mathbb{R}^n)} \leqslant c\left[\|g\|^2_{H^s(\mathbb{R}^n)} + \|h\|^2_{H^{s-1}(\mathbb{R}^n)} + t\int_0^t \|F(\cdot,\tau)\|^2_{H^{s-1}(\mathbb{R}^n)}d\tau\right].$$

Hence the result follows. □

For our fixed point argument for (2.33) we choose the solution space

$$X(t_*) := C([0,t_*],H^2(\mathbb{R}))\cap C^1([0,t_*],H^1(\mathbb{R})) \text{ with the norm } \|u\|_X := \sup_{t\in[0,t_*]}\left(\|u(\cdot,t)\|_{H^2(\mathbb{R})} + \|\partial_t u(\cdot,t)\|_{H^1(\mathbb{R})}\right).$$

For a given $\rho \in X(t_*)$ we denote by $\Phi(\rho)$ the solution of (2.34) with $F := f(\rho)$ and the initial data $u(x,0) = \partial_t u(x,0) = 0$. Fixed points of Φ are then solutions R of (2.33). By Banach fixed point theorem a fixed point will exist if there are constants $r, t_* > 0$ and $L \in (0,1)$ such that

(i) $\Phi(B_r) \subset B_r$, where $B_r := \{u \in X(t_*) : \|u\|_X \leqslant r\}$,

(ii) $\|\Phi(\rho_1) - \Phi(\rho_2)\|_X \leqslant L\|\rho_1 - \rho_2\|_X$ for all $\rho_1, \rho_2 \in X(t_*)$.

Moreover, if (i) and (ii) hold, the fixed point in B_r is unique. The basic assumption we will need in order to prove (i) and (ii) is the existence of a $T_0 > 0$ such that

$$A \in C^2([0,T_0],H^1(\mathbb{R},\mathbb{C})) \cap C^1([0,T_0],H^3(\mathbb{R},\mathbb{C})).$$

Lemma 2.2 provides the estimate

$$\|\Phi(\rho)\|_X \leqslant ct_* \sup_{t\in[0,t_*]} \|f(\rho)(\cdot,t)\|_{H^1}. \tag{2.36}$$

Hence, we need to control $\|f(\rho)(\cdot,t)\|_{H^1(\mathbb{R})}$. Let us start with the residual, which is explicitly

$$\begin{aligned} \mathrm{R}es = &- \varepsilon^4 \frac{6\mathrm{i}k_0}{9k_0^2 - 9\omega_0^2 + 1}\partial_X(A^3)e^{3\mathrm{i}\theta} + \varepsilon^5\left(\partial_T^2 A e^{\mathrm{i}\theta} - \frac{1}{9k_0^2 - 9\omega_0^2+1}\left(\partial_X^2(A^3) + 6\mathrm{i}\omega_0\partial_T(A^3)\right)\right) \\ &+ \varepsilon^7 \frac{1}{9k_0^2 - 9\omega_0^2 + 1}\partial_T^2(A^3)e^{3\mathrm{i}\theta} + \sum_{j\in\{5,7,9\}} \varepsilon^j N_j(A,\overline{A}) + \text{c.c.}, \end{aligned} \tag{2.37}$$

where $\theta := k_0x - \omega_0 t$ and N_j is a polynomial in A and $\overline{A}$ of total degree j. The polynomials $\varepsilon^j N_j$ come from the nonlinear term $(\varepsilon A e^{\mathrm{i}\theta} + \varepsilon^3 A^3 e^{3\mathrm{i}\theta} + \text{c.c.})$. Therefore all the functions $G_1 := \partial_X(A^3), G_2 := \partial_T^2 A, G_3 := \partial_X^2(A^3), G_4 := \partial_T(A^3), G_5 := \partial_T^2(A^3)$ and $G_6 := N_5, G_7 := N_7, G_8 := N_9$ consist of functions of A (and its derivatives) which depend only on the variables $\varepsilon(x - v_g t)$ and $\varepsilon^2 t$. These terms are possibly multiplied by $e^{mi\theta}$ with $m \in \mathbb{N}$. Hence, we have

$$\begin{aligned} \|\mathrm{R}es(\cdot,t)\|_{H^1(\mathbb{R})} &\leqslant c\varepsilon^4 \sum_m \|G_m(\varepsilon\cdot,\varepsilon^2 t)\|_{H^1} \\ &= c\varepsilon^{7/2}\sum_m \|G_m(\cdot,\varepsilon^2 t)\|_{H^1} \leqslant C\varepsilon^{7/2} \end{aligned}$$

because $A \in C^2([0,T_0],H^1(\mathbb{R},\mathbb{C})) \cap C^1([0,T_0],H^3(\mathbb{R},\mathbb{C}))$. For instance, for the term with $\partial_X^2(A^3)$ we have

$$\begin{aligned} \|\partial_X^2(A^3)(\cdot,T)\|_{H^1} &\leqslant c(\|A(\cdot,t)\|_{C_b^1}^2\|\partial_X^2 A(\cdot,t)\|_{H^1} + \|A(\cdot,t)\|_{C_b^1}\|(\partial_X A)^2(\cdot,t)\|_{H^1}) \\ &\leqslant c(\|A(\cdot,t)\|_{C_b^1}^2\|A(\cdot,t)\|_{H^3} + \|A(\cdot,t)\|_{C_b^2}^2\|A(\cdot,t)\|_{H^2}) \\ &\leqslant c\|A(\cdot,t)\|_{H^3}^3 \end{aligned}$$

because of the embedding $H^s(\mathbb{R}^n) \hookrightarrow C_b^k(\mathbb{R}^n)$ for $s > n/2 + k$.

The above loss of an $\varepsilon^{1/2}$ in the estimate of G_m is caused by the L^2-part of the norm and follows from $\left(\int_{\mathbb{R}} |G_m(\varepsilon x,\varepsilon^2 t)|^2 dx\right)^{1/2} = \varepsilon^{-1/2}\left(\int_{\mathbb{R}} |G_m(y,\varepsilon^2 t)|^2 dy\right)^{1/2}$. The remaining terms in $f(\rho)$ can be estimated as follows.

$$\|u_{\mathrm{as}}^2\rho\|_{H^1} \leqslant \|u_{\mathrm{as}}\|_{C_b^1}^2\|\rho\|_{H^1} \leqslant c\varepsilon^2\|\rho\|_{H^1} \quad \text{for all } t \in [0, T_0\varepsilon^{-2}],$$

$$\|u_{\rm as}\rho^2\|_{H^1} \leqslant \|u_{\rm as}\|_{C_b^1}\|\rho\|_{C_b^1}\|\rho\|_{H^1} \leqslant \|u_{\rm as}\|_{C_b^1}\|\rho\|_{H^2}\|\rho\|_{H^1} \leqslant c\varepsilon\|\rho\|_{H^2}\|\rho\|_{H^1} \quad \text{for all } t \in [0, T_0\varepsilon^{-2}]$$

using again the embedding $H^s(\mathbb{R}) \hookrightarrow C_b^k(\mathbb{R})$ for $s > 1/2 + k$. Similarly,

$$\|\rho^3\|_{H^1} \leqslant c\|\rho\|_{C_b^1}^2\|\rho\|_{H^1} \leqslant c\|\rho\|_{H^2}^2\|\rho\|_{H^1} \quad \text{for all } t \in [0, T_0\varepsilon^{-2}].$$

In summary, there is a $c > 0$ such that

$$\|f(\rho)(\cdot, t)\|_{L^2(\mathbb{R})} \leqslant c\left(\varepsilon^2\|\rho\|_{H^1} + \varepsilon^{5/2}\|\rho\|_{H^2}\|\rho\|_{H^1} + \varepsilon^3\|\rho\|_{H^2}^2\|\rho\|_{H^1} + \varepsilon^2\right) \quad \text{for all } t \in [0, T_0\varepsilon^{-2}].$$

Note that alternatively one can use the algebra property of $H^1(\mathbb{R})$, more generally $H^s(\mathbb{R}^n)$ is an algebra for $s > n/2$, i.e.

$$\|\varphi\psi\|_{H^s(\mathbb{R}^n)} \leqslant c\|\varphi\|_{H^s(\mathbb{R}^n)}\|\psi\|_{H^s(\mathbb{R}^n)} \quad \text{if } s > n/2,$$

see Theorem 4.39 in [3]. This leads to

$$\|f(\rho)(\cdot, t)\|_{H^1(\mathbb{R})} \leqslant c\left(\varepsilon^2\|\rho\|_{H^1} + \varepsilon^{5/2}\|\rho\|_{H^1}^2 + \varepsilon^3\|\rho\|_{H^1}^3 + \varepsilon^2\right) \quad \text{for all } t \in [0, T_0\varepsilon^{-2}].$$

Hence

$$\begin{aligned}
\|\Phi(\rho)\|_X &\leqslant ct_* \sup_{t\in[0,t,*]} \|f(\rho)(\cdot, t)\|_{L^2} \leqslant ct_*\left(\varepsilon^2\|\rho\|_X + \varepsilon^{5/2}\|\rho\|_X^2 + \varepsilon^3\|\rho\|_X^3 + \varepsilon^2\right) \\
&\leqslant ct_*(\varepsilon^2 r + \varepsilon^{5/2}r^2 + \varepsilon^3 r^3 + \varepsilon^2) \quad \text{if } \rho \in B_r \\
&\leqslant r \quad \text{for } \varepsilon > 0 \text{ small enough and } r, t_* = O(1)(\varepsilon \to 0).
\end{aligned}$$

In other words (i) holds for fixed $t_*, r > 0$ if $\varepsilon > 0$ is chosen small enough.

For the contraction property (ii) we estimate

$$\begin{aligned}
\|\Phi(\rho_1) - \Phi(\rho_2)\|_X &\leqslant c\sqrt{t_*}\int_0^{t_*} \|f(\rho_1)(\cdot, \tau) - f(\rho_2)(\cdot, \tau)\|_{H^1} d\tau \\
&\leqslant c\sqrt{t_*}\int_0^{t_*} \varepsilon^2\|(\rho_1 - \rho_2)(\cdot, \tau)\|_{H^1} + \varepsilon^{5/2}\|(\rho_1^2 - \rho_2^2)(\cdot, \tau)\|_{H^1} + \varepsilon^3\|(\rho_1^3 - \rho_2^3)(\cdot, \tau)\|_{H^1} d\tau \\
&\leqslant ct_*\left(\varepsilon^2 + \varepsilon^{5/2}\|\rho_1 + \rho_2\|_X + 2\varepsilon^3(\|\rho_1\|_X^2 + \|\rho_2\|_X^2)\right)\|\rho_1 - \rho_2\|_X,
\end{aligned}$$

where the algebra property of $H^1(\mathbb{R})$ and $H^2(\mathbb{R})$ and the identity $\rho_1^m - \rho_2^m = (\rho_1 - \rho_2)\sum_{k=0}^{m-1}\rho_1^k\rho_2^{m-1-k}$ for $m = 3$ have been used. For $\rho_1, \rho_2 \in B_r$ we have

$$\|\Phi(\rho_1) - \Phi(\rho_2)\|_X \leqslant ct_*\left(\varepsilon^2 + 2\varepsilon^{5/2}r + 4\varepsilon^3 r^2\right)\|\rho_1 - \rho_2\|_X.$$

Clearly, for $\varepsilon > 0$ small enough and $t_*, r = O(1)$ is $ct_*\left(\varepsilon^2 + 2\varepsilon^{5/2}r + 4\varepsilon^3 r^2\right) < 1$.

We conclude that for each $t_*, r > 0$ there exists $\varepsilon_0 > 0$ such that (2.33) with $\varepsilon \in (0, \varepsilon_0)$ has a unique solution $R \in B_r \subset X(t_*)$, i.e.

$$\sup_{t\in[0,t_*]}\left(\|R(\cdot, t)\|_{H^2(\mathbb{R})} + \|\partial_t R(\cdot, t)\|_{H^1(\mathbb{R})}\right) \leqslant r.$$

Next, the aim is to show that the estimate can be extended onto $[0, T_0\varepsilon^{-2}]$. We will, however, restrict only to the H^1 norm of $R(t)$ and the L^2 norm of $\partial_t R(t)$. We show that there exist $K > 0$ and $\varepsilon_0 > 0$ such that

$$E(t) := \|R(\cdot, t)\|_{H^1(\mathbb{R})}^2 + \|\partial_t R(\cdot, t)\|_{L^2(\mathbb{R})}^2 \leqslant K \qquad \text{for all } t \in [0, T_0\varepsilon^{-2}] \text{ and } \varepsilon \in (0, \varepsilon_0).$$

The idea is to formulate an integral inequality for $E(t)$ and apply a Gronwall's inequality argument.

