CHAPTER 1

Ordinary differential equations

Science is a differential equation. Religion is a boundary condition.
(Alan Turing, quoted in J.D. Barrow, "Theories of everything")

This monograph is primarily concerned with the global *Cauchy problem* (or *initial value problem*) for partial differential equations (PDE), but in order to assemble some intuition on the behaviour of such equations, and on the power and limitations of the various techniques available to analyze these equations, we shall first study these phenomena and methods in the much simpler context of *ordinary differential equations* (ODE), in which many of the technicalities in the PDE analysis are not present. Conversely, the theory of ODEs, particularly Hamiltonian ODEs, has a very rich and well-developed structure, the extension of which to nonlinear dispersive PDEs is still far from complete. For instance, phenomena from Hamiltonian dynamics such as Kolmogorov-Arnold-Moser (KAM) invariant tori, symplectic non-squeezing, Gibbs and other invariant measures, or Arnold diffusion are well established in the ODE setting, but the rigorous theory of such phenomena for PDEs is still its infancy.

One technical advantage of ODE, as compared with PDE, is that with ODE one can often work entirely in the category of classical (i.e. smooth) solutions, thus bypassing the need for the theory of distributions, weak limits, and so forth. However, even with ODE it is possible to exhibit blowup in finite time, and in high-dimensional ODE (which begin to approximate PDE in the infinite dimensional limit) it is possible to have the solution stay bounded in one norm but become extremely large in another norm. Indeed, the quantitative study of expressions such as mass, energy, momentum, etc. is almost as rich in the ODE world as it is in the PDE world, and thus the ODE model does serve to illuminate many of the phenomena that we wish to study for PDE.

A common theme in both nonlinear ODE and nonlinear PDE is that of *feedback* - the solution to the equation at any given time generates some forcing term, which in turn feeds back into the system to influence the solution at later times, usually in a nonlinear fashion. The tools we will develop here to maintain control of this feedback effect - the Picard iteration method, Gronwall's inequality, the bootstrap principle, conservation laws, monotonicity formulae, and Duhamel's formula - will form the fundamental tools we will need to analyze nonlinear PDE in later chapters. Indeed, the need to deal with such feedback gives rise to a certain "nonlinear way of thinking", in which one continually tries to control the solution in terms of itself, or derive properties of the solution from (slightly weaker versions of) themselves. This way of thinking may initially seem rather unintuitive, even circular, in nature, but it can be made rigorous, and is absolutely essential to proceed in this theory.

1.1. General theory

> *It is a capital mistake to theorise before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.* (Sir Arthur Conan Doyle, "A Study in Scarlet")

In this section we introduce the concept of an ordinary differential equation and the associated Cauchy problem, but then quickly specialise to an important subclass of such problems, namely the Cauchy problem (1.7) for autonomous first-order quasilinear systems.

Throughout this chapter, $\mathcal{D}$ will denote a (real or complex) finite dimensional vector space, which at times we will endow with some norm $\|\|_{\mathcal{D}}$; the letter $\mathcal{D}$ stands for "data". An *ordinary differential equation* (ODE) is an equation which governs certain functions $u : I \to \mathcal{D}$ mapping a (possibly infinite) time interval $I \subseteq \mathbf{R}$ to the vector space[1] $\mathcal{D}$. In this setup, the most general form of an ODE is that of a *fully nonlinear ODE*

$$G(u(t), \partial_t u(t), \ldots, \partial_t^k u(t), t) = 0 \tag{1.1}$$

where $k \geq 1$ is an integer, and $G : \mathcal{D}^{k+1} \times I \to X$ is a given function taking values in another finite-dimensional vector space X. We say that a function $u \in C^k_{\mathrm{loc}}(I \to \mathcal{D})$ is a *classical solution*[2] (or *solution* for short) of the ODE if (1.1) holds for all $t \in I$. The integer k is called the *order* of the ODE, thus for instance if $k = 2$ then we have a second-order ODE. One can think of $u(t)$ as describing the state of some physical system at a given time t; the dimension of $\mathcal{D}$ then measures the degrees of freedom available. We shall refer to $\mathcal{D}$ as the *state space*, and sometimes refer to the ODE as the *equation(s) of motion*, where the plural reflects the fact that X may have more than one dimension. While we will occasionally consider the *scalar case*, when $\mathcal{D}$ is just the real line $\mathbf{R}$ or the complex plane $\mathbf{C}$, we will usually be more interested in the case when the dimension of $\mathcal{D}$ is large. Indeed one can view PDE as a limiting case of ODE as $\dim(\mathcal{D}) \to \infty$.

In this monograph we will primarily consider those ODE which are *time-translation-invariant* (or *autonomous*), in the sense that the function G does not actually depend explicitly on the time parameter t, thus simplifying (1.1) to

$$\tilde{G}(u(t), \partial_t u(t), \ldots, \partial_t^k u(t)) = 0 \tag{1.2}$$

for some function $\tilde{G} : \mathcal{D}^{k+1} \to X$. One can in fact convert any ODE into a time-translation-invariant ODE, by the trick of embedding the time variable itself into the state space, thus replacing $\mathcal{D}$ with $\mathcal{D} \times \mathbf{R}$, X with $X \times \mathbf{R}$, u with the function

[1] One could generalise the concept of ODE further, by allowing $\mathcal{D}$ to be a smooth manifold instead of a vector space, or even a smooth bundle over the time interval I. This leads for instance to the theory of *jet bundles*, which we will not pursue here. In practice, one can descend from this more general setup back to the original framework of finite-dimensional vector spaces - locally in time, at least - by choosing appropriate local coordinate charts, though often the choice of such charts is somewhat artifical and makes the equations messier; see Chapter 6 for some related issues.

[2] We will discuss non-classical solutions shortly. As it turns out, for finite-dimensional ODE there is essentially no distinction between a classical and non-classical solution, but for PDE there will be a need to distinguish between classical, strong, and weak solutions. See Section 3.2 for further discussion.

$\tilde{u}(t) := (u(t), t)$, and G with the function[3]

$$\tilde{G}((u_0, s_0), (u_1, s_1), \ldots, (u_k, s_k)) := (G(u_0, \ldots, u_k), s_1 - 1).$$

For instance, solutions to the non-autonomous ODE

$$\partial_t u(t) = F(t, u(t))$$

are equivalent to solutions to the system of autonomous ODE

$$\partial_t u(t) = F(s(t), u(t)); \quad \partial_t s(t) - 1 = 0$$

provided that we also impose a new initial condition $s(0) = 0$. This trick is not always without cost; for instance, it will convert a non-autonomous linear equation into an autonomous nonlinear equation.

By working with time translation invariant equations we obtain our first symmetry, namely the *time translation symmetry*

$$u(t) \mapsto u(t - t_0). \tag{1.3}$$

More precisely, if $u : I \to \mathcal{D}$ solves the equation (1.2), and $t_0 \in \mathbf{R}$ is any time shift parameter, then the time-translated function $u_{t_0} : I + t_0 \to \mathcal{D}$ defined by $u_{t_0}(t) := u(t - t_0)$, where $I + t_0 := \{t + t_0 : t \in I\}$ is the time translation of I, is also a solution to (1.2). This symmetry tells us, for instance, that the initial value problem for this equation starting from time $t = 0$ will be identical (after applying the symmetry (1.3)) to the initial value problem starting from any other time $t = t_0$.

The equation (1.2) implicitly determines the value of the top-order derivative $\partial_t^k u(t)$ in terms of the lower order derivatives $u(t), \partial_t u(t), \ldots, \partial_t^{k-1} u(t)$. If the hypotheses of the implicit function theorem[4] are satisfied, then we can solve for $\partial_t^k u(t)$ uniquely, and rewrite the ODE as an *autonomous quasilinear ODE of order k*

$$\partial_t^k u(t) = F(u(t), \partial_t u(t), \ldots, \partial_t^{k-1} u(t)), \tag{1.4}$$

for some function $F : \mathcal{D}^k \to \mathcal{D}$. Of course, there are times when the implicit function theorem is not available, for instance if the domain $\mathcal{Y}$ of $\tilde{G}$ has a different dimension than that of $\mathcal{D}$. If $\mathcal{Y}$ has larger dimension than $\mathcal{D}$ then the equation is often *over-determined*; it has more equations of motion than degrees of freedom, and one may require some additional hypotheses on the initial data before a solution is guaranteed. If $\mathcal{Y}$ has smaller dimension than $\mathcal{D}$ then the equation is often *under-determined*; it has too few equations of motion, and one now expects to have a multiplicity of solutions for any given initial datum. And even if $\mathcal{D}$ and $\mathcal{Y}$ have the same dimension, it is possible for the ODE to sometimes be *degenerate*, in that the Jacobian that one needs to invert for the implicit function theorem becomes singular.

---

[3] Informally, what one has done is added a "clock" s to the system, which evolves at the fixed rate of one time unit per time unit ($\frac{ds}{dt} - 1 = 0$), and then the remaining components of the system are now driven by clock time rather than by the system time. The astute reader will note that this new ODE not only contains all the solutions to the old ODE, but also contains some additional solutions; however these new solutions are simply time translations of the solutions coming from the original ODE.

[4] An alternate approach is to differentiate (1.2) in time using the chain rule, obtaining an equation which is linear in $\partial_t^{k+1} u(t)$, and provided that a certain matrix is invertible, one can rewrite this in the form (1.4) but with k replaced by $k + 1$.