As $E(t) = \int_{\mathbb{R}} (\partial_t R)^2(x,t) + (\partial_x R)^2(x,t) + R^2(x,t)dx$, we get

$$\frac{1}{2}\frac{d}{dt}E(t) = \int_{\mathbb{R}} \partial_t R \partial_t^2 R + \partial_x R \partial_t \partial_x R + R\partial_t R dx = \int_{\mathbb{R}} \partial_t R(\partial_x^2 R - R + f(R)) + \partial_x R \partial_t \partial_x R + R\partial_t R dx$$
$$= \int_{\mathbb{R}} -\partial_t \partial_x R \partial_x R + f(R)\partial_t R + \partial_x R \partial_t \partial_x R dx = \int_{\mathbb{R}} f(R)\partial_t R dx,$$

where the differential equation in (2.33) and partial integration have been used. This leads to

$$\frac{d}{dt}E(t) \leqslant 2\|f(R)\|_{L^2(\mathbb{R})}\|\partial_t R\|_{L^2(\mathbb{R})} \leqslant 2c\|\partial_t R\|_{L^2}\left(\varepsilon^2\|R\|_{L^2} + \varepsilon^{5/2}\|R\|_{H^1}^2 + \varepsilon^3\|R\|_{H^1}^3 + \varepsilon^2\right)$$
$$\leqslant 2c(\varepsilon^2 + \varepsilon^{5/2}E^{1/2} + \varepsilon^3 E)E + 2c\varepsilon^2 E^{1/2}$$

because $\|R\|_{H^1}^2, \|R\|_{L^2}^2, \|\partial_t R\|_{L^2}^2 \leqslant E$. Estimating now $E^{1/2} \leqslant 1 + E$ in the last term, we arrive at

$$\frac{d}{dt}E(t) \leqslant 2c(2\varepsilon^2 + \varepsilon^{5/2}E^{1/2} + \varepsilon^3 E)E + 2c\varepsilon^2$$

and after integration (due to $E(0) = 0$)

$$E(t) \leqslant 2ct\varepsilon^2 + 2c\int_0^t (2\varepsilon^2 + \varepsilon^{5/2}E^{1/2}(\tau) + \varepsilon^3 E(\tau))E(\tau)d\tau.$$

Based on the above fixed point argument we know that if we select $t_*, K > 0$, then there is $\varepsilon_0 > 0$ such that $E(t) \leqslant K$ for all $\varepsilon \in (0, \varepsilon_0)$ and $t \in [0, t_*]$. Then also

$$E(t) \leqslant 2ct\varepsilon^2 + 2c\int_0^t \left(2\varepsilon^2 + \varepsilon^2(\varepsilon^{1/2}K^{1/2} + \varepsilon K)\right) E(\tau)d\tau.$$

If we now choose ε_0 so small that $\varepsilon_0^{1/2}K^{1/2} + \varepsilon_0 K \leqslant 1$, then Gronwall's inequality produces

$$E(t) \leqslant 2ct\varepsilon^2 e^{6c\varepsilon^2 t}.$$

It remains to make sure that $E(t) \leqslant K$ for all $t \in [0, \varepsilon^{-2}T_0]$. This can be easily done by the choice of K. The sequence of steps is

1. define $K := 2cT_0 e^{6cT_0}$,
2. choose $\varepsilon_0 > 0$ so small that $\varepsilon_0^{1/2}K^{1/2} + \varepsilon_0 K \leqslant 1$.

Then for all $\varepsilon \in (0, \varepsilon_0)$

$$E(t) \leqslant K \quad \text{for all } t \in [0, T_0\varepsilon^{-2}],$$

which translates into an estimate for R because $\|R(\cdot, t\|_{H^1} \leqslant \sqrt{E(t)}$.

We have thus proved

Theorem 2.3. *Let $A \in C^2([0,T_0], H^1(\mathbb{R},\mathbb{C})) \cap C^1([0,T_0], H^3(\mathbb{R},\mathbb{C}))$ with some $T_0 > 0$ be a solution of the NLS* (2.32) *and u be a solution of* (2.30) *with $u(x,0) = u_{as}(x,0)$, $\partial_t u(x,0) = \partial_t u_{as}(x,0)$ with u_{as} given by* (2.31)*. There exist $K, \varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0)$*

$$\|u(\cdot,t) - u_{as}(\cdot,t)\|_{H^1(\mathbb{R})} \leqslant K\varepsilon^{3/2} \quad \textit{for all } t \in [0, T_0\varepsilon^{-2}].$$

The existence of smooth NLS solutions A can be found, for instance in Prop. 3.8 of [23].

In fact, as we can see by the triangle inequality even the first term in the asymptotic ansatz (2.31) produces an approximation with an error of the same order.

Corollary 2.4. *With the same assumption as in Theorem 2.3 it is also*

$$\|u(\cdot,t) - u_{as}^{(0)}(\cdot,t)\|_{H^1(\mathbb{R})} \leqslant (K+1)\varepsilon^{3/2} \quad \text{for all } t \in [0, T_0\varepsilon^{-2}],$$

where $u_{as}^{(0)}(x,t) := \varepsilon A(\varepsilon(x - v_g t), \varepsilon^2 t)e^{\mathrm{i}(k_0 x - \omega_0 t)} + c.c.$.

Proof. Clearly, by the triangle inequality it suffices to estimate $\|u_{\text{as}}(\cdot,t) - u_{\text{as}}^{(0)}(\cdot,t)\|_{H^1(\mathbb{R})}$.

$$\begin{aligned}\|u_{\text{as}}(\cdot,t) - u_{\text{as}}^{(0)}(\cdot,t)\|_{H^1} &= \frac{\varepsilon^3}{9k_0^2 - 9\omega_0^2 + 1}\|A^3(\varepsilon(\cdot - v_g t), \varepsilon^2 t)e^{3\mathrm{i}(k_0\cdot - \omega_0 t)} + \text{c.c.}\|_{H^1} \\ &\leqslant c\varepsilon^3\|A^3(\varepsilon\cdot, \varepsilon^2 t)\|_{H^1} \leqslant c\varepsilon^{5/2}\|A^3(\cdot, \varepsilon^2 t)\|_{H^1} \leqslant C\varepsilon^{5/2}\|A(\cdot, \varepsilon^2 t)\|_{H^1}^3,\end{aligned}$$

where the last step follows by the algebra property and the one but last by the scaling property of the L^2 norm: $\|B(\varepsilon\cdot)\|_{L^2} = \left(\varepsilon^{-1}\int_{\mathbb{R}}|B(y)|^2 dy\right)^{1/2} = \varepsilon^{-1/2}\|B\|_{L^2}$. ☐

As we have shown, the approximation error is $O(\varepsilon^{3/2})$ in the H^1-norm. This can be considered as satisfactory because the approximation u_{as} (and $u_{\text{as}}^{(0)}$) is one order larger, namely $O(\varepsilon^{1/2})$ in the H^1-norm. This is again due to the scaling property of the L^2 norm: $\|\varepsilon A(\varepsilon\cdot, \varepsilon^2 t)\|_{L^2(\mathbb{R})} = \varepsilon^{1/2}\|A(\cdot, \varepsilon^2 t)\|_{L^2(\mathbb{R})}$.

Note that Theorem 2.3 assumes $u(x,0) = u_{\text{as}}(x,0)$, $\partial_t u(x,0) = \partial_t u_{\text{as}}(x,0)$. This can be easily generalized to $u(x,0) = u_{\text{as}}(x,0) + \varphi(x)$, $\partial_t u(x,0) = \partial_t u_{\text{as}}(x,0) + \psi(x)$, where $\|\psi\|_{H^2(\mathbb{R})}, \|\psi\|_{H^1(\mathbb{R})} \leqslant c\varepsilon^{3/2}$. This generates no difficulties as the estimate in (2.36) is simply extended by $O(\varepsilon^{3/2})$ terms independent of ρ.

2.2.2.2 Justification in Higher Spatial Dimension in $H^s(\mathbb{R}^n)$

Let us now consider the $n-$dimensional problem (2.29). As we will see, the difficulty of a justification in $H^s(\mathbb{R}^n)$ is the scaling property of the L^2-norm

$$\|f(\varepsilon\cdot)\|_{L^2(\mathbb{R}^n)} = \varepsilon^{-\frac{n}{2}}\|f\|_{L^2(\mathbb{R}^n)},$$

due to which $\frac{n}{2}$ powers of ε are lost in the estimate of the residual (and hence also the error).

In $\mathbb{R}^n$ the dispersion relation $\omega^2 = |k|^2 + 1$ leads to

$$\nabla\omega(k) = \frac{1}{\omega}k, \quad D^2\omega(k) = \frac{1}{\omega}(I - \nabla\omega(k)(\nabla\omega(k))^T).$$

We select again $k_0 \in \mathbb{R}$ and set $\omega_0 := \omega(k_0), v_g := \nabla\omega(k_0)$. Ansatz (2.31) generalizes to

$$u_{\text{as}}(x,t) = \varepsilon A(X,T)e^{\mathrm{i}(k_0\cdot x - \omega_0 t)} + \varepsilon^3\frac{A^3(X,T)}{9|k_0|^2 - 9\omega_0^2 + 1}e^{3\mathrm{i}(k_0\cdot x - \omega_0 t)} + \text{c.c.} \tag{2.38}$$

and the effective NLS is

$$\mathrm{i}\partial_T A + \frac{1}{2\omega_0}\nabla\cdot\left((I - v_g v_g^T)\nabla A\right) + \frac{3}{2\omega_0}|A|^2 A = 0. \tag{2.39}$$

The error is now denoted by

$$\varepsilon^\alpha R := u - u_{\text{as}},$$

where the optimal $\alpha \in \mathbb{R}$ is to be determined. The ansatz is now $O(\varepsilon^{1-\frac{n}{2}})$ in $\|\cdot\|_{L^2}$ (and any $\|\cdot\|_{H^s}, s \geqslant 0$,) because $\|\varepsilon A(\varepsilon\cdot, \varepsilon^2 t)\|_{L^2(\mathbb{R}^n)} = \varepsilon^{1-n/2}\|A(\cdot, \varepsilon^2 t)\|_{L^2(\mathbb{R}^n)}$. Hence, we would like to show that for some $\alpha > 1 - n/2$ it is $\|R(\cdot,t)\|_{H^s(\mathbb{R}^n)} = O(1)$ for $t \in [0, T_0\varepsilon^{-2}]$ and ε small enough assuming that A is a smooth solution of (2.39) on the time interval $[0, T_0]$. The equation for R is analogously to (2.33)

$$\begin{aligned}&\partial_t^2 R = \Delta R - R + f(R), \; x \in \mathbb{R}^n, t > 0 \quad \text{where} \;\; f(R) := 3u_{\text{as}}^2 R + 3\varepsilon^{\alpha+1}u_{\text{as}}R^2 + \varepsilon^{2\alpha}R^3 - \varepsilon^{-\alpha}\mathrm{R}es \\ &R(x,0) = \partial_t R(x,0) = 0.\end{aligned} \tag{2.40}$$

Once again, the first step is to apply a fixed point argument in order to show the existence of a solution of (2.40) on an interval $[0, t_*]$ with $t_* > 0$. However, the H^1 solution space, which we used in the case $n = 1$, is not sufficient since it is not an algebra in $\mathbb{R}^n, n \geqslant 2$. Hence, we choose

$$X(t_*) := C([0, t_*], H^s(\mathbb{R}^n)) \cap C^1([0, t_*], H^{s-1}(\mathbb{R}^n)), \quad s > n/2$$

as the solution space. For the estimates below we need an assumption on the solution A of (2.39). We require

$$A \in C^2([0, T_0], H^{s+1+\frac{n}{2}}(\mathbb{R}^n)) \quad \text{for some } T_0 > 0. \tag{2.41}$$

For the estimate of $\|f(\rho)(\cdot, t)\|_{H^{s-1}(\mathbb{R}^n)}$ we start again with the residual. It has the same form as in (2.37) except for the spatial derivatives replaced by their higher dimensional counterparts. For the L^2 norm we get

$$\|\mathrm{R}es(\cdot, t)\|_{L^2(\mathbb{R}^n)} \leqslant \varepsilon^4 \sum_m \left(\varepsilon^{-n} \int_{\mathbb{R}^n} |\tilde{G}_m(y, \varepsilon^2 t)|^2 dy\right)^{1/2} = c(A)\varepsilon^{4-n/2} \quad \text{for all } t \in [0, T_0\varepsilon^{-2}],$$

where $\tilde{G}_m$ are the analogues to G_m in Sec. 2.2.2.1 and $c(A)$ is a constant dependent only of the L^2 norm of A and its derivatives. The same ε scaling as for the L^2 norm applies also for H^s and hence

$$\|\mathrm{R}es(\cdot, t)\|_{H^s(\mathbb{R}^n)} \leqslant \tilde{c}(A)\varepsilon^{4-n/2} \quad \text{for all } t \in [0, T_0\varepsilon^{-2}],$$

where

$$\tilde{c}(A) = \tilde{c}\left(\max_{\substack{0 \leqslant p \leqslant 2 \\ 0 \leqslant q \leqslant 2}} \sup_{T \in [0, T_0]} \|\partial_T^q A(\cdot, T)\|_{H^{s+q}(\mathbb{R}^n)}\right).$$

For the other terms in $f(\rho)$ we have

$$\begin{aligned}
&\|u_{\text{as}}^2 \rho\|_{H^{s-1}} \leqslant c\|u_{\text{as}}^2\|_{C_0^{\lceil s-1 \rceil}} \|\rho\|_{H^{s-1}} \leqslant c\varepsilon^2 \|\rho\|_{H^{s-1}} \\
&\|u_{\text{as}} \rho^2\|_{H^{s-1}} \leqslant c\|u_{\text{as}}\|_{C_0^{\lceil s-1 \rceil}} \|\rho^2\|_{H^s} \leqslant c\varepsilon \|\rho\|_{H^s}^2 \\
&\|\rho^3\|_{H^{s-1}} \leqslant \|\rho^3\|_{H^s} \leqslant c\varepsilon \|\rho\|_{H^s}^3,
\end{aligned}$$

where in the first and second line assumption (2.41) has been used and $\lceil s \rceil$ is the smallest integer larger or equal to s. In summary

$$\|f(\rho)(\cdot, t)\|_{H^{s-1}(\mathbb{R}^n)} \leqslant c\left(\varepsilon^2 \|\rho\|_{H^{s-1}} + \varepsilon^{\alpha+1} \|\rho\|_{H^s}^2 + \varepsilon^{2\alpha} \|\rho\|_{H^s}^3 + \varepsilon^{4-\frac{n}{2}-\alpha}\right) \tag{2.42}$$

and the fixed point argument follows in complete analogy to the case $n = 1$. Once again, we can conclude that for each $t_*, r > 0$ there exists $\varepsilon_0 > 0$ such that (2.40) with $\varepsilon \in (0, \varepsilon_0)$ has a unique solution $R \in B_r \subset X(t_*)$, i.e.

$$\sup_{t \in [0, t_*]} \left(\|R(\cdot, t)\|_{H^s(\mathbb{R}^n)} + \|\partial_t R(\cdot, t)\|_{H^{s-1}(\mathbb{R}^n)}\right) \leqslant r.$$

It remains to extend the interval on which we can control the H^s norm of R onto an interval of length $O(\varepsilon^{-2})$. We attempt this again using a Gronwall argument. Let us define the energy

$$E(t) := \int_{\mathbb{R}^n} |\widehat{R}(k, t)|^2 (1 + |k|^2)^s + |\partial_t \widehat{R}(k, t)|^2 (1 + |k|^2)^{s-1} dk.$$

Note that although the integrals in E do not exactly match the H^s and H^{s-1} norms of $R(\cdot, t)$ and $\partial_t R(\cdot, t)$ as defined in Definition A.13, we have the equivalence (with some $c_1, c_2 > 0$)

$$c_1(\|R(\cdot, t)\|_{H^s(\mathbb{R}^n)} + \|\partial_t R(\cdot, t)\|_{H^{s-1}(\mathbb{R}^n)}) \leqslant E(t) \leqslant c_2(\|R(\cdot, t)\|_{H^s(\mathbb{R}^n)} + \|\partial_t R(\cdot, t)\|_{H^{s-1}(\mathbb{R}^n)}).$$

To obtain an inequality for E, we first differentiate

$$\begin{aligned}E'(t) &= 2\mathrm{Re}\int_{\mathbb{R}^n} \widehat{R}(k,t)\overline{\partial_t\widehat{R}}(k,t)(1+|k|^2)^s + \overline{\partial_t\widehat{R}}(k,t)\partial_t^2\widehat{R}(k,t)(1+|k|^2)^{s-1}dk \\ &= 2\mathrm{Re}\int_{\mathbb{R}^n} \overline{\partial_t\widehat{R}}(k,t)\widehat{f(R)}(k,t)(1+|k|^2)^{s-1}dk \\ &\leqslant 2\|\partial_t\widehat{R}(\cdot,t)(1+|\cdot|^2)^{\frac{s-1}{2}}\|_{L^2(\mathbb{R}^n)}\|\widehat{f(R)}(\cdot,t)(1+|\cdot|^2)^{\frac{s-1}{2}}\|_{L^2(\mathbb{R}^n)} \leqslant c\|\partial_t R\|_{H^{s-1}(\mathbb{R}^n)}\|f(R)\|_{H^{s-1}(\mathbb{R}^n)}.\end{aligned}$$

With (2.42) this leads to

$$\begin{aligned}E'(t) &\leqslant c\|\partial_t R\|_{H^{s-1}(\mathbb{R}^n)}\left(\varepsilon^2\|R\|_{H^{s-1}} + \varepsilon^{\alpha+1}\|R\|_{H^s}^2 + \varepsilon^{2\alpha}\|R\|_{H^s}^3 + \varepsilon^{4-\frac{n}{2}-\alpha}\right) \\ &\leqslant cE^{1/2}\left(\varepsilon^2E^{1/2} + \varepsilon^{\alpha+1}E + \varepsilon^{2\alpha}E^{3/2} + \varepsilon^{4-\frac{n}{2}-\alpha}\right).\end{aligned}$$

After integration and using $E^{1/2} \leqslant 1 + E$ for the term $cE^{1/2}\varepsilon^{4-\frac{n}{2}-\alpha}$, we arrive at

$$E(t) \leqslant ct\varepsilon^{4-\frac{n}{2}-\alpha} + c\int_0^t \left(\varepsilon^2 + \varepsilon^{4-\frac{n}{2}-\alpha} + \varepsilon^{\alpha+1}E^{1/2}(s) + \varepsilon^{2\alpha}E(s)\right)E(s)ds. \tag{2.43}$$

Next, we choose $t_* > 0$ and $K > 0$ and exploiting the existence result obtained by the fixed point argument, we know that there is $\varepsilon_0 > 0$ such that $E(t) \leqslant K$ for all $\varepsilon \in (0, \varepsilon_0)$ ad $t \in [0, t_*]$. Hence

$$E(t) \leqslant ct\varepsilon^{4-\frac{n}{2}-\alpha} + c\int_0^t \left(\varepsilon^2 + \varepsilon^{4-\frac{n}{2}-\alpha} + \varepsilon^{\alpha+1}K^{1/2} + \varepsilon^{2\alpha}K\right)E(s)ds \quad \text{for all } t \in [0,t_*], \varepsilon \in (0,\varepsilon_0).$$

In order to be able to extend the estimate $E(t) \leqslant K$ for some suitable ε-independent K to an interval of length $O(\varepsilon^{-2})$, we certainly need $\varepsilon^{-2}\varepsilon^{4-\frac{n}{2}-\alpha} = O(1)$, i.e. $2 - n/2 - \alpha \geqslant 0$. Let us set

$$\alpha := 2 - n/2 - \gamma \quad \text{with } \gamma \geqslant 0.$$

Then

$$E(t) \leqslant ct\varepsilon^{2+\gamma} + c\int_0^t \left(\varepsilon^2 + \varepsilon^{2+\gamma} + \varepsilon^2(\varepsilon^{1-\frac{n}{2}-\gamma}K^{1/2} + \varepsilon^{2-n-2\gamma}K)\right)E(s)ds \quad \text{for all } t \in [0,t_*], \varepsilon \in (0,\varepsilon_0). \tag{2.44}$$

In order to be able to redefine K such that $E(t) \leqslant K$ on $[0, T_0\varepsilon^{-2}]$ for ε small enough, we need $\varepsilon^{1-\frac{n}{2}-\gamma}K^{1/2} + \varepsilon^{2-n-2\gamma}K$ to be bounded by a $K-$independent $O(1)$ constant. This can be achieved by choosing ε small but only if $1 - \frac{n}{2} - \gamma > 0$ and $2 - n - 2\gamma > 0$, i.e. if $\gamma < 1 - \frac{n}{2}$. This is impossible for $n \geqslant 2$ because $\gamma \geqslant 0$.

Apparently, for $n \geqslant 2$ an estimate on $[0, T_0\varepsilon^{-2}]$ cannot be obtained. The problem lies in the residual - if the residual was of higher order in ε, the argument would work on $[0, T_0\varepsilon^{-2}]$. A smaller residual can be achieved only by improving the ansatz (2.38). In other words, in the formal derivation in Sec. 2.2.1 the expansions have to carried out to a higher order in ε such that the residual at the higher order can be calculated and set to zero by a suitable extension of the ansatz. We refrain from this tedious calculation here.

Nevertheless, for $n = 2$ we can obtain an estimate on $[0, T_1\varepsilon^{-2}]$ with some $T_1 \leqslant T_0$, which is independent of ε. If we set $\gamma = 0$, i.e. $\alpha = 2 - n/2 = 1$, then (2.44) becomes

$$E(t) \leqslant ct\varepsilon^2 + c\int_0^t \varepsilon^2\left(2 + K^{1/2} + K\right)E(s)ds \quad \text{for all } t \in [0,t_*], \varepsilon \in (0,\varepsilon_0).$$

By Gronwall's inequality $E(t) \leqslant ct\varepsilon^2 e^{ct\varepsilon^2(2+K^{1/2}+K)}$. Because

$$cT_1e^{cT_1(2+K^{1/2}+K)} \leqslant K \quad \text{for some } K > 0 \text{ if } T_1 > 0 \text{ is small enough,}$$

we have an estimate on $[0, T_1\varepsilon^{-2}]$. In detail, one first finds $T_1 > 0$ such that $cT_1e^{cT_1(2+K^{1/2}+K)} \leqslant K$ for one $K > 0$. An arbitrary such K is then selected. This leads to

Theorem 2.5. *Let $s > 1$, let $A \in C^2([0,T_0], H^{s+2}(\mathbb{R}^2))$ with some $T_0 > 0$ be a solution of the NLS (2.39) with $n = 2$ and u be a solution of (2.29) with $n = 2$, $u(x,0) = u_{as}(x,0)$, $\partial_t u(x,0) = \partial_t u_{as}(x,0)$ with u_{as} given by (2.38). There exist $T_1 \in (0,T_0], K = K(T_1) > 0$, and $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0,\varepsilon_0)$*

$$\|u(\cdot,t) - u_{as}(\cdot,t)\|_{H^s(\mathbb{R}^2)} \leqslant K\varepsilon \quad \text{for all } t \in [0, T_1\varepsilon^{-2}].$$

Remark 2.3. *Note that the estimate $E^{1/2} \leqslant 1 + E$, which we have used above, is relatively crude. Unfortunately, improving on this step does not lead to a better final error estimate. Without this step we get instead of (2.43) the inequality*

$$E(t) \leqslant c\int_0^t \varepsilon^2 \left(\varepsilon^{2-\frac{n}{2}-\alpha} + E^{1/2}(s) + \varepsilon^{\alpha-1}E(s) + \varepsilon^{2(\alpha-1)}E^{3/2}(s)\right) E^{1/2}(s)ds.$$

If $\alpha \geqslant 1$ and $2 - \frac{n}{2} - \alpha \geqslant 1$ (which can be satisfied only for $n \leqslant 2$ and for $n = 2$ it requires $\alpha = 1$), then

$$E(t) \leqslant c\int_0^t \varepsilon^2 \left(1 + K^{1/2} + K + K^{3/2}\right) E^{1/2}(s)ds \quad \text{for all } t \in [0,t_*], \varepsilon \in (0,\varepsilon_0).$$

We can now use a nonlinear version of Gronwall's inequality in Lemma 2.6 and get

$$E(t) \leqslant \left(\frac{c}{2}t\varepsilon^2 \left(1 + K^{1/2} + K + K^{3/2}\right)\right)^2 \quad \text{for all } t \in [0,t_*], \varepsilon \in (0,\varepsilon_0).$$

Again, by a choice of small enough $T_1 \leqslant T_0$ we can achieve $E(t) \leqslant K$ for all $t \in [0, T_1\varepsilon^{-2}]$ and $\varepsilon \in (0,\varepsilon_0)$. In summary, we arrive again at Theorem 2.5.

Since for $n > 2$ the conditions $\alpha \geqslant 1$ and $2 - \frac{n}{2} - \alpha \geqslant 1$ cannot be satisfied, this approach works only for $n \leqslant 2$.

Lemma 2.6 (Nonlinear Gronwall's inequality). *If $u(t) \leqslant \int_0^t a(s)\sqrt{u(s)}ds$ for some $a \in C((0,\infty),[0,\infty))$, then $u(t) \leqslant \frac{1}{4}(\int_0^t a(s)ds)^2$.*

Proof. Letting $v(t) := \int_0^t a(s)\sqrt{u(s)}ds$, we get $v'(t) = a(t)\sqrt{u(t)} \leqslant a(t)\sqrt{v(t)}$ and hence

$$\frac{d}{dt}\sqrt{v(t)} = \left(2\sqrt{v(t)}\right)^{-1} v'(t) \leqslant \frac{1}{2}a(t)$$

such that after integration

$$\sqrt{v(t)} \leqslant \tfrac{1}{2}\int_0^t a(s)ds.$$

The statement follows because $u(t) \leqslant v(t)$. □

Remark 2.4. *For $n > 2$ ansatz (2.38) produces an estimate only on a time scale $O(\varepsilon^{-\beta}), \beta < 2$. For this time scale we need in (2.43) $4 - \frac{n}{2} - \alpha - \beta \geqslant 0$. Setting $\alpha := 4 - \frac{n}{2} - \beta$, we have*

$$E(t) \leqslant ct\varepsilon^\beta + c\int_0^t \left(\varepsilon^2 + \varepsilon^\beta + \varepsilon^\beta(\varepsilon^{5-\frac{n}{2}-2\beta}K^{1/2} + \varepsilon^{8-n-3\beta}K\right) E(s)ds \quad \text{for all } t \in [0,t_*], \varepsilon \in (0,\varepsilon_0).$$

If $\beta < \min\{\frac{5}{2} - \frac{n}{4}, \frac{8}{3} - \frac{n}{3}\}$, then we can

(1) set $K := cT_0e^{3cT_0}$,

(2) choose $\varepsilon_0 > 0$ such that $\varepsilon_0^{5-\frac{n}{2}-2\beta}K^{1/2} + \varepsilon_0^{8-n-3\beta}K < 1$

and with Gronwall's inequality obtain

$$E(t) \leqslant ct\varepsilon^{\beta} e^{3ct\varepsilon^{\beta}} \leqslant cT_0 e^{3cT_0} = K \quad \textit{for all } t \in [0, T_0\varepsilon^{-\beta}], \varepsilon \in (0, \varepsilon_0).$$

This leads to

Theorem 2.7. *Let $s > n/2, \beta < \min\{2, \frac{5}{2} - \frac{n}{4}, \frac{8}{3} - \frac{n}{3}\}$, let $A \in C^2([0, T_0], H^{s+1+\frac{n}{2}}(\mathbb{R}^n))$ with some $T_0 > 0$ be a solution of the NLS* (2.39) *and u be a solution of* (2.29)*, $u(x,0) = u_{as}(x,0)$, $\partial_t u(x,0) = \partial_t u_{as}(x,0)$ with u_{as} given by* (2.38)*. There exist $K > 0$ and $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0)$*

$$\|u(\cdot, t) - u_{as}(\cdot, t)\|_{H^s(\mathbb{R}^n)} \leqslant K\varepsilon^{4-\frac{n}{2}-\beta} \quad \textit{for all } t \in [0, T_0\varepsilon^{-\beta}].$$

Of course, the time scale $O(\varepsilon^{-\beta}), \beta < 2$ is not satisfactory as the time scale of the modulation in u_{as} is $O(\varepsilon^{-2})$.