Degenerate ODE are rather difficult to study and will not be addressed here. Both under-determined and over-determined equations cause difficulties for analysis, which are resolved in different ways. An over-determined equation can often be made determined by "forgetting" some of the constraints present in (1.2), for instance by projecting Y down to a lower-dimensional space. In many cases, one can then recover the forgotten constraints by using some additional hypothesis on the initial datum, together with an additional argument (typically involving Gronwall's inequality); see for instance Exercises 1.13, (6.4). Meanwhile, an under-determined equation often enjoys a large group of "gauge symmetries" which help "explain" the multiplicity of solutions to the equation; in such a case one can often fix a special gauge, thus adding additional equations to the system, to make the equation determined again; see for instance Section 6.2 below. In some cases, an ODE can contain both over-determined and under-determined components, requiring one to perform both of these types of tricks in order to recover a determined equation, such as one of the form (1.4).

Suppose that u is a classical solution to the quasilinear ODE (1.4), and that the nonlinearity $F : \mathcal{D}^k \to \mathcal{D}$ is smooth. Then one can differentiate (1.4) in time, use the chain rule, and then substitute in (1.4) again, obtain an equation of the form

$$\partial_t^{k+1} u(t) = F_{k+1}(u(t), \partial_t u(t), \ldots, \partial_t^{k-1} u(t))$$

for some smooth function $F_{k+1} : \mathcal{D}^k \to \mathcal{D}$ which can be written explicitly in terms of G. More generally, by an easy induction we obtain equations of the form

$$\partial_t^{k'} u(t) = F_{k'}(u(t), \partial_t u(t), \ldots, \partial_t^{k-1} u(t)) \tag{1.5}$$

for any $k' \geq k$, where $F_{k'} : \mathcal{D}^k \to \mathcal{D}$ is a smooth function which depends only on G and k'. Thus, if one specifies the initial data $u(t_0), \ldots, \partial_t^{k-1} u(t_0)$ at some fixed time t_0, then all higher derivatives of u at t_0 are also completely specified. This shows in particular that if u is $k-1$-times continuously differentiable and F is smooth, then u is automatically smooth. If u is not only smooth but analytic, then from Taylor expansion we see that u is now fixed uniquely. Of course, it is only reasonable to expect u to be analytic if F is also analytic. In such a case, we can complement the above uniqueness statement with a (local) existence result:

THEOREM 1.1 (Cauchy-Kowalevski theorem). *Let $k \geq 1$. Suppose $F : \mathcal{D}^k \to \mathcal{D}$ is real analytic, let $t_0 \in \mathbf{R}$, and let $u_0, \ldots, u_{k-1} \in \mathcal{D}$ be arbitrary. Then there exists an open time interval I containing t_0, and a unique real analytic solution $u : I \to \mathcal{D}$ to* (1.4)*, which obeys the initial value conditions*

$$u(t_0) = u_0; \quad \partial_t u(t_0) = u_1, \ldots, \partial_t^{k-1} u(t_0) = u_{k-1}.$$

We defer the proof of this theorem to Exercise 1.1. This beautiful theorem can be considered as a complete local existence theorem for the ODE (1.4), in the case when G is real analytic; it says that the initial position $u(t_0)$, and the first $k-1$ derivatives, $\partial_t u(t_0), \ldots, \partial_t^{k-1} u(t_0)$, are precisely the right amount of *initial data*[5] needed in order to have a wellposed initial value problem (we will define wellposedness more precisely later). However, it turns out to have somewhat

[5]Conventions differ on when to use the singular "datum" and the plural "data". In this text, we shall use the singular "datum" for ODE and PDE that are first-order in time, and the plural "data" for ODE and PDE that are higher order (or unspecified order) in time. Of course, in both cases we use the plural when considering an ensemble or class of data.

limited application when we move from ODE to PDE (though see Exercise 3.25). We will thus rely instead primarily on a variant of the Cauchy-Kowalevski theorem, namely the *Picard existence theorem*, which we shall discuss below.

REMARK 1.2. The fact that the solution u is restricted to lie in a open interval I, as opposed to the entire real line $\mathbf{R}$, is necessary. A basic example is the initial value problem

$$u_t = u^2; \quad u(0) = 1 \tag{1.6}$$

where u takes values on the real line $\mathbf{R}$. One can easily verify that the function $u(t) := \frac{1}{1-t}$ solves this ODE with the given initial datum as long as $t < 1$, and thus is the unique real-analytic solution to this ODE in this region. But this solution clearly blows up (i.e. ceases to be smooth) at $t = 1$, and so cannot be continued[6] real analytically beyond this point.

There is a simple trick available to reduce a k^{th} order ODE such as (1.4) to a first order ODE, at the cost of multiplying the number of degrees of freedom by k, or more precisely, replacing the state space $\mathcal{D}$ by the *phase space* $\mathcal{D}^k$. Indeed, if one defines the new function $\tilde{u} : I \to \mathcal{D}^k$ by

$$\tilde{u}(t) := (u(t), \partial_t u(t), \ldots, \partial_t^{k-1} u(t)),$$

then the equation (1.4) is equivalent to

$$\partial_t \tilde{u}(t) = \tilde{F}(\tilde{u}(t))$$

where $\tilde{F} : \mathcal{D}^k \to \mathcal{D}^k$ is the function

$$\tilde{F}(u_0, \ldots, u_{k-1}) = (u_1, \ldots, u_{k-1}, F(u_0, \ldots, u_{k-1})).$$

Furthermore, $\tilde{u}$ is continuously differentiable if and only if u is k times continuously differentiable, and the k initial conditions

$$u(t_0) = u_0; \quad \partial_t u(t_0) = u_1; \quad \ldots; \quad \partial_t^{k-1} u(t_0) = u_{k-1}$$

collapse to a single initial condition

$$\tilde{u}(t_0) = \tilde{u}_0 := (u_0, \ldots, u_{k-1}).$$

Thus for the remainder of this chapter, we shall focus primarily on the initial value problem (or *Cauchy problem*) of obtaining solutions $u(t)$ to the first-order ODE[7]

$$\partial_t u(t) = F(u(t)) \text{ for all } t \in I; \quad u(t_0) = u_0. \tag{1.7}$$

where the interval I, the initial time t_0, the initial datum $u_0 \in \mathcal{D}$, and the nonlinearity $F : \mathcal{D} \to \mathcal{D}$ are given. We will distinguish three types of solutions:

---

[6]One can of course consider a meromorphic continuation beyond $t = 1$, but this would require complexifying time, which is a somewhat non-physical operation. Also, this complexification now relies very heavily on the analyticity of the situation, and when one goes from ODE to PDE, it is unlikely to work for non-analytic initial data. The question of whether one can continue a solution in some weakened sense beyond a singularity is an interesting and difficult one, but we will not pursue it in this text.

[7]One can interpret F as a vector field on the state space $\mathcal{D}$, in which case the ODE is simply integrating this vector field; see Figure 1. This viewpoint is particularly useful when considering the Hamiltonian structure of the ODE as in Section 1.4, however it is not as effective a conceptual framework when one passes to PDE.

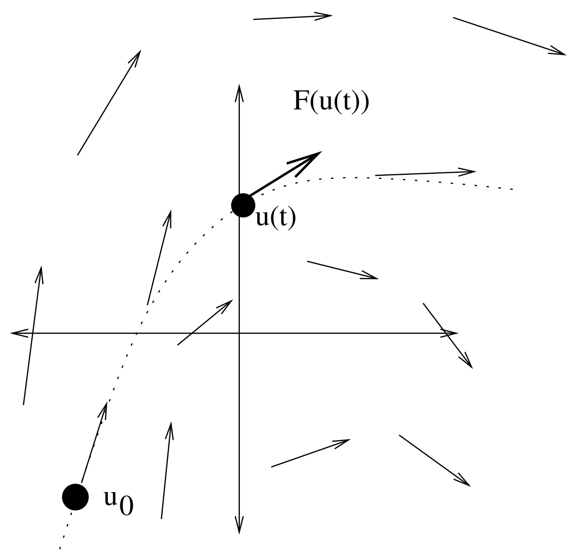

FIGURE 1. Depicting F as a vector field on $\mathcal{D}$, the trajectory of the solution $u(t)$ to the first order ODE (1.7) thus "follows the arrows" and *integrates* the vector field F. Contrast this "classical solution" interpretation of an ODE with the rather different "strong solution" interpretation in Figure 2.

- A *classical solution* of (1.7) is a function $u \in C^1_{\mathrm{loc}}(I \to \mathcal{D})$ which solves (1.7) for all $t \in I$ in the classical sense (i.e. using the classical notion of derivative).
- A *strong solution* of (1.7) is a function $u \in C^0_{\mathrm{loc}}(I \to \mathcal{D})$ which solves (1.7) in the integral sense that

$$u(t) = u_0 + \int_{t_0}^t F(u(s))\, ds \tag{1.8}$$

  holds for all[8] $t \in I$;
- A *weak solution* of (1.7) is a function $u \in L^\infty(I \to \mathcal{D})$ which solves (1.8) in the sense of distributions, thus for any test function $\psi \in C^\infty_0(I)$, one has

$$\int_I u(t)\psi(t)\, dt = u_0 \int_I \psi(t) + \int_I \psi(t) \int_{t_0}^t F(u(s))\, dsdt.$$

Later, when we turn our attention to PDE, these three notions of solution shall become somewhat distinct; see Section 3.2. In the ODE case, however, we fortunately have the following equivalence (under a very mild assumption on F):

LEMMA 1.3. *Let $F \in C^0_{\mathrm{loc}}(\mathcal{D} \to \mathcal{D})$. Then the notions of classical solution, strong solution, and weak solution are equivalent.*

PROOF. It is clear that a classical solution is strong (by the fundamental theorem of calculus), and that a strong solution is weak. If u is a weak solution, then

[8]Recall that we are adopting the convention that $\int_s^t = -\int_t^s$ if $t < s$.