2.2.2.3 Justification in $L^1(\mathbb{R}^n)$ in the Fourier Variables

An alternative approach to Sec. 2.2.2.1 and 2.2.2.2 is to consider the problem in Fourier variables and estimate the error in the $L^1(\mathbb{R}^n)$ norm. Via the Riemann-Lebesgue lemma (Theorem A.8) this then implies an estimate of the supremum norm of the error. A major advantage of this approach is the fact that in L^1 no powers of ε are lost when estimating terms of the form $g(x) := f(\varepsilon x)$. Indeed,

$$\widehat{g}(k) = \varepsilon^{-n}\widehat{f}(\tfrac{k}{\varepsilon})$$

such that

$$\|\widehat{g}\|_{L^1(\mathbb{R}^n)} = \|\varepsilon^{-n}\widehat{f}(\tfrac{\cdot}{\varepsilon})\|_{L^1(\mathbb{R}^n)} = \|\widehat{f}\|_{L^1(\mathbb{R}^n)}.$$

Of course, this approach gives no estimate on derivatives of the error.

This approach is carried out, for example, in Section 2 of [12].

2.3 Hamiltonian Structure of KdV and NLS

The KdV and NLS are examples of Hamiltonian partial differential equations as we show below. Let us first recall the definition of Hamiltonian ODEs.

Definition 2.8. *An ODE system $\frac{dy}{dt} = f(y)$ with $f : \mathbb{R}^M \to \mathbb{R}^M$ for some $M \in \mathbb{N}$ is called Hamiltonian if there exists a function $H : \mathbb{R}^M \to \mathbb{R}$ and a skew-symmetric matrix $\omega \in \mathbb{R}^{M\times M}$ such that*

$$f = \omega\nabla H.$$

H is called the Hamiltonian function.

Darboux's theorem on symplectic manifolds [7] implies that there is always a transformation to the canonical variables, i.e. $\varphi : y \to (q, p, c_1, \ldots, c_r)$ with $q, p \in \mathbb{R}^N, 2N + r = M$ such that

$$\frac{d}{dx}\begin{pmatrix} q \\ p \end{pmatrix} = \begin{pmatrix} 0 & I_N \\ -I_N & 0 \end{pmatrix} \nabla\tilde{H}(q,p), \qquad \frac{dc_j}{dt} = 0, j = 1, \ldots, r,$$

where $\tilde{H}(q, p, c_1, \ldots, c_r) = H(\varphi^{-1}(q, p, c_1, \ldots, c_r))$.

The Hamiltonian function is a conserved quantity of the ODE $\frac{dy}{dt} = f(y) = \omega\nabla H$ because

$$\frac{d}{dt}H(y) = \nabla H(y)^T\frac{dy}{dt} = \nabla H(y)^T\omega H(y) = -(\omega\nabla H(y))^T\nabla H(y) = -\left(\frac{dy}{dt}\right)^T \nabla H(y)$$

and hence $\frac{d}{dt}H(y) = 0$.

Hamiltonian PDEs are defined analogously as problems of the form

$$\partial_t u = B\nabla H(u),$$

where $u(\cdot,t) \in X^M$, $M \in \mathbb{N}$, X is a real Hilbert space, $u(\cdot,t) : \mathbb{R}^n \to \mathbb{R}^M$ for some $n, M \in \mathbb{N}$, B is a skew-adjoint operator, $H : X^M \to \mathbb{R}$ is a functional and

$$\nabla H(u) = (\partial_{u_1} H(u), \ldots, \partial_{u_M} H(u))^T,$$

where $\partial_{u_j} H(u), j = 1, \ldots, M$ is the so called variational derivative of H with respect to u_j, i.e. $\partial_{u_j} H(u)$ satisfies

$$\langle \partial_{u_j} H(u), w \rangle_{L^2(\mathbb{R}^n)} = D_{u_j} H(u)\langle w \rangle \quad \text{for all } w \in X$$

and $D_{u_j} H$ is the Fréchet derivative of H with respect to u_j.

Let us briefly recall the concept of the Fréchet derivative. It can be defined using the more general Gâteaux derivative.

Definition 2.9. *Let e_j be the $j-$th Euclidean unit vector in $\mathbb{R}^M$. If $d_{u_j} H(u)\langle w \rangle := \lim_{\varepsilon \to 0} \frac{H(u+\varepsilon w e_j) - H(u)}{\varepsilon}$ exists for every $w \in X$, then we call $d_{u_j} H(u) : X \to \mathbb{R}, w \mapsto d_{u_j} H(u)\langle w \rangle$ the Gâteaux derivative of H with respect to u and $d_{u_j} H(u)\langle w \rangle$ the Gâteaux derivative of H with respect to u in the direction w.*

Definition 2.10. *When $u \mapsto d_{u_j} H(u)$ is continuous, then $d_{u_j} H$ is called the Freéchet derivative of H with respect to u_j.*

Remark 2.5. *Note that $\nabla H(u)$ is not the variational gradient of H. The variational gradient is defined as the Riesz representation of $D_u H(u)$, i.e. as the element v of X^M such that $\langle v, w \rangle_{X^M} = D_u H(u)\langle w \rangle := \sum_{j=1}^M D_{u_j} H(u)\langle w_j \rangle$ for all $w \in X^M$.*

As we see next, it is easy to show that H is a conserved quantity also for Hamiltonian PDEs:

$$\begin{aligned}\frac{dH(u)}{dt} &= D_u H(u)\langle \partial_t u \rangle = \int_{\mathbb{R}^n} \nabla H(u) \cdot \partial_t u dx = \int_{\mathbb{R}^n} \nabla H(u) \cdot (B\nabla H(u)) dx \\ &= \langle \nabla H(u), B\nabla H(u) \rangle_{L^2(\mathbb{R}^n)} = -\langle B\nabla H(u), \nabla H(u) \rangle_{L^2(\mathbb{R}^n)} = -\int_{\mathbb{R}^n} \partial_t u \cdot \nabla H(u) dx = -\frac{dH(u)}{dt}.\end{aligned}$$

As we show next, the KdV

$$\partial_t u = -\partial_x^3 u - 6u\partial_x u, \qquad x \in \mathbb{R}$$

is Hamiltonian with

$$H = \int_{\mathbb{R}} \frac{1}{2}(\partial_x u)^2 - u^3 dx.$$

In the general definition of Hamiltonian PDEs we thus have $M = 1$ and, for example $X = H^1(\mathbb{R})$. To check that H is really the corresponding Hamiltonian functional, we have per definition of the Fréchet derivative

$$\begin{aligned}D_u H(u)\langle w \rangle &= \lim_{\varepsilon \to 0} \varepsilon^{-1} \int_{\mathbb{R}} \frac{1}{2}\left(\partial_x(u + \varepsilon w)\right)^2 - (u + \varepsilon w)^3 - \frac{1}{2}\left(\partial_x u\right)^2 + u^3 dx \\ &= \lim_{\varepsilon \to 0} \varepsilon^{-1} \int_{\mathbb{R}} \varepsilon \partial_x u \partial_x w - 3u^2 w + O(\varepsilon^2) dx = \int_{\mathbb{R}} \left(-\partial_x^2 u - 3u^2\right) w dx,\end{aligned}$$

such that $\partial_u H = -\partial_x^2 u - 3u^2$. Clearly, KdV is equivalent to

$$\partial_t u = B\partial_u H(u) \quad \text{with } B = \partial_x.$$

For the cubic NLS

$$\mathrm{i}\partial_t \psi = -\Delta\psi + \gamma|\psi|^2\psi, \qquad \gamma \in \mathbb{R}, x \in \mathbb{R}^n$$

we first pose the equation in the real variables $u := \mathrm{Re}(\psi), v := \mathrm{Im}(\psi)$

$$\begin{aligned}\partial_t u &= -\Delta v + \gamma(u^2 + v^2)v \\ \partial_t v &= \Delta u - \gamma(u^2 + v^2)u.\end{aligned}$$

Here the corresponding Hamiltonian is

$$H(u, v) := \int_{\mathbb{R}^n} \frac{1}{2}\left(|\nabla u|^2 + |\nabla v|^2\right) + \frac{\gamma}{4}(u^2 + v^2)^2 dx$$

because

$$\begin{aligned}D_u H(u, v)\langle w\rangle &= \lim_{\varepsilon\to 0} \varepsilon^{-1} \int_{\mathbb{R}^n} \varepsilon\nabla u \cdot \nabla w + \gamma\varepsilon(uwv^2 + u^2vw) + O(\varepsilon^2)dx \\ &= \int_{\mathbb{R}^n} \left(-\Delta u + \gamma u(v^2 + u^2)\right) w dx,\end{aligned}$$

such that $\partial_u H(u, v) = -\Delta u + \gamma(v^2 + u^2)u$. Similarly, $\partial_v H(u, v) = -\Delta v + \gamma(v^2 + u^2)v$. This leads to

$$\partial_t \begin{pmatrix} u \\ v \end{pmatrix} = B\nabla H(u, v), \quad \text{where } B := \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}.$$

2.4 Orbital Stability of the KdV 1-Soliton

Before discussing the stability of PDE solutions, let us recall Lyapunov's result on stability of critical points $y_c \in \mathbb{R}^n$ of autonomous ODE systems

$$\frac{dy}{dt} = f(y), \quad f \in C(\mathbb{R}^n, \mathbb{R}^n),$$

where $f(y_c) = 0$.

The critical point y_c is called stable if for every $\varepsilon > 0$ one can find $\delta > 0$ such that if $|y(0) - y_c| < \delta$, then $|y(t) - y_c| < \varepsilon$ for all $t > 0$. Lyapunov proved [25]

Theorem 2.11. *If there is a function $V \in C^1(\mathbb{R}^n)$ such that $V(y_c) = 0$ and a neighborhood $U(y_c) \subset \mathbb{R}^n$, for which*

(i) $V(y) > 0$ for every $y \in U(y_c)\backslash\{y_c\}$,

(ii) $\nabla V(y) \cdot f(y) \leqslant 0$ for every $y \in U(y_c)$,

then y_c is stable.

A function satisfying the assumption of the theorem is called a Lyapunov function (on the neighborhood U).

Critical points of Hamiltonian ODE systems are given by extrema of the Hamiltonian function H. For minima and maxima of H a simple shift of H then produces a Laypunov function.

Theorem 2.12. *Let $\frac{dy}{dt} = \omega\nabla H(y)$ be a Hamiltonian ODE-system. Local maxima and minima of H are stable critical points.*

Proof. Let us first consider the case that y_c is a local minimum of H. Then $V(y) := H(y) - H(y_c)$ is a Lyapunov function on a neighborhood of y_c because clearly $V(y_c) = 0, V(y) > 0$ on a punctured neighborhood of y_c and

$$\nabla V(y) \cdot f(y) = \nabla H(y) \cdot (\omega \nabla H(y)) = 0$$

since $\omega = -\omega^T$.

For a local maximum one defines $V(y) := H(y_c) - H(y)$. □

Let us now turn to PDEs. We consider the KdV

$$\partial_t u + \partial_x^3 u + 6u\partial_x u = 0, \quad x \in \mathbb{R}$$

and study the satbility of localized traveling waves of the form

$$u(x,t) = u_c(x,t) := g_c(x - ct) \quad \text{with } c \in \mathbb{R}, g_c^{(n)}(x) \to 0 \quad \text{for } n = 0,1,2 \quad \text{as } |x| \to \infty.$$

As shown in Section 2.1.1.3, the only nontrivial solution is

$$g_c(\xi) = \frac{c}{2}\text{sech}^2(\sqrt{c}\xi), \quad c > 0, \tag{2.45}$$

modulo an arbitrary constant shift in the argument, see (2.15).

The profile g_c satisfies $(g_c'' - cg_c + 3g_c^2)' = 0$ and because of the decay of $g_c^{(n)}$, $n = 0,1,2$ we have after integration

$$g_c'' - cg_c + 3g_c^2 = 0. \tag{2.46}$$

As we have shown above, KdV is Hamiltonian with

$$H(u) = \int_{\mathbb{R}} \tfrac{1}{2}(\partial_x u)^2 - u^3 dx.$$

One can easily check that another conserved quantity is

$$N(u) := \int_{\mathbb{R}} u^2 dx.$$

The profile g_c is a critical point of H under the constraint of fixed L^2-norm, i.e. restricted to $\{w \in H^1(\mathbb{R}) : N(w) = N(g_c)\}$. This follows from the Lagrange multiplier rule because for $E(w,\lambda) := H(w) + \lambda(N(w) - N(g_c))$ we get

$$\begin{aligned} \partial_w E(g_c, c) &= -g_c'' - 3g_c^2 + cg_c = 0 \quad \text{due to (2.46)}, \\ \partial_c E(g_c, c) &= N(w) - N(g_c) = 0. \end{aligned}$$

In analogy to ODEs one could expect that minima and maxima of H restricted to $\{w \in H^1(\mathbb{R}) : N(w) = N(g_c)\}$ are stable. One can hope hat because H is a conserved quantity, if $|H(u(\cdot,0)) - H(g_c)|$ is small, then $u(\cdot,t)$ stays close to $u_c(\cdot,t)$. The problem is that in infinite dimensional problems the smallness of $|H(u(\cdot,t)) - H(u_c(\cdot,t))|$ does not imply the smallness of $\|u(\cdot,t) - u_c(\cdot,t)\|_{H^1}$ (or in any other generic norm). In particular, any perturbation of the initial data which leads to a change in the velocity of the solution means that u propagates at a different velocity than u_c and even if the shape of u remains close to that of the shape of u_c, the difference $\|u(\cdot,t) - u_c(\cdot,t)\|_{H^1}$ does not remain small. In fact it converges to $\|u(\cdot,t)\|_{H^1} + \|u_c(\cdot,t)\|_{H^1}$ as $t \to \infty$.

This motivates a generalized definition of stability, which describes stability of the shape of u.

Definition 2.13. *u_c is orbitally stable if for each $\varepsilon > 0$ one can find $\delta > 0$ such that if $\|u(\cdot,0) - g_c\|_{H^1(\mathbb{R})} < \delta$, then*

$$d(u(\cdot,t), u_c(\cdot,t)) := \inf_{x_0 \in \mathbb{R}} \|u(\cdot + x_0, t) - u_c(\cdot,t)\|_{H^1(\mathbb{R})} < \varepsilon \quad \textit{for all } t > 0. \tag{2.47}$$

Orbital stability of u_c was proved by Benjamin in 1972, cf. [4]. The main result is the following statement. Let $A > 0$. Then there are constants $\alpha, \beta > 0$ such that

$$0 \leqslant \alpha d(g, g_c)^2 \leqslant H(g) - H(g_c) \leqslant \beta \|g - g_c\|^2_{H^1(\mathbb{R})} \quad \text{for all } g \in \{w \in H^1(\mathbb{R}) : N(w) = N(g_c), \|w - g_c\|_{H^1(\mathbb{R})} \leqslant A\}. \tag{2.48}$$

We prove only the second inequality in (2.48). We write $g = g_c + \varphi$ and because $N(g) = N(g_c)$, we have

$$\begin{aligned} H(g) - H(g_c) &= \int_{\mathbb{R}} \partial_x g_c \partial_x \varphi + \frac{1}{2}(\partial_x \varphi)^2 - 3g_c^2 \varphi + 3 g_c \varphi^2 - \varphi^3 dx + \frac{c}{2} \int_{\mathbb{R}} (g_c + \varphi)^2 - g_c^2 dx \\ &= \int_{\mathbb{R}} \underbrace{\left(-\partial_x^2 u_c - 3g_c^2 + c g_c\right)}_{=0} \varphi dx + \int_{\mathbb{R}} \tfrac{1}{2}(\partial_x \varphi)^2 - \varphi^3 + (\tfrac{c}{2} - 3g_c)\varphi^2 dx \\ &\leqslant \tfrac{1}{2}\|\partial_x \varphi\|^2_{L^2(\mathbb{R})} + (\tfrac{c}{2} + 3\|g_c\|_{C_b^0(\mathbb{R})})\|\varphi\|^2_{L^2(\mathbb{R})} + \|\varphi\|_{C_b^0(\mathbb{R})}\|\varphi\|^2_{L^2(\mathbb{R})} \leqslant c_1 \|\varphi\|^2_{H^1(\mathbb{R})} + c_2 \|\varphi\|^3_{H^1(\mathbb{R})} \\ &\leqslant (c_1 + c_2 A)\|\varphi\|^2_{H^1(\mathbb{R})} \quad \text{because } \|\varphi\|_{H^1(\mathbb{R})} \leqslant A. \end{aligned}$$

In the one but last step we have used the Sobolev embedding in Theorem A.14.

The proof of the inequality $d(g, g_c)^2 \leqslant H(g) - H(g_c)$ is substantially lengthier, see [4].

1) Orbital stability within initial data $\{w \in H^1(\mathbb{R}) : N(w) = N(g_c)\}$ Orbital stability within initial data that have the same L^2 norm as $u_c(\cdot, t)$ (i.e. as g_c) follows directly from (2.48) and the conservation of H:

$$\alpha d(u(\cdot, t), u_c(\cdot, t))^2 \leqslant H(u(\cdot, t)) - H(u_c(\cdot, t)) = H(u(\cdot, 0)) - H(u_c(\cdot, 0)) \leqslant \beta \|u(\cdot, 0) - u_c(\cdot, 0)\|^2_{H^1(\mathbb{R})}.$$

2) Orbital stability within $H^1(\mathbb{R})$ Here we use the property that for any $c > 0$ and any $\delta > 0$ small enough and initial data $u(\cdot, 0)$ such that $\|u(\cdot, 0) - g_c\|_{H^1(\mathbb{R})} < \delta$ there exists $c_* > 0$ with $N(g_{c_*}) = N(u(\cdot, 0))$ and $\|g_c - g_{c_*}\|_{H^1(\mathbb{R})} \to 0$ as $\delta \to 0$. This can be checked directly from the explicit form $g_c(\xi) = \frac{c}{2}\text{sech}^2(\sqrt{c}\xi)$, for which $N(g_c) = \frac{8}{3}c^{2/3}$. Clearly, N is continuous and surjective onto $(0, \infty)$ and g_c is smooth in c.

In other words, for any initial data $\delta-$close (in $\|\cdot\|_{H^1}$) to g_c there is a soliton g_{c_*} with the same L^2 norm as $u(\cdot, 0)$ and $o(1)$ close to g_c as $\delta \to 0$.

Using step 1, we have

$$d(u(\cdot, t), u_{c_*}(\cdot, t)) \to 0 \text{ as } \delta \to 0.$$

By the triangle inequality of Lemma 2.14, we get

$$d(u(\cdot, t), u_c(\cdot, t)) \leqslant d(u(\cdot, t), u_{c_*}(\cdot, t)) + \underbrace{d(u_{c_*}(\cdot, t), u_c(\cdot, t))}_{=:d_2} \to d_2 \text{ as } \delta \to 0.$$

Finally, for d_2 we have

$$d_2 = \inf_{x_0 \in \mathbb{R}} \|g_c(\cdot - ct) - g_{c_*}(\cdot + x_0 - c_* t)\|_{H^1(\mathbb{R})} = \inf_{x_0 \in \mathbb{R}} \|g_c(\cdot) - g_{c_*}(\cdot + x_0)\|_{H^1(\mathbb{R})} \to 0 \text{ as } \delta \to 0$$

because $\|g_c - g_{c_*}\|_{H^1(\mathbb{R})} \to 0$ as $\delta \to 0$. This concludes the proof of the orbital stability of u_c for any $c > 0$.

Lemma 2.14. *The distance function d in* (2.47) *satisfies the triangle inequality.*

Proof. For any $u, v, w \in H^1(\mathbb{R})$

$$\begin{aligned} d(u, v) &\leqslant \|u(\cdot + a) - v(\cdot + b)\|_{H^1(\mathbb{R})} \quad \text{for all } a, b \in \mathbb{R} \\ &\leqslant \|u(\cdot + a) - w(\cdot)\|_{H^1(\mathbb{R})} + \|w(\cdot) - v(\cdot + b)\|_{H^1(\mathbb{R})} \quad \text{for all } a, b \in \mathbb{R}. \end{aligned}$$

Taking the infimum over $a, b \in \mathbb{R}$ yields the result. ☐

Appendix A

Fourier Transform and Sobolev Spaces

Let us first recall the standard definitions of the weak derivative and the Sobolev space H^s.

Definition A.1. *For $u \in L^1_{loc}(\mathbb{R}^n)$ and a multiindex $\alpha \in \mathbb{N}_0^n$ a function v is a weak α-derivative if*

$$\int_{\mathbb{R}^n} u D^\alpha \varphi dx = (-1)^{|\alpha|} \int_{\mathbb{R}^n} v \varphi dx \quad \text{for all } \varphi \in C_c^\infty(\mathbb{R}^n).$$

We denote a weak derivative by $v = D^\alpha u$. We are using the standard notation $D^\alpha u = \partial_{x_1}^{\alpha_1} \partial_{x_2}^{\alpha_2} \ldots \partial_{x_n}^{\alpha_n} u$.

Weak derivatives are unique up to sets of measure zero.

Definition A.2. *The Sobolev space $H^s(\mathbb{R}^n)$ is*

$$H^s(\mathbb{R}^n) = \{u : \mathbb{R}^n \to \mathbb{C} \quad \text{such that } D^\alpha u \in L^2(\mathbb{R}^n) \text{ for all } \alpha \in \mathbb{N}_0^n \text{ with } |\alpha| \leqslant s.\}$$

The corresponding norm is

$$\|u\|_{H^s(\mathbb{R}^n)} = \left(\sum_{|\alpha| \leqslant s} \|D^\alpha u\|_{L^2(\mathbb{R}^n)} \right)^{\frac{1}{2}}.$$

A.1 Fourier Transform

One of the major tools in the mathematics of waves is the Fourier transformation.

Definition A.3. *For $u \in L^1(\mathbb{R}^n)$ we define the forward and backward Fourier transforms*

$$\widehat{u}(k) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{-\mathrm{i}k \cdot x} u(x) \, \mathrm{d}x \qquad \text{for } k \in \mathbb{R}^n,$$
$$\check{u}(x) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{\mathrm{i}k \cdot x} u(k) \, \mathrm{d}k \qquad \text{for } x \in \mathbb{R}^n$$

respectively.

The Fourier transformation can be also applied to L^2 functions by choosing an approximating sequence $(u_j)_{j \in \mathbb{N}} \subset L^2(\mathbb{R}^n) \cap L^1(\mathbb{R}^n)$ such that $u_j \to u$ in L^2 as $j \to \infty$ and defining $\widehat{u} := \lim_{j \to \infty} \widehat{u}_j$, see Sec. 4.3.1 in [9]. In other words there exists a unique bounded linear extension of $\widehat{\ }$ to $L^2(\mathbb{R}^n)$. The same holds for $\check{\ }$.

The following properties of the Fourier transform for arbitrary functions $u, v \in L^2(\mathbb{R}^n)$ will be fundamental for our analysis.

$$\|u\|_{L^2(\mathbb{R}^n)} = \|\widehat{u}\|_{L^2(\mathbb{R}^n)} = \|\check{u}\|_{L^2(\mathbb{R}^n)}, \tag{A.1}$$

$$(u, v) = (\widehat{u}, \widehat{v}) \text{ where } (\cdot, \cdot) \text{ is the standard inner product in } L^2(\mathbb{R}^n), \tag{A.2}$$

$$\widehat{D^\alpha u}(k) = (\mathrm{i}k)^\alpha \widehat{u}(k) \quad \text{for all multi-indices } \alpha \in \mathbb{N}_0^n \text{ with } |\alpha| \leqslant s \text{ if } u \in H^s(\mathbb{R}^n), \tag{A.3}$$

$$\widehat{(u * v)} = (2\pi)^{n/2}\widehat{u}\widehat{v}, \quad \widehat{(uv)} = (2\pi)^{-n/2}\widehat{u} * \widehat{v} \text{ if } u, v \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n), \tag{A.4}$$

$$u = \check{\widehat{u}}. \tag{A.5}$$

The identity (A.1) is called the Plancherel identity. For the proofs see Sec. 4.3.1 in [9]. Here we show the proofs of (A.2), (A.3), and (A.5).

First, using the Plancherel identity and the polarization identity we prove (A.2). *Proof.* (of (A.2)) The polarization identity is

$$(u, v) = \tfrac{1}{2}\left(\|u + v\|_{L^2}^2 - \mathrm{i}\|u + \mathrm{i}v\|_{L^2}^2 - (1 - \mathrm{i})\|u\|_{L^2}^2 - (1 - \mathrm{i})\|v\|_{L^2}^2\right).$$

Using the Plancherel identity, we thus get

$$(u, v) = \tfrac{1}{2}\left(\|\widehat{u} + \widehat{v}\|_{L^2}^2 - \mathrm{i}\|\widehat{u} + \mathrm{i}\widehat{v}\|_{L^2}^2 - (1 - \mathrm{i})\|\widehat{u}\|_{L^2}^2 - (1 - \mathrm{i})\|\widehat{v}\|_{L^2}^2\right) = (\widehat{u}, \widehat{v}).$$

□

The identity (A.3) shows that the Fourier transform acts on weak derivatives the same way as it does on classical derivatives. We prove this next.