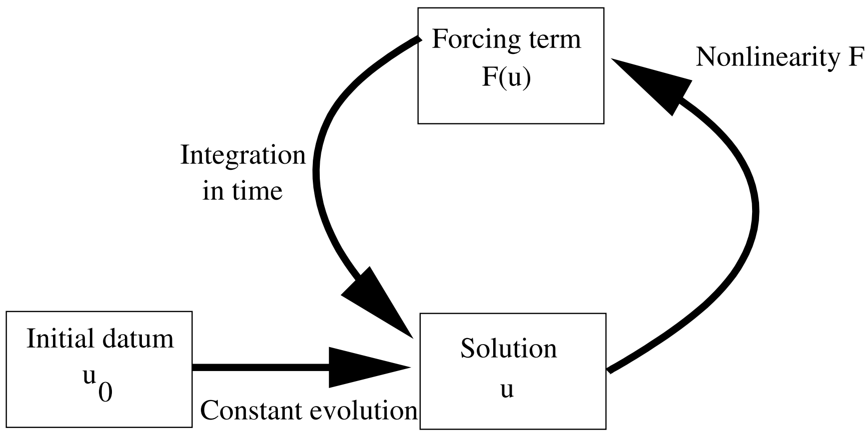

FIGURE 2. A schematic depiction of the relationship between the initial datum u_0, the solution $u(t)$, and the nonlinearity $F(u)$. The main issue is to control the "feedback loop" in which the solution influences the nonlinearity, which in turn returns to influence the solution.

it is bounded and measurable, hence $F(u)$ is also bounded and measurable. Thus the integral $\int_{t_0}^t F(u(s))\, ds$ is Lipschitz continuous, and (since u solves (1.8) in the sense of distributions) $u(t)$ is also Lipschitz continuous, so it is a strong solution (we allow ourselves the ability to modify u on a set of measure zero). Then $F(u)$ is continuous, and so the fundamental theorem of calculus and (1.8) again, u is in fact in C^1_{loc} and is a classical solution. $\square$

The three perspectives of classical, strong, and weak solutions are all important in the theory of ODE and PDE. The classical solution concept, based on the differential equation (1.7), is particularly useful for obtaining conservation laws (Section 1.4) and monotonicity formulae (Section 1.5), and for understanding symmetries of the equation. The strong solution concept, based on the integral equation (1.8), is more useful for constructing solutions (in part because it requires less *a priori* regularity on the solution), and establishing regularity and growth estimates on the solution. It also leads to a very important perspective on the equation, viewing the solution $u(t)$ as being the combination of two influences, one coming from the initial datum u_0 and the other coming from the forcing term $F(u)$; see Figure 2. Finally, the concept of a weak solution arises naturally when constructing solutions via compactness methods (e.g. by considering weak limits of classical solutions), since continuity is not *a priori* preserved by weak limits.

To illustrate the strong solution concept, we can obtain the first fundamental theorem concerning such Cauchy problems, namely the *Picard existence theorem*. We begin with a simplified version of this theorem to illustrate the main point.

THEOREM 1.4 (Picard existence theorem, simplified version). *Let $\mathcal{D}$ be a finite-dimensional normed vector space. Let $F \in \dot C^{0,1}(\mathcal{D} \to \mathcal{D})$ be a Lipschitz function on $\mathcal{D}$ with Lipschitz constant $\|F\|_{\dot C^{0,1}} = M$. Let $0 < T < 1/M$. Then for any $t_0 \in \mathbf{R}$ and $u_0 \in \mathcal{D}$, there exists a strong (hence classical) solution $u : I \to \mathcal{D}$ to the Cauchy problem* (1.7)*, where I is the compact time interval $I := [t_0 - T, t_0 + T]$.*

PROOF. Fix $u_0 \in \mathcal{D}$ and $t_0 \in \mathbf{R}$, and let $\Phi : C^0(I \to \mathcal{D}) \to C^0(I \to \mathcal{D})$ be the map

$$\Phi(u)(t) := u_0 + \int_{t_0}^{t} F(u(t'))\, dt'.$$

Observe from (1.8) that a strong solution is nothing more than a fixed point of the map Φ. It is easy to verify that Φ is indeed a map from $C^0(I \to \mathcal{D})$ to $C^0(I \to \mathcal{D})$. Using the Lipschitz hypothesis on F and the triangle inequality, we obtain

$$\|\Phi(u)(t) - \Phi(v)(t)\|_{\mathcal{D}} = \| \int_{t_0}^{t} F(u(t')) - F(v(t'))\, dt' \|_{\mathcal{D}} \leq \int_{t_0}^{t} M \|u(t') - v(t')\|_{\mathcal{D}}\, dt'$$

for all $t \in I$ and $u, v \in C^0(I \to \Omega_\varepsilon)$, and thus

$$\|\Phi(u) - \Phi(v)\|_{C^0(I \to \mathcal{D})} \leq TM \|u - v\|_{C^0(I \to \mathcal{D})}.$$

Since we have $TM < 1$, we see that Φ will be a strict contraction on the complete metric space $C^0(I \to \mathcal{D})$. Applying the contraction mapping theorem (Exercise 1.2) we obtain a fixed point to Φ, which gives the desired strong (and hence classical) solution to the equation (1.7). $\square$

REMARK 1.5. An inspection of the proof of the contraction mapping theorem reveals that the above argument in fact gives rise to an explicit iteration scheme that will converge to the solution u. Indeed, one can start with the constant solution $u^{(0)}(t) := u_0$, and then define further iterates $u^{(n)} \in C^0(I \to \Omega_\varepsilon)$ by $u^{(n)} := \Phi(u^{(n-1)})$, or in other words

$$u^{(n)}(t) := u_0 + \int_{t_0}^{t} F(u^{(n-1)}(t')\, dt'.$$

These *Picard iterates* do not actually solve the equation (1.7) in any of the above senses, but they do converge uniformly on I to the actual solution. See Figure 3.

REMARK 1.6. The above argument is perhaps the simplest example of the *iteration method* (also known as the *contraction mapping principle method* or the *inverse function theorem method*), constructing a nonlinear solution as the strong limit of an iterative procedure. This type of method will be our primary means of generating solutions which obey a satisfactory set of existence, uniqueness, and regularity properties. Note that one needs to select a norm $\|\|_{\mathcal{D}}$ in order to obtain a quantitative estimate on the time of existence. For finite-dimensional ODE, the exact choice of norm is not terribly important (as all norms are equivalent), but selecting the norm in which to apply the contraction mapping theorem will become decisive when studying PDE.

Because F is assumed to be globally Lipschitz ($\dot{C}^{0,1}$), one can actually construct a global solution to (1.7) in this case, by iterating the above theorem; see Exercise 1.10. However, in most applications F will only be locally Lipschitz ($\dot{C}^{0,1}_{\mathrm{loc}}$), and so we shall need a more general version of the above existence theorem. One such version (which also gives some further information, namely some Lipschitz continuity properties on the solution map) is as follows.

THEOREM 1.7 (Picard existence theorem, full version). *Let $\mathcal{D}$ be a finite-dimensional normed vector space. Let $t_0 \in \mathbf{R}$, let Ω be a non-empty subset of $\mathcal{D}$, and let $N_\varepsilon(\Omega) := \{u \in \mathcal{D} : \|u - v\|_{\mathcal{D}} < \varepsilon$ for some $v \in \Omega\}$ be the ε-neighbourhood of Ω for some $\varepsilon > 0$. Let $F : \mathcal{D} \to \mathcal{D}$ be a function which is Lipschitz on the closed*

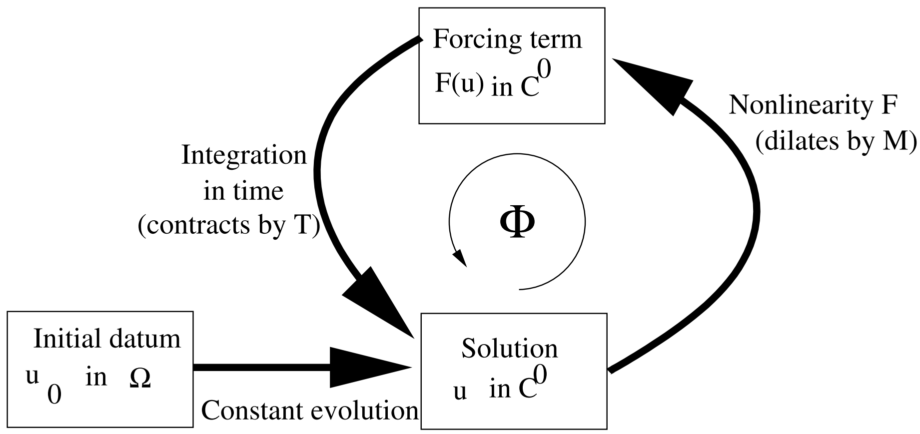

FIGURE 3. The Picard iteration scheme. The map Φ is basically the loop from the solution u to itself. To obtain the fixed point, start with the initial datum u_0 as the first approximant to u, and apply Φ repeatedly to obtain further approximations to u. As long as the net contraction factor $T \times M$ is less than 1, the iteration scheme will converge to an actual solution.

neighbourhood $\overline{N_\varepsilon(\Omega)}$ with some Lipschitz constant $\|F\|_{\dot C^{0,1}(\overline{N_\varepsilon(\Omega)})} = M > 0$, and which is bounded by some $A > 0$ on this region. Let $0 < T < \min(\varepsilon/A, 1/M)$, and let I be the interval $I := [t_0 - T, t_0 + T]$. Then for every $u_0 \in \Omega$, there exists a strong (hence classical) solution $u : I \to \overline{N_\varepsilon(\Omega)}$ to the Cauchy problem (1.7)*. Furthermore, if we then define the solution maps $S_{t_0}(t) : \Omega \to \mathcal{D}$ for $t \in I$ and $S_{t_0} : \Omega \to C^0(I \to \mathcal{D})$ by setting $S_{t_0}(t)(u_0) := u(t)$ and $S_{t_0}(u_0) := u$, then $S_{t_0}(t)$ and S_{t_0} are Lipschitz continuous maps, with Lipschitz constant at most $\frac{1}{1-TM}$.*

PROOF. Write $\Omega_\varepsilon := \overline{N_\varepsilon(\Omega)}$ for short. For each $u_0 \in \Omega$ let $\Phi_{u_0} : C^0(I \to \Omega_\varepsilon) \to C^0(I \to \Omega_\varepsilon)$ be the map

$$\Phi_{u_0}(u)(t) := u_0 + \int_{t_0}^{t} F(u(s))\, ds.$$

As before, a strong solution to (1.7) is nothing more than a fixed point of the map Φ_{u_0}. Since F is bounded by A on Ω_ε and $T < \varepsilon/A$, we see from the triangle inequality that Φ_{u_0} will indeed map $C^0(I \to \Omega_\varepsilon)$ to $C^0(I \to \Omega_\varepsilon)$. Also, since F has Lipschitz constant at most M on Ω_ε, we may argue as in the proof of Theorem 1.4 and conclude that Φ_{u_0} will thus be a strict contraction on the complete metric space $C^0(I \to \Omega_\varepsilon)$ with contraction constant $c := TM < 1$, and hence will have a fixed point $u = \Phi_{u_0}(u) \in C^0(I \to \Omega_\varepsilon)$. This gives a strong (and hence classical) solution to the equation (1.7).

Now let u_0 and $\tilde u_0$ be two initial data in Ω, with corresponding solutions $S_{t_0}(u_0) = u \in C^0(I \to \mathcal{D})$, $S_{t_0}(\tilde u_0) = \tilde u \in C^0(I \to \mathcal{D})$ constructed above. Observe from construction that $\Phi_{u_0}(u) = u$ and $\Phi_{u_0}(\tilde u) = \Phi_{\tilde u_0}(\tilde u) + u_0 - \tilde u_0 = \tilde u + u_0 - \tilde u_0$, thus

$$u - \tilde u = \Phi_{u_0}(u) - \Phi_{u_0}(\tilde u) + u_0 - \tilde u_0.$$

Taking norms and applying the contraction property and the triangle inequality, we conclude

$$\|u - \tilde{u}\|_{C^0(I \to \mathcal{D})} \leq c\|u - \tilde{u}\|_{C^0(I \to \mathcal{D})} + \|u_0 - \tilde{u}_0\|_{\mathcal{D}}$$

and hence

$$\|u - \tilde{u}\|_{C^0(I \to \mathcal{D})} \leq \frac{1}{1-c}\|u_0 - \tilde{u}_0\|_{\mathcal{D}}.$$

This proves the desired Lipschitz property on S_{t_0}, and hence on each individual $S_{t_0}(t)$. □

REMARK 1.8. The above theorem illustrates a basic point in nonlinear differential equations: in order to construct solutions, one does not need to control the nonlinearity $F(u)$ for *all* choices of state u, but only for those u that one expects to encounter in the evolution of the solution. For instance, if the initial datum is small, one presumably only needs to control $F(u)$ for small u in order to obtain a local existence result. This observation underlies many of the "perturbative" arguments which we shall see in this text (see for instance Proposition 1.24 below).

REMARK 1.9. In the next section we shall complement the Picard existence theorem with a uniqueness theorem. The hypothesis that F is locally Lipschitz can be weakened, but at the cost of losing the uniqueness; see Exercise 1.23.

EXERCISE 1.1. Begin the proof of the Cauchy-Kowalevski theorem by reducing to the case $k = 1$, $t_0 = 0$, and $u_0 = 0$. Then, use induction to show that if the higher derivatives $\partial_t^m u(0)$ are derived recursively as in (1.5), then we have some bound of the form

$$\|\partial_t^m u(0)\|_{\mathcal{D}} \leq K^{m+1} m!$$

for all $m \geq 0$ and some large $K > 0$ depending on F, where $\|\|_{\mathcal{D}}$ is some arbitrary norm on the finite-dimensional space $\mathcal{D}$. Then, define $u : I \to \mathcal{D}$ for some sufficiently small neighbourhood I of the origin by the power series

$$u(t) = \sum_{m=0}^{\infty} \frac{\partial_t^m u(0)}{m!} t^m$$

and show that $\partial_t u(t) - G(u(t))$ is real analytic on I and vanishes at infinite order at zero, and is thus zero on all of I.

EXERCISE 1.2. (Contraction mapping theorem) Let (X, d) be a complete non-empty metric space, and let $\Phi : X \to X$ be a strict contraction on X, thus there exists a constant $0 < c < 1$ such that $d(\Phi(u), \Phi(v)) \leq c d(u, v)$ for all $u, v \in X$. Show that Φ has a unique fixed point, thus there is a unique $u \in X$ such that $u = \Phi(u)$. Furthermore, if u_0 is an arbitrary element of X and we construct the sequence $u_1, u_2, \ldots \in X$ iteratively by $u_{n+1} := \Phi(u_n)$, show that u_n will converge to the fixed point u. Finally, we have the bound

$$d(v, u) \leq \frac{1}{1-c} d(v, \Phi(v)) \tag{1.9}$$

for all $v \in X$.

EXERCISE 1.3. (Inverse function theorem) Let $\mathcal{D}$ be a finite-dimensional vector space, and let $\Phi \in C^1_{\mathrm{loc}}(\mathcal{D} \to \mathcal{D})$ be such that $\nabla\Phi(x_0)$ has full rank for some $x_0 \in \mathcal{D}$. Using the contraction mapping theorem, show that there exists an open neighbourhood U of x_0 and an open neighbourhood V of $\Phi(x_0)$ such that Φ is a bijection from U to V, and that Φ^{-1} is also C^1_{loc}.

EXERCISE 1.4. Suppose we make the further assumption in the Picard existence theorem that $F \in C^k_{\mathrm{loc}}(\mathcal{D} \to \mathcal{D})$ for some $k \geq 1$. Show that the maps $S_{t_0}(t)$ and $S(t)$ are then also continuously k-times differentiable, and that $u \in C^{k+1}_{\mathrm{loc}}(I \to \mathcal{D})$.

EXERCISE 1.5. How does the Picard existence theorem generalise to higher order quasilinear ODE? What if there is time dependence in the nonlinearity (i.e. the ODE is non-autonomous)? The latter question can also be asked of the Cauchy-Kowaleski theorem. (These questions can be answered quickly by using the reduction tricks mentioned in this section.)

EXERCISE 1.6. One could naively try to extend the local solution given by the Picard existence theorem to a global solution by iteration, as follows: start with the initial time t_0, and use the existence theorem to construct a solution all the way up to some later time t_1. Then use $u(t_1)$ as a new initial datum and apply the existence theorem again to move forward to a later time t_2, and so forth. What goes wrong with this strategy, for instance when applied to the problem (1.6)?

1.2. Gronwall's inequality

It takes money to make money. (Proverbial)

As mentioned earlier, we will be most interested in the behaviour of ODE in very high dimensions. However, in many cases one can compress the key features of an equation to just a handful of dimensions, by isolating some important scalar quantities arising from the solution $u(t)$, for instance by inspecting some suitable norm $\|u(t)\|_{\mathcal{D}}$ of the solution, or looking at special quantities related to conservation or pseudoconservation laws such as energy, centre-of-mass, or variance. In many cases, these scalar quantities will not obey an exact differential equation themselves, but instead obey a *differential inequality*, which places an upper limit on how quickly these quantities can grow or decay. One is then faced with the task of "solving" such inequalities in order to obtain good bounds on these quantities for extended periods of time. For instance, if a certain quantity is zero or small at some time t_0, and one has some upper bound on its growth rate, one would like to say that it is still zero or small at later times. Besides the iteration method used already in the Picard existence theorem, there are two very useful tools for achieving this. One is *Gronwall's inequality*, which deals with linear growth bounds and is treated here. The other is the *continuity method*, which can be used with nonlinear growth bounds and is treated in Section 1.3.

We first give Gronwall's inequality in an integral form.

THEOREM 1.10 (Gronwall inequality, integral form). *Let $u : [t_0, t_1] \to \mathbf{R}^+$ be continuous and non-negative, and suppose that u obeys the integral inequality*

$$u(t) \leq A + \int_{t_0}^{t} B(s)u(s)\ ds \tag{1.10}$$

for all $t \in [t_0, t_1]$, where $A \geq 0$ and $B : [t_0, t_1] \to \mathbf{R}^+$ is continuous and nonnegative. Then we have

$$u(t) \leq A \exp\Big(\int_{t_0}^{t} B(s)\ ds\Big) \tag{1.11}$$

for all $t \in [t_0, t_1]$.

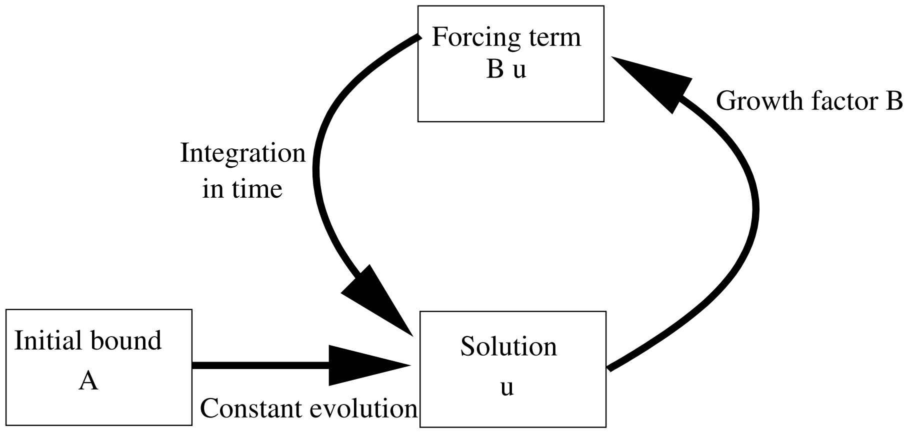

FIGURE 4. The linear feedback encountered in Theorem 1.10, that causes exponential growth by an amount depending on the growth factor B. Contrast this with Figure 2.