Proof. (of (A.3)) First we note that for $\varphi \in C_c^\infty(\mathbb{R}^n)$ it is clear by integration by parts that $\widehat{D^\alpha\varphi}(k) = (\mathrm{i}k)^\alpha\widehat{\varphi}(k)$ We calculate next for all $\varphi \in C_c^\infty(\mathbb{R}^n)$

$$(\widehat{D^\alpha u}, \widehat{\varphi}) = (D^\alpha u, \varphi) = (-1)^{|\alpha|}(u, D^\alpha\varphi) = (-1)^{|\alpha|}(\widehat{u}, \widehat{D^\alpha\varphi}) = (-1)^{|\alpha|}\overline{\mathrm{i}^{|\alpha|}}(\widehat{u}, k^\alpha\widehat{\varphi}) = \mathrm{i}^{|\alpha|}(k^\alpha\widehat{u}, \widehat{\varphi}),$$

where the first and the third equation hold by the A.2. Hence with $\widehat{h} := \widehat{D^\alpha u} - \mathrm{i}^{|\alpha|}k^\alpha\widehat{u}$

$$0 = (\widehat{h}, \widehat{\varphi}) = (h, \varphi) \quad \text{for all } \varphi \in C_c^\infty(\mathbb{R}^n).$$

Because $C_c^\infty(B_R)$ is dense in $L^2(B_R)$ for all $R > 0$, we get $h = 0$ almost everywhere in B_R for all $R > 0$ and hence $h = 0$ almost everywhere in $\mathbb{R}^n$.

For (A.5) we first note that for $u, v \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ it is

$$(2\pi)^{-n/2}\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} u(x)v(y)e^{\mathrm{i}x\cdot y}\,\mathrm{d}x\,\mathrm{d}y = \int_{\mathbb{R}^n}\check{u}v\,\mathrm{d}x = \int_{\mathbb{R}^n}u\check{v}\,\mathrm{d}x$$

thanks to the theorems of Tonelli and Fubini (use Tonelli to show $\int_{\mathbb{R}^{2n}}|u((x)||v(y)|\,\mathrm{d}(x, y) < \infty$. Then $\int_{\mathbb{R}^{2n}}u((x)v(y)e^{\mathrm{i}x\cdot y}\,\mathrm{d}(x, y) = \int_{\mathbb{R}^n}u(x)\int_{\mathbb{R}^n}v(y)e^{\mathrm{i}x\cdot y}\,\mathrm{d}y\,\mathrm{d}x = \int_{\mathbb{R}^n}v(y)\int_{\mathbb{R}^n}u(x)e^{\mathrm{i}x\cdot y}\,\mathrm{d}x\,\mathrm{d}y$ by Fubini). Let now $u, v \in L^2(\mathbb{R}^n)$. The identity

$$\int_{\mathbb{R}^n}\check{u}v\,\mathrm{d}x = \int_{\mathbb{R}^n}u\check{v}\,\mathrm{d}x \tag{A.6}$$

holds also in that case as can be shown by approximation in $L^1 \cap L^2$.

Besides, the definition of the $\check{}$-transform implies

$$\check{v} = \overline{\widehat{\overline{v}}}. \tag{A.7}$$

Finally

$$\int_{\mathbb{R}^n} \check{\widehat{u}} v \,\mathrm{d}x \overset{(A.6)}{=} \int_{\mathbb{R}^n} \widehat{u}\check{v} \,\mathrm{d}x \overset{(A.7)}{=} \int_{\mathbb{R}^n} \widehat{u}\overline{\widehat{\overline{v}}} \,\mathrm{d}x \overset{(A.2)}{=} \int_{\mathbb{R}^n} u\overline{\overline{v}} \,\mathrm{d}x = \int_{\mathbb{R}^n} uv \,\mathrm{d}x.$$

Since $v \in L^2$ is arbitrary, we are done. □

The Fourier transform can be defined also for tempered distributions. Let us first define the Schwartz space and the space of tempered distributions.

Definition A.4. *The Schwartz space in* $\mathbb{R}^n$ *is* $S(\mathbb{R}^n) = \{\varphi \in C^\infty(\mathbb{R}^n) : \sup_{x\in\mathbb{R}^n} |x^\alpha D^\beta \varphi(x)| < \infty$ *for all* $\alpha, \beta \in \mathbb{N}_0^n\}$.

The space $S(\mathbb{R}^n)$ is a complete metric space with the metric $d(\varphi,\psi) := \sum_{|p|,|q|=0}^\infty 2^{-|p|-|q|} \frac{\|\varphi-\psi\|_{p,q}}{1+\|\varphi-\psi\|_{p,q}}$, where $\|\eta\|_{p,q} := \sup_{x\in\mathbb{R}^n} x^p |D^q(\varphi)(x)|$, see Sec. 7.3 in [19].

Definition A.5. *The space of tempered distributions* $S'(\mathbb{R}^n)$ *is the dual space of* $S(\mathbb{R}^n)$, *i.e. the space of all linear continuous mappings from* $S(\mathbb{R}^n)$ *to* $\mathbb{C}$.

Note that any locally integrable function $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ defines a tempered distribution T_f via

$$T_f(\varphi) = \int_{\mathbb{R}^n} f(x)\varphi(x)\,\mathrm{d}x \qquad \text{for all } \varphi \in S(\mathbb{R}^n).$$

The Fourier transform is a bijection from $S(\mathbb{R}^n)$ to $S(\mathbb{R}^n)$, see Theorem IX.1 in [18]. Hence, the following definition of the Fourier transform for tempered distributions makes sense.

Definition A.6. *For any* $T \in S'(\mathbb{R}^n)$ *we define the Fourier transform* $\widehat{T}$ *via*

$$\widehat{T}(\psi) = T(\widehat{\psi}) \qquad \textit{for all } \psi \in S(\mathbb{R}^n).$$

The Fourier transform is also a bijection from $S'(\mathbb{R}^n)$ to $S'(\mathbb{R}^n)$, see Theorem IX.2 in [18].

As the following lemma shows, the product rule (A.4) for the Fourier transform generalizes also to distributions.

Lemma A.7. *Let* $\varphi \in S(\mathbb{R}^n)$ *and* $T \in S'(\mathbb{R}^n)$ *and define*

$$(T * \varphi)(x) := T(\varphi(x - \cdot)).$$

Then $T * \varphi \in C^\infty(\mathbb{R}^n) \cap S'(\mathbb{R}^n)$ *and*

$$\widehat{T * \varphi} = \widehat{T}\widehat{\varphi}, \ \textit{where } \widehat{T}\widehat{\varphi} \in S'(\mathbb{R}^n), (\widehat{T}\widehat{\varphi})(\psi) := \widehat{T}(\widehat{\varphi}\psi) \textit{ for all } \psi \in S(\mathbb{R}^n).$$

The proof is left as an exercise.

A fundamental result relating decay in the physical space with smoothness in the Fourier space is the Riemann-Lebesgue lemma, cf. Theorem IX.7 in [18].

Theorem A.8. *(Riemann-Lebesgue lemma) The Fourier transform is a bounded map from* $L^1(\mathbb{R}^n)$ *to* $C_\infty(\mathbb{R}^n)$, *i.e. the space of continuous functions* $k \mapsto f(k)$ *decaying to* 0 *as* $|k| \to \infty$.

An easy important corollary of the Riemann-Lebesgue lemma is

Corollary A.9. *If* $f \in C(\Omega)$ *for some compact* $\Omega \subset \mathbb{R}^n$, *then*

$$\int_\Omega e^{-\mathrm{i}k\cdot x} f(x)\,\mathrm{d}x \to 0 \quad \textit{as} \quad |k| \to \infty.$$

Finally, we state an important result relating the support of a function (or generally a tempered distribution) in the physical $x-$space and the smoothness of its Fourier transform. This is one of the famous Paley-Wiener theorems. First, let us, however, define the support of a tempered distribution

Definition A.10. *The support of* $T \in S'(\mathbb{R}^n)$ *is*

$$\text{supp}(T) = \mathbb{R}^n \setminus \bigcup \{V \subset \mathbb{R}^n : T(\varphi) = 0 \textit{ for all } \varphi \in S(\mathbb{R}^n) \textit{ with } \text{supp}(\varphi) \subset V\}.$$

It is an easy exercise to check that for a function T this definition coincides with the definition of the support for functions.

Theorem A.11. *(Paley-Wiener Theorem) The following two statements are equivalent*

1. $\psi \in S'(\mathbb{R}^n), \text{supp}(\psi) \subset B_R = \{x \in \mathbb{R}^n : |x| \leqslant R\}$ *for some* $R > 0$,
2. $\hat{\psi} : \mathbb{C}^n \to \mathbb{C}$ *is an entire function on* $\mathbb{C}^n$ *and there are* $C > 0, N \in \mathbb{N}$ *such that*

$$|\widehat{\psi}(k)| \leqslant C(1 + |k|)^N e^{R|\operatorname{Im}(k)|} \quad \textit{for all} \quad k \in \mathbb{C}^n.$$

For the proof see Theorem IX.12 in [18] and Theorem 7.23 in [19].

A.2 Sobolev Spaces and their Definition in the Fourier Variables

As the following theorem shows, Sobolev spaces can be characterized via the decay rate in Fourier variables.

Theorem A.12. *(Characterization of Sobolev spaces in Fourier Variables) Let* $u \in L^2(\mathbb{R}^n)$ *and* $s \in \mathbb{N}_0$.

1. $$u \in H^s(\mathbb{R}^n) \quad \textit{if and only if } (1 + |k|^s)\widehat{u} \in L^2(\mathbb{R}^n).$$
2. *There are constants* $c_1, c_2 > 0$ *such that*

$$c_1 \|u\|_{H^s(\mathbb{R}^n)} \leqslant \|(1 + |k|^s)\widehat{u}\|_{L^2(\mathbb{R}^n)} \leqslant c_2 \|u\|_{H^s(\mathbb{R}^n)}$$

for all $u \in H^s(\mathbb{R}^n)$.

Proof. First let $u \in H^s(\mathbb{R}^n)$. Identity (A.3) implies $k^\alpha \widehat{u} \in L^2(\mathbb{R}^n)$ for all $\alpha \in \mathbb{N}_0^n$ such that $|\alpha| \leqslant s$. Choosing $\alpha = qe_j$ with $\mathbb{N}_0 \ni q \leqslant s, j \in \{1, \ldots, n\}$, we have by the Plancherel identity

$$\int_{\mathbb{R}^n} k_j^{2q} |\widehat{u}|^2 dk = \int_{\mathbb{R}^n} |D^{qe_j} u|^2 dx.$$

From this we get $\int_{\mathbb{R}^n} |k|^{2q} |\widehat{u}|^2 dk \leqslant c \sum_{j=1}^n \int_{\mathbb{R}^n} |D^{qe_j} u|^2 dx$ because $|k|^{2q} = (k_1^2 + \ldots, k_n^2)^q \leqslant c(k_1^{2q} + \cdots + k_n^{2q})$. We thus have

$$\int_{\mathbb{R}^n} (1 + |k|^s)^2 |\widehat{u}|^2 dk \leqslant c \|u\|^2_{H^s(\mathbb{R}^n)}.$$

This proves the implication $\Rightarrow$ in 1. and one part of the norm equivalence in 2.

Now assume $(1 + |k|^s)\widehat{u} \in L^2(\mathbb{R}^n)$. For $|\alpha| \leqslant s$

$$\|(\mathrm{i}k)^\alpha \widehat{u}\|^2_{L^2(\mathbb{R}^n)} \leqslant \int_{\mathbb{R}^n} |k|^{2\alpha} |\widehat{u}|^2 dk \leqslant c \|(1 + |k|^s)\widehat{u}\|^2_{L^2(\mathbb{R}^n)}, \tag{A.8}$$

where the first inequality follows from $|(\mathrm{i}k)^\alpha|^2 = k_1^{2\alpha_1} k_2^{2\alpha_2} \ldots k_n^{2\alpha_n} \leqslant (k_1^2 + \cdots + k_n^2)^{\alpha_1 + \cdots + \alpha_n} = |k|^{2|\alpha|}$. We set next

$$u_\alpha := ((\mathrm{i}k)^\alpha \widehat{u})\check{}.$$

Note that $u_\alpha \in L^2(\mathbb{R}^n)$. For all $\varphi \in C_c^\infty(\mathbb{R}^n)$ we obtain

$$(D^\alpha \varphi, u) = (\widehat{D^\alpha \varphi}, \widehat{u}) = ((\mathrm{i}k)^\alpha \widehat{\varphi}, \widehat{u}) = (-1)^{|\alpha|}(\widehat{\varphi}, (\mathrm{i}k)^\alpha \widehat{u}) = (-1)^{|\alpha|}(\varphi, u_\alpha),$$

i.e. u_α is the weak derivative $D^\alpha u$. From (A.8) we conclude that $\|u\|_{H^s(\mathbb{R}^n)} \leqslant c\|(1+|k|^s)\widehat{u}\|_{L^2(\mathbb{R}^n)}$, which proves the implication $\Leftarrow$ in 1. as well as the other inequality in 2. $\square$

Theorem A.12 lets us define Sobolev spaces for real (positive) indices s.