REMARK 1.11. This estimate is absolutely sharp, since the function $u(t) := A\exp(\int_{t_0}^t B(s)\ ds)$ obeys the hypothesis (1.10) with equality.

PROOF. By a limiting argument it suffices to prove the claim when $A > 0$. By (1.10) and the fundamental theorem of calculus, (1.10) implies

$$\frac{d}{dt}(A + \int_{t_0}^t B(s)u(s)\ ds) \leq B(t)(A + \int_{t_0}^t B(s)u(s)\ ds)$$

and hence by the chain rule

$$\frac{d}{dt}\log(A + \int_{t_0}^t B(s)u(s)\ ds) \leq B(t).$$

Applying the fundamental theorem of calculus again, we conclude

$$\log(A + \int_{t_0}^t B(s)u(s)\ ds) \leq \log A + \int_{t_0}^t B(s)\ ds.$$

Exponentiating this and applying (1.10) again, the claim follows. □

There is also a differential form of Gronwall's inequality in which B is allowed to be negative:

THEOREM 1.12 (Gronwall inequality, differential form). *Let $u : [t_0, t_1] \to \mathbf{R}^+$ be absolutely continuous and non-negative, and suppose that u obeys the differential inequality*

$$\partial_t u(t) \leq B(t)u(t)$$

for almost every $t \in [t_0, t_1]$, where $B : [t_0, t_1] \to \mathbf{R}^+$ is continuous and nonnegative. Then we have

$$u(t) \leq u(t_0)\exp(\int_{t_0}^t B(s)\ ds)$$

for all $t \in [t_0, t_1]$.

PROOF. Write $v(t) := u(t) \exp(-\int_{t_0}^t B(s)\, ds)$. Then v is absolutely continuous, and an application of the chain rule shows that $\partial_t v(t) \leq 0$. In particular $v(t) \leq v(t_0)$ for all $t \in [t_0, t_1]$, and the claim follows. $\square$

REMARK 1.13. This inequality can be viewed as controlling the effect of linear feedback; see Figure 4. As mentioned earlier, this inequality is sharp in the "worst case scenario" when $\partial_t u(t)$ equals $B(t)u(t)$ for all t. This is the case of "adversarial feedback", when the forcing term $B(t)u(t)$ is always acting to increase $u(t)$ by the maximum amount possible. Many other arguments in this text have a similar "worst-case analysis" flavour. In many cases (in particular, supercritical defocusing equations) it is suspected that the "average-case" behaviour of such solutions (i.e. for generic choices of initial data) is significantly better than what the worst-case analysis suggests, thanks to self-cancelling oscillations in the nonlinearity, but we currently have very few tools which can separate the average case from the worst case.

As a sample application of this theorem, we have

THEOREM 1.14 (Picard uniqueness theorem). *Let I be an interval. Suppose we have two classical solutions $u, v \in C^1_{\mathrm{loc}}(I \to \mathcal{D})$ to the ODE*

$$\partial_t u(t) = F(u(t))$$

for some $F \in \dot{C}^{0,1}_{\mathrm{loc}}(\mathcal{D} \to \mathcal{D})$. If u and v agree at one time $t_0 \in I$, then they agree for all times $t \in I$.

REMARK 1.15. Of course, the same uniqueness claim follows for strong or weak solutions, thanks to Lemma 1.3.

PROOF. By a limiting argument (writing I as the union of compact intervals) it suffices to prove the claim for compact I. We can use time translation invariance to set $t_0 = 0$. By splitting I into positive and negative components, and using the change of variables $t \mapsto -t$ if necessary, we may take $I = [0, T]$ for some $T > 0$.

Here, the relevant scalar quantity to analyze is the distance $\|u(t) - v(t)\|_{\mathcal{D}}$ between u and v, where $\|\|_{\mathcal{D}}$ is some arbitrary norm on $\mathcal{D}$. We then take the ODE for u and v and subtract, to obtain

$$\partial_t(u(t) - v(t)) = F(u(t)) - F(v(t)) \text{ for all } t \in [0, T]$$

Applying the fundamental theorem of calculus, the hypothesis $u(0) = v(0)$, and the triangle inequality, we conclude the integral inequality

$$\|u(t) - v(t)\|_{\mathcal{D}} \leq \int_0^t \|F(u(s)) - F(v(s))\|_{\mathcal{D}}\, ds \text{ for all } t \in [0, T]. \tag{1.12}$$

Since I is compact and u, v are continuous, we see that $u(t)$ and $v(t)$ range over a compact subset of $\mathcal{D}$. Since F is locally Lipschitz, we thus have a bound of the form $|F(u(s)) - F(v(s))| \leq M|u(s) - v(s)|$ for some finite $M > 0$. Inserting this into (1.12) and applying Gronwall's inequality (with $A = 0$), the claim follows. $\square$

REMARK 1.16. The requirement that F be Lipschitz is essential; for instance the non-Lipschitz Cauchy problem

$$\partial_t u(t) = p|u(t)|^{(p-1)/p}; \quad u(0) = 0 \tag{1.13}$$

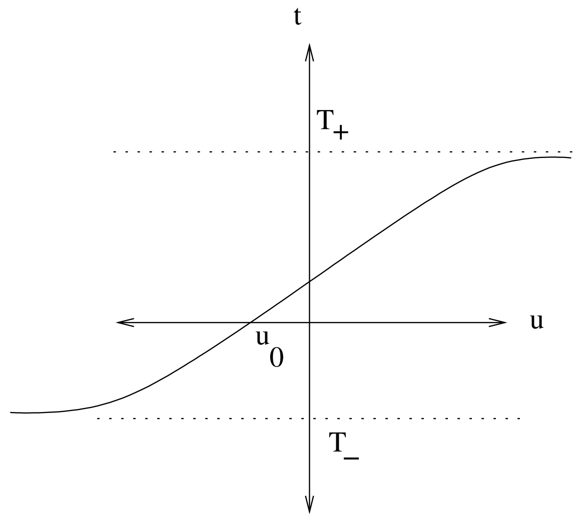

Figure 5. The maximal Cauchy development of an ODE which blows up both forwards and backwards in time. Note that in order for the time of existence to be finite, the solution $u(t)$ must go to infinity in finite time; thus for instance oscillatory singularities cannot occur (at least when the nonlinearity F is smooth).

for some $p > 1$ has the two distinct (classical) solutions $u(t) := 0$ and $v(t) := \max(0,t)^p$. Note that a modification of this example also shows that one cannot expect any continuous or Lipschitz dependence on the initial data in such cases.

By combining the Picard existence theorem with the Picard uniqueness theorem, we obtain

Theorem 1.17 (Picard existence and uniqueness theorem). *Let $F \in \dot{C}^{0,1}_{\mathrm{loc}}(\mathcal{D} \to \mathcal{D})$ be a locally Lipschitz function, let $t_0 \in \mathbf{R}$ be a time, and let $u_0 \in \mathcal{D}$ be an initial datum. Then there exists a maximal interval of existence $I = (T_-, T_+)$ for some $-\infty \leq T_- < t_0 < T_+ \leq +\infty$, and a unique classical solution $u : I \to \mathcal{D}$ to the Cauchy problem (1.7). Furthermore, if T_+ is finite, we have $\|u(t)\|_{\mathcal{D}} \to \infty$ as $t \to T_+$ from below, and similarly if T_- is finite then we have $\|u(t)\|_{\mathcal{D}} \to \infty$ as $t \to T_-$ from above.*

Remark 1.18. This theorem gives a *blowup criterion* for the Cauchy problem (1.7): a solution exists globally if and only if the $\|u(t)\|_{\mathcal{D}}$ norm does not go to infinity[9] in finite time; see Figure 5. (Clearly, if $\|u(t)\|_{\mathcal{D}}$ goes to infinity in finite time, u is not a global classical solution.) As we shall see later, similar blowup criteria (for various norms $\mathcal{D}$) can be established for certain PDE.

Proof. We define I to be the union of all the open intervals containing t_0 for which one has a classical solution to (1.7). By the existence theorem, I contains a

[9]We sometimes say that a solution *blows up at infinity* if the solution exists globally as $t \to \infty$, but that the norm $\|u(t)\|_{\mathcal{D}}$ is unbounded; note that Theorem 1.17 says nothing about whether a global solution will blow up at infinity or not, and indeed both scenarios are easily seen to be possible.

neighbourhood of t_0 and is clearly open and connected, and thus has the desired form $I = (T_-, T_+)$ for some $-\infty \leq T_- < t_0 < T_+ \leq +\infty$. By the uniqueness theorem, we may glue all of these solutions together and obtain a classical solution $u : I \to \mathcal{D}$ on (1.7). Now suppose for contradiction that T_+ was finite, and that there was some sequence of times t_n approaching T_+ from below for which $\|u(t)\|_{\mathcal{D}}$ stayed bounded. On this bounded set (or on any slight enlargement of this set) F is Lipschitz. Thus we may apply the existence theorem and conclude that one can extend the solution u to a short time beyond T_+; gluing this solution to the existing solution (again using the uniqueness theorem) we contradict the maximality of I. This proves the claim for T_+, and the claim for T_- is proven similarly. □

The Picard theorem gives a very satisfactory local theory for the existence and uniqueness of solutions to the ODE (1.7), assuming of course that F is locally Lipschitz. The issue remains, however, as to whether the interval of existence (T_-, T_+) is finite or infinite. If one can somehow ensure that $\|u(t)\|_{\mathcal{D}}$ does not blow up to infinity at any finite time, then the above theorem assures us that the interval of existence is all of $\mathbf{R}$; as we shall see in the exercises, Gronwall's inequality is one method in which one can assure the absence of blowup. Another common way to ensure global existence is to obtain a suitably "coercive" conservation law (e.g. energy conservation), which manages to contain the solution to a bounded set; see Proposition 1.24 below, as well as Section 1.4 for a fuller discussion. A third way is to obtain decay estimates, either via monotonicity formulae (see Section 1.5) or some sort of dispersion or dissipation effect. We shall return to all of these themes throughout this monograph, in order to construct global solutions to various equations.

Gronwall's inequality is *causal* in nature; in its hypothesis, the value of the unknown function $u(t)$ at times t is controlled by its value at previous times $0 < s < t$, averaged against a function $B(t)$ which can be viewed as a measure of the *feedback* present in the system; thus it is excessive feedback which leads to exponential growth (see Figure 4). This is of course very compatible with one's intuition regarding cause and effect, and our interpretation of t as a time variable. However, in some cases, when t is not being interpreted as a time variable, one can obtain integral inequalities which are *acausal* in that $u(t)$ is controlled by an integral of $u(s)$ both for $s < t$ and $s > t$. In many such cases, these inequalities lead to no useful conclusion. However, if the feedback is sufficiently weak, and one has some mild growth condition at infinity, one can still proceed as follows.

THEOREM 1.19 (Acausal Gronwall inequality). *Let $0 < \alpha' < \alpha$, $0 < \beta' < \beta$ and $\varepsilon, \delta > 0$ be real numbers. Let $u : \mathbf{R} \to \mathbf{R}^+$ be measurable and non-negative, and suppose that u obeys the integral inequality*

$$u(t) \leq A(t) + \delta \int_{\mathbf{R}} \min(e^{-\alpha(s-t)}, e^{-\beta(t-s)}) u(s)\, ds \tag{1.14}$$

for all $t \in \mathbf{R}$, where $A : \mathbf{R} \to \mathbf{R}^+$ is an arbitrary function. Suppose also that we have the subexponential growth condition

$$\sup_{t \in \mathbf{R}} e^{-\varepsilon|t|} u(t) < \infty.$$

Then if $\varepsilon < \min(\alpha, \beta)$ *and* δ *is sufficiently small depending on* $\alpha, \beta, \alpha', \beta', \varepsilon$, *we have*

$$u(t) \leq 2 \sup_{s \in \mathbf{R}} \min(e^{-\alpha'(s-t)}, e^{-\beta'(t-s)}) A(s). \tag{1.15}$$

for all $t \in \mathbf{R}$.

Proof. We shall use an argument similar in spirit to that of the contraction mapping theorem, though in this case there is no actual contraction to iterate as we have an integral *inequality* rather than an integral *equation*. By raising α' and β' (depending on $\varepsilon, \alpha, \beta$) if necessary we may assume $\varepsilon < \min(\alpha', \beta')$. We will assume that there exists $\sigma > 0$ such that $A(t) \geq \sigma e^{\varepsilon|t|}$ for all $t \in \mathbf{R}$; the general case can then be deduced by replacing $A(t)$ by $A(t) + \sigma e^{\varepsilon|t|}$ and then letting $\sigma \to 0$, noting that the growth of the $e^{\varepsilon|t|}$ factor will be compensated for by the decay of the $\min(e^{-\alpha'(s-t)}, e^{-\beta'(t-s)})$ factor since $\varepsilon < \min(\alpha', \beta')$. Let $B : \mathbf{R} \to \mathbf{R}^+$ denote the function

$$B(t) := \sup_{s \in \mathbf{R}} \min(e^{-\alpha'(s-t)}, e^{-\beta'(t-s)}) A(s).$$

Then we see that $\sigma e^{\varepsilon|t|} \leq A(t) \leq B(t)$, that B is strictly positive, and furthermore B obeys the continuity properties

$$B(s) \leq \max(e^{\alpha'(s-t)}, e^{\beta'(t-s)}) B(t) \tag{1.16}$$

for all $t, s \in \mathbf{R}$.

Let M be the smallest real number such that $u(t) \leq MB(t)$ for all $t \in \mathbf{R}$; our objective is to show that $M \leq 2$. Since B is bounded from below by $\sigma e^{\varepsilon|t|}$, we see from the subexponential growth condition that M exists and is finite. From (1.14) we have

$$u(t) \leq B(t) + \delta \int_{\mathbf{R}} \min(e^{-\alpha(s-t)}, e^{-\beta(t-s)}) u(s)\, ds.$$

Bounding $u(s)$ by $MB(s)$ and applying (1.16), we conclude

$$u(t) \leq B(t) + MB(t)\delta \int_{\mathbf{R}} \min(e^{-(\alpha-\alpha')(s-t)}, e^{-(\beta-\beta')(t-s)})\, ds.$$

Since $0 < \alpha' < \alpha$ and $0 < \beta' < \beta$, the integral is convergent and is independent of t. Thus if δ is sufficiently small depending on $\alpha, \beta, \alpha', \beta'$, we conclude that

$$u(t) \leq B(t) + \frac{1}{2} MB(t)$$

which by definition of M implies $M \leq 1 + \frac{1}{2}M$. Since M is finite, we have $M \leq 2$ as desired. □

The above inequality was phrased for a continuous parameter t, but it quickly implies a discrete analogue:

Corollary 1.20 (Discrete acausal Gronwall inequality). *Let* $0 < \alpha' < \alpha$, $0 < \beta' < \beta$, $\delta > 0$, *and* $0 < \varepsilon < \min(\alpha, \beta)$ *be real numbers. Let* $(u_n)_{n \in \mathbf{Z}}$ *be a sequence of non-negative numbers such that*

$$u_n \leq A_n + \delta \sum_{m \in \mathbf{Z}} \min(e^{-\alpha(m-n)}, e^{-\beta(n-m)}) u_m \tag{1.17}$$

for all $t \in \mathbf{R}$, where $(A_n)_{n\in\mathbf{Z}}$ is an arbitrary non-negative sequence. Suppose also that we have the subexponential growth condition

$$\sup_{n\in\mathbf{Z}} u_n e^{-\varepsilon|n|} < \infty.$$

Then if δ is sufficiently small depending on $\alpha, \beta, \alpha', \beta', \varepsilon$, we have

$$u_n \leq 2 \sup_{m\in\mathbf{Z}} \min(e^{-\alpha'(m-n)}, e^{-\beta'(n-m)}) A_m. \tag{1.18}$$

for all $n \in \mathbf{Z}$.

This corollary can be proven by modifying the proof of the previous theorem, or alternatively by considering the function $u(t) := u_{[t]}$, where $[t]$ is the nearest integer to t; we leave the details to the reader. This corollary is particularly useful for understanding the frequency distribution of solutions to nonlinear dispersive equations, in situations when the data is small (so the nonlinear effects of energy transfer between dyadic frequency ranges $|\xi| \sim 2^n$ are weak). See for instance [**Tao5**], [**Tao6**], [**Tao7**] for ideas closely related to this. One can also use these types of estimates to establish small energy regularity for various elliptic problems by working in frequency space (the smallness is needed to make the nonlinear effects weak).

Exercise 1.7 (Comparison principle). Let $I = [t_0, t_1]$ be a compact interval, and let $u : I \to \mathbf{R}$, $v : I \to \mathbf{R}$ be two scalar absolutely continuous functions. Let $F \in \dot{C}^{0,1}_{\mathrm{loc}}(I \times \mathbf{R} \to \mathbf{R})$, and suppose that u and v obey the differential inequalities

$$\partial_t u(t) \leq F(t, u(t)); \quad \partial_t v(t) \geq F(t, v(t))$$

for all $t \in I$. Show that if $u(t_0) \leq v(t_0)$, then $u(t) \leq v(t)$ for all $t \in [t_0, t_1]$, and similarly if $u(t_0) < v(t_0)$, then $u(t) < v(t)$ for all $t \in [t_0, t_1]$. (Hint: for the first claim, study the derivative of $\max(0, u(t) - v(t))^2$ and use Gronwall's inequality. For the second, perturb the first argument by an epsilon.)

Exercise 1.8. Give an example to show that Theorem 1.10 fails when B is permitted to be negative. Briefly discuss how this is consistent with the fact that Theorem 1.12 still holds for negative B.

Exercise 1.9 (Sturm comparison principle). Let I be a time interval, and let $u, v \in C^2_{\mathrm{loc}}(I \to \mathbf{R})$ and $a, f, g \in C^0_{\mathrm{loc}}(I \to \mathbf{R})$ be such that

$$\partial_t^2 u(t) + a(t)\partial_t u(t) + f(t)u(t) = \partial_t^2 v(t) + a(t)\partial_t v(t) + g(t)v(t) = 0$$

for all $t \in I$. Suppose also that v oscillates faster than u, in the sense that $g(t) \geq f(t)$ for all $t \in I$. Suppose also that u is not identically zero. Show that the zeroes of v *intersperse* the zeros of u, in the sense that whenever $t_1 < t_2$ are times in I such that $u(t_1) = u(t_2) = 0$, then v has at least one zero in the interval $[t_1, t_2]$. (Hint: reduce to the case when t_1 and t_2 are consecutive zeroes of u, and argue by contradiction. By replacing u or v with $-u$ or $-v$ if necessary one may assume that u, v are nonnegative on $[t_1, t_2]$. Obtain a first order equation for the *Wronskian* $u\partial_t v - v\partial_t u$.) This principle can be thought of as a substantial generalisation of the observation that the zeroes of the sine and cosine functions intersperse each other.