Definition A.13. *For $s \in (0,\infty)$ we define*

$$H^s(\mathbb{R}^n) := \{u \in L^2(\mathbb{R}^n) : \|u\|^2_{H^s(\mathbb{R}^n)} := \int_{\mathbb{R}^n} (1+|k|^s)^2 |\widehat{u}(k)|^2 dk < \infty\}.$$

The Fourier representation and the theorem of Riemann-Lebesgue allow a simple proof of the following Sobolev embedding result.

Theorem A.14. *Let $s > n/2$. Then there is a constant $c > 0$ such that for each $f \in H^s(\mathbb{R}^n)$ we have*

$$f \in C(\mathbb{R}^n), \quad \|f\|_{C_b^0(\mathbb{R}^n)} \leqslant c\|f\|_{H^s(\mathbb{R}^n)}, \quad f(x) \to 0 \quad \textit{as } |x| \to \infty.$$

Proof. By the Cauchy-Schwarz inequality

$$\begin{aligned}\int_{\mathbb{R}^n} |\widehat{f}(k)| dk &= \int_{\mathbb{R}^n} (1+|k|^2)^{s/2} |\widehat{f}(k)| (1+|k|^2)^{-s/2} dk \leqslant \left(\int_{\mathbb{R}^n} (1+|k|^2)^s |\widehat{f}(k)|^2 dk\right)^{1/2} \left(\int_{\mathbb{R}^n} (1+|k|^2)^{-s} dk\right)^{1/2} \\ &\leqslant K\|f\|_{H^s(\mathbb{R}^n)}\end{aligned}$$

for some $K \in (0,\infty)$, where in the last step we used the fact that $\int_{\mathbb{R}^n} (1+|k|^2)^{-s} dk < \infty$ for $s > n/2$.

Hence $\widehat{f} \in L^1(\mathbb{R}^n)$ and by Theorem A.8 we have the continuity and decay of f. The estimate of the supremum norm follows from the above calculation and the definition of the inverse Fourier transform $f(x) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} \widehat{f}(k) e^{\mathrm{i}k\cdot x} dk$:

$$\|f\|_{C_b^0(\mathbb{R}^n)} \leqslant (2\pi)^{-n/2} \int_{\mathbb{R}^n} |\widehat{f}(k)| dk \leqslant K(2\pi)^{-n/2} \|f\|_{H^s(\mathbb{R}^n)}.$$

$\square$

Appendix B

Asymptotics

B.1 Asymptotic Notation

Let $f, g \colon X \to Y$, where $X, Y \in \{\mathbb{R}, \mathbb{C}\}$.

Definition B.1. *f is big-O of g as $x \to x_0$ if*

i) $\exists$ *neighborhood U of x_0 in X*

ii) $\exists\ M > 0$: $x \in U \implies |f(x)| \leqslant M \cdot |g(x)|$

Notation: $f(x) = O(g(x))$ as $x \to x_0$.

Remark B.1. *When $g(x) \neq 0$ in a punctured neighborhood $U(x_0) \backslash \{x_0\}$, then $f(x) = O(g(x))$ as $x \to x_0$ if and only if* $\limsup\limits_{x \to x_0} \frac{|f(x)|}{|g(x)|} < \infty$.

Definition B.2. *(little o) f is little-o of g as $x \to x_0$ if* $\forall \varepsilon > 0\ \exists U_\varepsilon(x_0) \subset X : x \in U_\varepsilon \implies |f(x)| \leqslant \varepsilon \cdot |g(x)|$.

Notation: $f(x) = o(g(x))$ as $x \to x_0$.

Remark B.2. *When $g(x) \neq 0$ in a punctured neighborhood $U(x_0) \backslash \{x_0\}$, then $f(x) = o(g(x))$ as $x \to x_0$ if and only if* $\lim\limits_{x \to x_0} \frac{|f(x)|}{|g(x)|} = 0$.

The meaning of $f(x) = o(g(x))$ is that (in the given limit) f is asymptotically negligible with respect to g.

Definition B.3. *f is asymptotically equivalent to g as $x \to x_0$ if $f - g = o(g)$ as $x \to x_0$.*

Notation: $f \sim g$ as $x \to x_0$.

Remark B.3. *If $g(x) \neq 0$ on a punctured neighborhood $U(x_0) \backslash \{x_0\}$, then $f \sim g$ as $x \to x_0 \iff \lim\limits_{x \to x_0} \frac{f(x)}{g(x)} = 1$.*

Remark B.4. *$\sim$ is an equivalence relation since*

- $f \sim f$
- $f \sim g \implies g \sim f$
- $f \sim g,\ g \sim h \implies f \sim h$.

B.2 Gamma Function

The Gamma function is defined via

Definition B.4. *For* $z \in \mathbb{C}$ *with* $-\operatorname{Re}(z) \notin \mathbb{N} := \{0, 1, 2, \dots\}$ *we define*

$$\Gamma(z) := \begin{cases} \Gamma_+(z) = \int\limits_0^\infty e^{-t} t^{z-1}\,\mathrm{d}t, & \operatorname{Re}(z) > 0 \\ \frac{\Gamma_+(z+n)}{z(z+1)\cdot\ldots\cdot(z+n-1)}, & -n < \operatorname{Re}(z) \leqslant -n+1, \;\; z \neq -n+1, \;\; n \in \mathbb{N} \end{cases}$$

Some basic properties of the Gamma function are

$$\begin{aligned} \Gamma(n) &= (n-1)! && \text{for all } n \in \mathbb{N}, \\ \Gamma(z+1) &= z\Gamma(z) && \text{for all } z \in \mathbb{C} \text{ with } -\operatorname{Re}(z) \notin \mathbb{N}, \\ \Gamma(\tfrac{1}{2}) &= \sqrt{\pi}. \end{aligned}$$

Example B.1. Let $\alpha \in (0,1)$, $x > 0$. Then

$$\int\limits_0^\infty e^{\mathrm{i}xt} t^{\alpha-1}\,\mathrm{d}t = \frac{\mathrm{i}^\alpha \Gamma(\alpha)}{x^\alpha}.$$

Proof. Define $f : \mathbb{C} \to \mathbb{C}$ via $f(z) := e^{\mathrm{i}xz} z^{\alpha-1}$. f is discontinuous across the branch cut on the negative real axis. We choose a contour as in Fig. B.2. Inside the contour f is holomorphic, so that Cauchy Integral

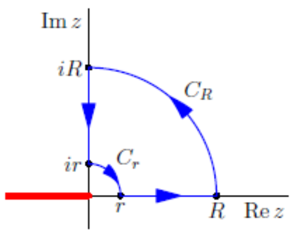

Figure B.1: from [24]

Theorem yields

$$0 = \int\limits_r^R f(z)\,\mathrm{d}z + \int\limits_{C_R} f(z)\,\mathrm{d}z + \int\limits_{\mathrm{i}R}^{\mathrm{i}r} f(z)\,\mathrm{d}z + \int\limits_{C_r} f(z)\,\mathrm{d}z.$$

We have

$$\begin{aligned} \left| \int\limits_{C_R} f(z)\,\mathrm{d}z \right| &\leqslant \int\limits_0^{\pi/2} |e^{-xR\sin\theta + \mathrm{i}R\cos\theta} R^{\alpha-1} R|\,\mathrm{d}\theta = R^\alpha \int\limits_0^{\pi/2} e^{-xR\sin\theta}\,\mathrm{d}\theta \qquad \leqslant R^\alpha \int\limits_0^{\pi/2} e^{-xR\frac{2\theta}{\pi}}\,\mathrm{d}\theta \\ &= R^{\alpha-1}\frac{\pi}{2x}(1 - e^{-xR}) \xrightarrow{R\to\infty} 0 \quad \text{since } \alpha < 1. \end{aligned}$$

For the curve C_r we argue similarly and get

$$\left| \int\limits_{C_r} f(z)\,\mathrm{d}z \right| \leqslant r^\alpha \frac{\pi}{2x}\left(\frac{1 - e^{-xr}}{r}\right) \xrightarrow{r\to\infty} 0 \quad \text{since } \alpha > 0.$$

Hence for $r \to 0$ and $R \to \infty$

$$\int_0^\infty f(z)\,\mathrm{d}z = -\int_{\mathrm{i}R}^{\mathrm{i}r} f(z)\,\mathrm{d}z = -\mathrm{i}\int_\infty^0 f(\mathrm{i}t)\,\mathrm{d}t = \mathrm{i}^\alpha \int_0^\infty e^{-xt}t^{\alpha-1}\,\mathrm{d}t = \frac{\mathrm{i}^\alpha\Gamma(\alpha)}{x^\alpha}.$$

□

B.3 Method of Stationary Phase

Consider

$$I(x) = \int_a^b e^{\mathrm{i}xh(t)} f(t)\,\mathrm{d}t,$$

where $a, b \in \mathbb{R}$, $x \in \mathbb{R}$, and $h\colon \mathbb{R} \to \mathbb{R}$.

The aim is to determine the asymptotics of $I(x)$ as $x \to \infty$. The main idea is that for large values of x the function $e^{\mathrm{i}xh(t)}$ is highly oscillatory in t so that cancellation occurs in the integral. The main contribution to the integral thus comes from the neighborhood of points t, where $h'(t) = 0$, i.e. where the phase is stationary.

Example B.2. A common example is the Fourier transform integral $\int_{\mathbb{R}} e^{\mathrm{i}kx}\hat{g}(k)\,\mathrm{d}t = g(x)$ for large values of x.

Lemma B.5. *Let $a, b \in \mathbb{R}, a < b$ and $f \in C([a,b]), h \in C^1([a,b]), \frac{f}{h'} \in C^1([a,b])$ with $h'(t) \neq 0$ on $[a,b]$. Then*

$$\int_a^b e^{\mathrm{i}xh(t)} f(t)\,\mathrm{d}t \sim \frac{\mathrm{i}}{x}\left[\frac{e^{\mathrm{i}xh(a)}f(a)}{h'(a)} - \frac{e^{\mathrm{i}xh(b)}f(b)}{h'(b)}\right] \quad (x \to \infty).$$

Let $a \in \mathbb{R}$. If $f \in C([a,\infty)), h \in C^1([a,\infty))$ with $h'(t) \neq 0$ on $[a,\infty)$, $\left(\frac{f}{h'}\right)' \in L^1(a,\infty)$ and $\frac{f(t)}{h'(t)} \to 0$ as $t \to \infty$, then

$$\int_a^\infty e^{\mathrm{i}xh(t)} f(t)\,\mathrm{d}t \sim \frac{\mathrm{i}}{x}\frac{e^{\mathrm{i}xh(a)}f(a)}{h'(a)} \quad (x \to \infty).$$

Proof. For the statement on the compact interval use the substitution $t = h^{-1}(s)$ and partial integration to get

$$\begin{aligned}\int_a^b f(t)e^{\mathrm{i}xh(t)}\,\mathrm{d}t &= \int_{h(a)}^{h(b)} e^{\mathrm{i}xs}\frac{f(h^{-1}(s))}{h'(h^{-1}(s))}\,\mathrm{d}s \\ &= -\frac{\mathrm{i}}{x}\left[e^{\mathrm{i}xs}\frac{f(h^{-1}(s))}{h'(h^{-1}(s))}\right]_{h(a)}^{h(b)} + \frac{\mathrm{i}}{x}\int_{h(a)}^{h(b)} e^{\mathrm{i}xs}\frac{d}{ds}\left(\frac{f(h^{-1}(s))}{h'(h^{-1}(s))}\right)\mathrm{d}s.\end{aligned}$$

Since $\frac{f}{h'} \in C^1([a,b])$, the integral on the right hand side converges to zero for $x \to \infty$ by the Riemann-Lebesgue lemma A.8. Hence the result follows.

For the unbounded interval the same calculation applies. We note that $\left(\frac{f}{h'}\right)' \in L^1(a,\infty)$ is equivalent to $\frac{d}{ds}\left(\frac{f(h^{-1}(s))}{h'(h^{-1}(s))}\right) \in L^1(h(a), \lim_{b\to\infty} h(b))$. Thus the last integral converges to zero due the Riemann-Lebesgue lemma again. With $\frac{f(t)}{h'(t)} \to 0$ as $t \to \infty$ we get the result. □

Lemma B.6. *If $f \sim g$ and g is bounded as $x \to x_0$, then*

$$e^f \sim e^g \quad \textit{as } x \to x_0.$$

Proof. Let $\varepsilon > 0$. Then $\exists\, \delta > 0$ so that

$$|x - x_0| < \delta \implies |f(x) - g(x)| < \frac{\varepsilon}{M}|g(x)| < \varepsilon,$$

where $M := \max_{x \in [x_0 - \delta, x_0 + \delta]} |g(x)|$. This implies $\lim_{x \to x_0} (f(x) - g(x)) = 0$. The continuity of $\exp(\cdot)$ then yields

$$1 = e^0 = \lim_{x \to x_0} (e^{f(x) - g(x)}) = \lim_{x \to x_0} \frac{e^{f(x)}}{e^{g(x)}}.$$

□

Note that it is not generally true that if $f(x) \sim g(x)$ as $x \to x_0$ and if H is a continuous function, then $H(f(x)) \sim H(g(x))$ as $x \to x_0$. A counterexample is $f(x) = x, g(x) = \pi, H(y) = \sin(y)$ and $x_0 = \pi$ because, clearly, $|\sin(x) - \sin(\pi)| > \varepsilon \sin(\pi) = 0$ for all x in a punctured neighborhood of π and all $\varepsilon > 0$.