Exercise 1.10. Let $F \in \dot{C}^{0,1}_{\mathrm{loc}}(\mathcal{D} \to \mathcal{D})$ have at most linear growth, thus $\|F(u)\|_{\mathcal{D}} \lesssim 1 + \|u\|_{\mathcal{D}}$ for all $u \in \mathcal{D}$. Show that for each $u_0 \in \mathcal{D}$ and $t_0 \in \mathbf{R}$ there exists a unique classical global solution $u : \mathbf{R} \to \mathcal{D}$ to the Cauchy problem

(1.7). Also, show that the solution maps $S_{t_0}(t) : \mathcal{D} \to \mathcal{D}$ defined by $S_{t_0}(u_0) = u(t_0)$ are locally Lipschitz, obey the time translation invariance $S_{t_0}(t) = S_0(t - t_0)$, and the group laws $S_0(t)S_0(t') = S_0(t + t')$ and $S_0(0) = \mathrm{id}$. (Hint: use Gronwall's inequality to obtain bounds on $\|u(t)\|_{\mathcal{D}}$ in the maximal interval of existence (T_-, T_+) given by Theorem 1.17.) This exercise can be viewed as the limiting case $p = 1$ of Exercise 1.11 below.

EXERCISE 1.11. Let $p > 1$, let $\mathcal{D}$ be a finite-dimensional normed vector space, and let $F \in \dot{C}^{0,1}_{\mathrm{loc}}(\mathcal{D} \to \mathcal{D})$ have at most p^{th}-power growth, thus $\|F(u)\|_{\mathcal{D}} \lesssim 1 + \|u\|^p_{\mathcal{D}}$ for all $u \in \mathcal{D}$. Let $t_0 \in \mathbf{R}$ and $u_0 \in \mathcal{D}$, and let $u : (T_-, T_+) \to \mathcal{D}$ be the maximal classical solution to the Cauchy problem (1.7) given by the Picard theorem. Show that if T_+ is finite, then we have the lower bound

$$\|u(t)\|_{\mathcal{D}} \gtrsim_p (T_+ - t)^{-1/(p-1)}$$

as t approaches T_+ from below, and similarly for T_-. Give an example to show that this blowup rate is best possible.

EXERCISE 1.12 (Slightly superlinear equations). Suppose $F \in \dot{C}^{0,1}_{\mathrm{loc}}(\mathcal{D} \to \mathcal{D})$ has at most $x \log x$ growth, thus

$$\|F(u)\|_{\mathcal{D}} \lesssim (1 + \|u\|_{\mathcal{D}}) \log(2 + \|u\|_{D})$$

for all $u \in \mathcal{D}$. Do solutions to the Cauchy problem (1.7) exist classically for all time (as in Exercise 1.10), or is it possible to blow up (as in Exercise 1.11)? In the latter case, what is the best bound one can place on the growth of $\|u(t)\|_{\mathcal{D}}$ in time; in the former case, what is the best lower bound one can place on the blow-up rate?

EXERCISE 1.13 (Persistence of constraints). Let $u : I \to \mathcal{D}$ be a (classical) solution to the ODE $\partial_t u(t) = F(u(t))$ for some time interval I and some $F \in C^0_{\mathrm{loc}}(\mathcal{D} \to \mathcal{D})$, and let $H \in C^1_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$ be such that $\langle F(v), dH(v)\rangle = G(v)H(v)$ for some $G \in C^0_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$ and all $v \in \mathcal{D}$; here we use

$$\langle u, dH(v)\rangle := \frac{d}{d\varepsilon} H(v + \varepsilon u)|_{\varepsilon=0} \tag{1.19}$$

to denote the directional derivative of H at v in the direction u. Show that if $H(u(t))$ vanishes for one time $t \in I$, then it vanishes for all $t \in I$. Interpret this geometrically, viewing F as a vector field and studying the level surfaces of H. Note that it is necessary that the ratio G between $\langle F, dH\rangle$ and H be continuous; it is not enough merely for $\langle F, dH\rangle$ to vanish whenever H does, as can be seen for instance from the counterexample $H(u) = u^2$, $F(u) = 2|u|^{1/2}$, $u(t) = t^2$.

EXERCISE 1.14 (Compatibility of equations). Let $F, G \in C^1_{\mathrm{loc}}(\mathcal{D} \to \mathcal{D})$ have the property that

$$\langle F(v), dG(v)\rangle - \langle G(v), dF(v)\rangle = 0 \tag{1.20}$$

for all $v \in \mathcal{D}$. (The left-hand side has a natural interpretation as the *Lie bracket* $[F, G]$ of the differential operators $F \cdot \nabla$, $G \cdot \nabla$ associated to the vector fields F and G.) Show that for any $u_0 \in \mathcal{D}$, there exists a neighbourhood $B \subset \mathbf{R}^2$ of the origin, and a map $u \in C^2(B \to \mathcal{D})$ which satisfies the two equations

$$\partial_s u(s, t) = F(u(s, t)); \quad \partial_t u(s, t) = G(u(s, t)) \tag{1.21}$$

for all $(s, t) \in B$, with initial datum $u(0, 0) = u_0$. Conversely, if $u \in C^2(B \to \mathcal{D})$ solves (1.21) on B, show that (1.20) must hold for all v in the range of u. (Hint:

use the Picard existence theorem to construct u locally on the s-axis $\{t = 0\}$ by using the first equation of (1.21), and then for each fixed s, extend u in the t direction using the second equation of (1.21). Use Gronwall's inequality and (1.20) to establish that $u(s,t) - u(0,t) - \int_0^s F(u(s',t))\, ds' = 0$ for all (s,t) in a neighbourhood of the origin.) This is a simple case of *Frobenius's theorem*, regarding when a collection of vector fields can be simultaneously integrated.

Exercise 1.15 (Integration on Lie groups). Let H be a finite-dimensional vector space, and let G be a Lie group in $\mathrm{End}(H)$ (i.e. a group of linear transformations on H which is also a smooth manifold). Let $\mathfrak{g}$ be the Lie algebra of G (i.e. the tangent space of G at the identity). Let $g_0 \in G$, and let $X \in \dot{C}^{0,1}_{\mathrm{loc}}(\mathbf{R} \to \mathfrak{g})$ be arbitrary. Show that there exists a unique function $g \in C^1_{\mathrm{loc}}(\mathbf{R} \to G)$ such that $g(0) = g_0$ and $\partial_t g(t) = X(t)g(t)$ for all $t \in \mathbf{R}$. (Hint: first use Gronwall's inequality and Picard's theorem to construct a global solution $g : \mathbf{R} \to M_n(\mathbf{C})$ to the equation $\partial_t g(t) = X(t)g(t)$, and then use Gronwall's inequality again, and local coordinate patches of G, to show that g stays on G.) Show that the same claim holds if the matrix product $X(t)g(t)$ is replaced by the Lie bracket $[g(t), X(t)] := g(t)X(t) - X(t)g(t)$.

Exercise 1.16 (Levinson's theorem). Let $L \in C^0(\mathbf{R} \to \mathrm{End}(\mathcal{D}))$ be a time-dependent linear transformation acting on a finite-dimensional Hilbert space $\mathcal{D}$, and let $F \in C^0(\mathbf{R} \to \mathcal{D})$ be a time-dependent forcing term. Show that for every $u_0 \in \mathcal{D}$ there exists a global solution $u \in C^0_{\mathrm{loc}}(\mathbf{R} \to \mathcal{D})$ to the ODE $\partial_t u = L(t)u + F(t)$, with the bound

$$|u(t)| \leq (|u_0|_{\mathcal{D}} + \int_0^t |F(s)|_{\mathcal{D}}\, ds) \exp(\int_0^t \|(L(t) + L^*(t))/2\|_{op}\, dt)$$

for all $t \geq 0$. (Hint: control the evolution of $|u(t)|^2 = \langle u(t), u(t)\rangle_{\mathcal{D}}$.) Thus one can obtain global control on a linear ODE with arbitrarily large coefficients, as long as the largeness is almost completely contained in the skew-adjoint component of the linear operator. In particular, if F and the self-adjoint component of L are both absolutely integrable, conclude that $u(t)$ is bounded uniformly in t.

Exercise 1.17. Give examples to show that Theorem 1.19 and Corollary 1.20 fail (even when A is identically zero) if ε or δ become too large, or if the hypothesis that u has subexponential growth is dropped.

Exercise 1.18. Let $\alpha, \delta > 0$, let $d \geq 1$ be an integer, let $0 \leq \gamma < d$, and let $u : \mathbf{R}^d \to \mathbf{R}^+$ and $A : \mathbf{R}^d \to \mathbf{R}^+$ be locally integrable functions such that one has the pointwise inequality

$$u(x) \leq A(x) + \delta \int_{\mathbf{R}^d} \frac{e^{-\alpha|x-y|}}{|x-y|^\gamma} u(y)\, dy$$

for almost every $x \in \mathbf{R}^d$. Suppose also that u is a tempered distribution in addition to a locally integrable function. Show that if $0 < \alpha' < \alpha$ and δ is sufficiently small depending on α, α', γ, then we have the bound

$$u(x) \leq 2\|e^{-\alpha'|x-y|} A(y)\|_{L^\infty_y(\mathbf{R}^d)}$$

for almost every $x \in \mathbf{R}^d$. (Hint: you will need to regularise u first, averaging on a small ball, in order to convert the tempered distribution hypothesis into a pointwise subexponential bound. Then argue as in Proposition 1.19. One can then take limits at the end using the Lebesgue differentiation theorem.)