Theorem B.7. *Let $a \in \mathbb{R}, b \in \mathbb{R} \cup \{\infty\}$, let $\operatorname{supp} f = S \subset \mathbb{R}$ be compact and $f \in C^1(S)$, $h \in C^N(S)$. If*

- $f(a) \neq 0$,
- $h'(t) \neq 0 \ \ \forall t \in S \backslash \{a\}$,
- $h'(a) = \cdots = h^{(N-1)}(a) = 0$, $h^{(N)}(a) \neq 0$ *for some* $N \geqslant 2$,

then

$$\begin{aligned} I(x) := \int_a^b e^{\mathrm{i}xh(t)} f(t)\,\mathrm{d}t &= \frac{1}{N}\Gamma\left(\frac{1}{N}\right) f(a) e^{\mathrm{i}xh(a)} \left(\frac{\mathrm{i}N!}{h^{(N)}(a)x}\right)^{\frac{1}{N}} + o(x^{-1/N}) \quad (x \to \infty) \\ &= \frac{1}{N}\Gamma\left(\frac{1}{N}\right) f(a) e^{\mathrm{i}xh(a)} e^{\mathrm{i}\frac{\pi}{2N}\operatorname{sign}(h^{(N)}(a))} \left(\frac{N!}{|h^{(N)}(a)|x}\right)^{\frac{1}{N}} + o(x^{-1/N}) \quad (x \to \infty). \end{aligned}$$

Proof. Using Lemma B.6, we have

$$e^{\mathrm{i}x(h(t) - h(a))} \frac{f(t)}{f(a)} \sim e^{\mathrm{i}xh^{(N)}(a)\frac{(t-a)^N}{N!}} =: F(x,t) \quad (t \to a^+).$$

That means

$$e^{\mathrm{i}x(h(t) - h(a))} f(t) = f(a)\,(F(x,t) + R(x,t)) \ \text{ with } \ R(x,t) = o(F(x,t)) \quad (t \to a^+).$$

Choosing $x > 0$, there exists $\delta > 0$ small enough so that if $a \leqslant t < a + \delta$, then

$$|R(x,t)| \leqslant \frac{1}{x}|F(x,t)|. \tag{B.1}$$

We split $I(x)$ into

$$I(x) = \underbrace{\int_a^{a+\delta} e^{\mathrm{i}xh(t)} f(t)\,\mathrm{d}t}_{I_1(x)} + \underbrace{\int_{a+\delta}^{b} e^{\mathrm{i}xh(t)} f(t)\,\mathrm{d}t}_{I_2(x)}.$$

For I_2 we use Lemma B.5 to conclude $I_2(x) = O\left(\frac{1}{x}\right)$ since $|h'(t)| > \alpha^2 > 0 \quad \forall t \in [a+\delta, b) \cap S$. For I_1 we get

$$I_1(x) = e^{\mathrm{i}xh(a)} f(a) \int_a^{a+\delta} F(x,t) + R(x,t)\,\mathrm{d}t \overset{(B.1)}{=} e^{\mathrm{i}xh(a)} f(a) \int_a^{a+\delta} F(x,t)\,\mathrm{d}t + O\left(\frac{1}{x}\right).$$

We write next $\int_a^{a+\delta} F(x,t)\,\mathrm{d}t = \int_a^{\infty} F(x,t)\,\mathrm{d}t - \int_{a+\delta}^{\infty} F(x,t)\,\mathrm{d}t$. By another application of Lemma B.5 the latter integral satisfies

$$\int_{a+\delta}^{\infty} F(x,t)\,\mathrm{d}t = O\left(\frac{1}{x}\right)$$

as $(t-a)^N$ has no extrema in $[a+\delta, \infty)$. It can be checked that all assumptions of Lemma B.5 (mainly the L^1 condition) are satisfied in the application to $\int_{a+\delta}^{\infty} F(x,t)\,\mathrm{d}t$.

Note that the existence of $\int_{a+\delta}^{\infty} F(x,t)\,\mathrm{d}t$ follows from the existence of $\int_a^{\infty} F(x,t)\,\mathrm{d}t$ shown next.

For the former integral, with the substitution $u = |h^{(N)}(a)| \frac{(t-a)^N}{N!}$, we have

$$\begin{aligned}\int_a^{\infty} F(x,t)\,\mathrm{d}t &= \left(\int_0^{\infty} e^{\mathrm{i}xu\,\mathrm{sign}(h^{(N)}(a))} u^{\left(\frac{1}{N}-1\right)}\,\mathrm{d}u\right) \frac{1}{N} \left(\frac{N!}{|h^{(N)}(a)|}\right)^{\frac{1}{N}} \\ &= \frac{1}{N} \left(\frac{N!}{|h^{(N)}(a)|}\right)^{\frac{1}{N}} \frac{\left(\mathrm{i}\,\mathrm{sign}\left(h^{(N)}(a)\right)\right)^{\frac{1}{N}} \Gamma\left(\frac{1}{N}\right)}{x^{\frac{1}{N}}},\end{aligned}$$

where the identity $\int_0^{\infty} e^{\mathrm{i}xt} t^{\alpha-1}\,\mathrm{d}t = \mathrm{i}^{\alpha} \frac{\Gamma(\alpha)}{x^{\alpha}}$ has been used. □

Remark B.5. *Similarly for an extremum of order N in the right end point $t = b \in \mathbb{R}$ with analogous assumptions as above we get*

$$I(x) = \frac{1}{N}\Gamma\left(\frac{1}{N}\right) f(b) e^{\mathrm{i}xh(b)} e^{(-1)^N \mathrm{i}\frac{\pi}{2N}\,\mathrm{sign}(h^{(N)}(b))} \left(\frac{N!}{|h^{(N)}(b)|x}\right)^{\frac{1}{N}} + o(x^{-1/N}) \quad (x \to \infty).$$

For an internal extremum in $c \in (a,b)$ one needs to add the contributions from $(a,c]$ and $[c,b)$. The resulting asymptotics for $x \to \infty$ are

$$\begin{aligned}I(x) \sim& \frac{2}{N}\Gamma\left(\frac{1}{N}\right) f(c) e^{\mathrm{i}xh(c)} \cos\left(\frac{\pi}{2N}\right) \left(\frac{N!}{|h^{(N)}(c)|x}\right)^{\frac{1}{N}} \quad \text{for } N \text{ odd,} \\ \sim& \frac{2}{N}\Gamma\left(\frac{1}{N}\right) f(c) e^{\mathrm{i}xh(c)} e^{\mathrm{i}\frac{\pi}{2N}\,\mathrm{sign}(h^{(N)}(c))} \left(\frac{N!}{|h^{(N)}(c)|x}\right)^{\frac{1}{N}} \quad \text{for } N \text{ even.}\end{aligned}$$

Remark B.6. *Also note that the condition of compact support of f is not necessary. The proof of Theorem B.7 can be easily adapted for the case $f \in C^1([a,\infty)) \cap L^1((a,\infty)), h \in C^N([a,\infty))$ with $\left(\frac{f}{h'}\right)' \in L^1(a+\delta,\infty)$ for each $\delta > 0$ and $\frac{f(t)}{h'(t)} \to 0$ as $t \to \infty$. These conditions are satisfied, e.g., if $f \in S(R)$, h grows at most exponentially and if for each $\delta > 0$ there is $\alpha > 0$ s.t. $|h'(t)| > \alpha > 0$ for all $t \in [a+\delta, \infty)$.*

There is also a multidimensional version of the stationary phase method. A classical result [15] is

Theorem B.8. *Let $f \in C^1(\mathbb{R}^n, \mathbb{R})$ have a compact support $S \subset \mathbb{R}^n$ and $y_0 \in S$. Let also $h \in C^2(S, \mathbb{R})$. If*

- $f(y_0) \neq 0$,
- $\nabla h(y_0) = 0, \ \nabla h(x) \neq 0 \ \ \forall x \in S \backslash \{y_0\}$,
- $\det(D^2 h(y_0)) \neq 0$,

then

$$\int\limits_{\mathbb{R}^n} e^{\mathrm{i}\mu h(x)} f(x) \, \mathrm{d}x = \left(\frac{2\pi}{\mu}\right)^{n/2} \frac{f(y_0)}{|\det(D^2 h(y_0))|^{\frac{1}{2}}} e^{\mathrm{i}(\mu h(y_0) + \sigma\pi/4)} + o(\mu^{-n/2}) \quad (\mu \to \infty),$$

where $D^2 h$ is the Hessian matrix of h and where σ is the signature of $D^2 h(y_0)$, i.e. the number of positive eigenvalues minus the number of negative eigenvalues.

Bibliography

[1] M. Ablowitz. Nonlinear Dispersive Waves: Asymptotic Analysis and Solitons. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2011.

[2] M. Ablowitz and H. Segur. Solitons and the Inverse Scattering Transform. SIAM Studies in Applied Mathematics. Society for Industrial and Applied Mathematics, 2006.

[3] R. A. Adams and J. J. F. Fournier. Sobolev spaces. Elsevier/Academic Press, Amsterdam, second edition, 2003.

[4] T. B. Benjamin. The stability of solitary waves. Proc. Roy. Soc. Lond. A, 328(1573):153–183, 1972.

[5] X. Carvajal. A note on local smoothing effects for the unitary group associated with the KdV equation. Electronic Journal of Differential Equations, 2008:1 − 7, 2008.

[6] P. Drazin and R. Johnson. Solitons: An Introduction. Cambridge Computer Science Texts. Cambridge University Press, 1989.

[7] J. Dufour and N. Zung. Poisson Structures and Their Normal Forms. Progress in Mathematics. Birkhäuser Basel, 2006.

[8] N. Dunford and J. Schwartz. Linear Operators: General theory. Pure and applied mathematics. Interscience Publishers, 1958.

[9] L. Evans. Partial Differential Equations. Graduate studies in mathematics. American Mathematical Society, 1999.

[10] E. Falcon, C. Laroche, and S. Fauve. Observation of depression solitary surface waves on a thin fluid layer. Phys. Rev. Lett., 89:204501, Oct 2002.

[11] E. Fermi, J. Pasta, M. Tsingou, and S. M. Ulam. Studies of nonlinear problems. I. Technical Report LA-1940, May 1955. Also in Enrico Fermi: Collected Papers, volume 2, edited by E. Amaldi, H. L. Anderson, E. Persico, E. Segre, and A. Wattenberg. Chicago: University of Chicago Press, 1965, pages 978–988.

[12] A. Hermann. The validity of the Nonlinear Schrödinger approximation in higher space dimensions. PhD thesis, Universität Stuttgart, Holzgartenstr. 16, 70174 Stuttgart, 2014.

[13] D. Lannes. The Water Waves Problem: Mathematical Analysis and Asymptotics. Mathematical Surveys and Monographs. Amer Mathematical Society, 2013.

[14] F. Linares and G. Ponce. Introduction to Nonlinear Dispersive Equations. Universitext - Springer-Verlag. Springer, 2009.

[15] A. Majda. Introduction to PDEs and Waves for the Atmosphere and Ocean. Courant lecture notes in mathematics. Courant Institute of Mathematical Sciences, 2003.

[16] A. Majda and A. Bertozzi. Vorticity and Incompressible Flow. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2002.

[17] J. Moloney and A. Newell. Nonlinear Optics. Advanced Book Program. Westview Press, 2004.

[18] M. Reed and B. Simon. Methods of modern mathematical physics. II. Fourier analysis, self adjointness. Academic Press, New York, 1975.

[19] W. Rudin. Functional analysis. International series in pure and applied mathematics. McGraw-Hill, 1991.

[20] J. S. Russell. Report on Waves. In Report of the fourteenth meeting of the British Association for the Advancement of Science, pages 311–390, Plates XLVII–LVII, York 1844 (London, 1845).

[21] G. Schneider. Nonlinear stability of taylor vortices in infinite cylinders. Archive for Rational Mechanics and Analysis, 144(2):121–200, 1998.

[22] K. Socha. Circles in circles: Creating a mathematical model of surface water waves. The American Mathematical Monthly, 114(3):202–216, 2007.

[23] T. Tao. Nonlinear Dispersive Equations: Local and Global Analysis. Number Nr. 106 in Conference Board of the Mathematical Sciences. Regional conference series in mathematics. American Mathematical Soc., 2006.

[24] H. van Roessel and J. Bowman. Asymptotic methods, lecture notes, 2012. http://www.math.ualberta.ca/~bowman/m538/m538.pdf.

[25] W. Walter. Gewöhnliche Differentialgleichungen: Eine Einführung. Springer-Lehrbuch. Springer Berlin Heidelberg, 2000. Sec. 30.

[26] G. Whitham. Linear and nonlinear waves. Pure and applied mathematics. Wiley, 1974.