EXERCISE 1.19 (Singular ODE). Let $F, G \in \dot{C}^{0,1}(\mathcal{D} \to \mathcal{D})$ be Lipschitz maps with $F(0) = 0$ and $\|F\|_{\dot{C}^{0,1}(\mathcal{D} \to \mathcal{D})} < 1$. Show that there exists a $T > 0$ for which there exists a unique classical solution $u : (0, T] \to \mathcal{D}$ to the singular non-autonomous ODE $\partial_t u(t) = \frac{1}{t} F(u(t)) + G(u(t))$ with the boundary condition $\limsup_{t \to 0} \|u(t)\|_{\mathcal{D}}/t < \infty$ as $t \to 0$. (Hint: For uniqueness, use a Gronwall inequality argument. For existence, construct iterates in the space of functions $\{tv : v \in C^0([0,T] \to \mathcal{D})\}$.) Show that u in fact extends to a C^1 function on $[0, T]$ with $u(0) = 0$ and $\partial_t u(0) = G(0)$. Also give an example to show that uniqueness can break down when the Lipschitz constant of F exceeds 1. (You can take a very simple example, for instance with F linear and G zero.)

1.3. Bootstrap and continuity arguments

If you have built your castles in the air, your work need not be lost; that is where they should be. Now put the foundations under them.
(Henry David Thoreau, "Walden")

The Picard existence theorem allows us to construct solutions to ODE such as $\partial_t u(t) = F(u(t))$ on various time intervals. Once these solutions have been constructed, it is natural to then ask what kind of quantitative estimates and asymptotics these solutions satisfy, especially over long periods of time. If the equation is fortunate enough to be solvable exactly (which can happen for instance if the equation is completely integrable), then one can read off the desired estimates from the exact solution, in principle at least. However, in the majority of cases no explicit solution is available[10]. Many times, the best one can do is to write the solution $u(t)$ in terms of itself, using the strong solution concept. For instance, if the initial condition is $u(t_0) = u_0$, then we have

$$u(t) = u_0 + \int_{t_0}^{t} F(u(s))\, ds. \tag{1.22}$$

This equation tells us that if we have some information on u (for instance, if we control some norm $\|\|_Y$ of $u(s)$), we can insert this information into the right-hand side of the above integral equation (together with some knowledge of the initial datum u_0 and the nonlinearity F), and conclude some further control of the solution u (either in the same norm $\|\|_Y$, or in some new norm).

Thus we can use equations such as (1.22) to obtain control on u - but only if one starts with some control on u in the first place. Thus it seems difficult to get started when executing this strategy, since one often starts with only very little control on u, other than continuity. Nevertheless there is a simple principle, of almost magical power, that allows one to assume "for free" that u already obeys some quantitative bound, in order to prove that u obeys another quantitative bound - as long as the bound one ends up proving is slightly stronger than the bound one used as a hypothesis (to avoid circularity). This principle - which is a continuous analogue

[10]Of course, the contraction mapping argument used in Theorem 1.7 does in principle give a description of the solution, at least locally, as the limit of iterates of a certain integral map Φ, and the Cauchy-Kowalevski theorem in principle gives a Taylor series expansion of the solution. However in practice these expansions are rather complicated, and only useful for analyzing the short-time behaviour and not long-time behaviour. Even if an explicit solution (e.g. involving special functions) is available, it may be easier to read off the asymptotics and other features of the equation from an analytic argument such as a bootstrap argument than from inspection of the explicit solution.

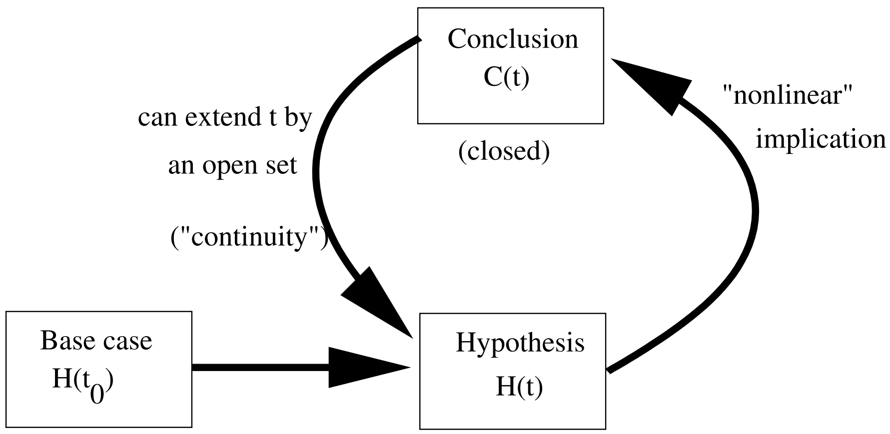

FIGURE 6. A schematic depiction of the relationship between the hypothesis $\mathbf{H}(t)$ and the conclusion $\mathbf{C}(t)$; compare this with Figure 2. The reasoning is noncircular because at each loop of the iteration we extend the set of times for which the hypothesis and conclusion are known to hold. The closure hypothesis prevents the iteration from getting stuck indefinitely at some intermediate time.

of the principle of mathematical induction - is known as the *bootstrap principle* or the *continuity method*[11]. Abstractly, the principle works as follows.

PROPOSITION 1.21 (Abstract bootstrap principle). *Let I be a time interval, and for each $t \in I$ suppose we have two statements, a "hypothesis" $\mathbf{H}(t)$ and a "conclusion" $\mathbf{C}(t)$. Suppose we can verify the following four assertions:*

(a) *(Hypothesis implies conclusion) If $\mathbf{H}(t)$ is true for some time $t \in I$, then $\mathbf{C}(t)$ is also true for that time t.*
(b) *(Conclusion is stronger than hypothesis) If $\mathbf{C}(t)$ is true for some $t \in I$, then $\mathbf{H}(t')$ is true for all $t' \in I$ in a neighbourhood of t.*
(c) *(Conclusion is closed) If $t_1, t_2, \ldots$ is a sequence of times in I which converges to another time $t \in I$, and $\mathbf{C}(t_n)$ is true for all t_n, then $\mathbf{C}(t)$ is true.*
(d) *(Base case) $\mathbf{H}(t)$ is true for at least one time $t \in I$.*

Then $\mathbf{C}(t)$ is true for all $t \in I$.

REMARK 1.22. When applying the principle, the properties $\mathbf{H}(t)$ and $\mathbf{C}(t)$ are typically chosen so that properties (b), (c), (d) are relatively easy to verify, with property (a) being the important one (and the "nonlinear" one, usually proven by exploiting one or more nonlinear feedback loops in the equations under study). The bootstrap principle shows that in order to prove a property $\mathbf{C}(t)$ obeying (c), it would suffice to prove the seemingly easier assertion $\mathbf{H}(t) \implies \mathbf{C}(t)$, as long as $\mathbf{H}$ is "weaker" than $\mathbf{C}$ in the sense of (b) and is true for at least one time.

[11]The terminology "bootstrap principle" arises because a solution u obtains its regularity from its own resources rather than from external assumptions - "pulling itself up by its bootstraps", as it were. The terminology "continuity method" is used because the continuity of the solution is essential to making the method work.

Proof. Let Ω be the set of times $t \in I$ for which $\mathbf{C}(t)$ holds. Properties (d) and (a) ensure that Ω is non-empty. Properties (b) and (a) ensure that Ω is open. Property (c) ensures that Ω is closed. Since the interval I is connected, we thus see that $\Omega = I$, and the claim follows. □

More informally, one can phrase the bootstrap principle as follows:

Principle 1.23 (Informal bootstrap principle). *If a quantity u can be bounded in a nontrivial way in terms of itself, then under reasonable conditions, one can conclude that u is bounded* unconditionally.

We give a simple example of the bootstrap principle in action, establishing global existence for a system in a locally stable potential well from small initial data.

Proposition 1.24. *Let $\mathcal{D}$ be a finite-dimensional Hilbert space, and let $V \in C^2_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$ be such that such that $V(0) = 0$, $\nabla V(0) = 0$, and $\nabla^2 V(0)$ is strictly positive definite. Then for all $u_0, u_1 \in \mathcal{D}$ sufficiently close to 0, there is a unique classical global solution $u \in C^2_{\mathrm{loc}}(\mathbf{R} \to \mathcal{D})$ to the Cauchy problem*

$$\partial_t^2 u(t) = -\nabla V(u(t)); \quad u(0) = u_0; \quad \partial_t u(0) = u_1. \tag{1.23}$$

Furthermore, this solution stays bounded uniformly in t.

Remark 1.25. The point here is that the potential well V is known to be stable near zero by hypothesis, but could be highly unstable away from zero; see Figure 7. Nevertheless, the bootstrap argument can be used to prevent the solution from "tunnelling" from the stable region to the unstable region.

Proof. Applying the Picard theorem (converting the second-order ODE into a first-order ODE in the usual manner) we see that there is a maximal interval of existence $I = (T_-, T_+)$ containing 0, which supports a unique classical solution $u \in C^2_{\mathrm{loc}}(I \to \mathcal{D})$ to the Cauchy problem (1.23). Also, if T_+ is finite, then we have $\lim_{t \to T_+} \|u(t)\|_{\mathcal{D}} + \|\partial_t u(t)\|_{\mathcal{D}} = \infty$, and similarly if T_- is finite.

For any time $t \in I$, let $E(t)$ denote the energy

$$E(t) := \frac{1}{2}\|\partial_t u(t)\|^2_{\mathcal{D}} + V(u(t)). \tag{1.24}$$

From (1.23) we see that

$$\partial_t E(t) = \langle \partial_t u(t), \partial_t^2 u(t)\rangle + \langle \partial_t u(t), \nabla V(u(t))\rangle = 0$$

and thus we have the conservation law

$$E(t) = E(0) = \frac{1}{2}\|u_1\|^2_{\mathcal{D}} + V(u_0).$$

If u_0, u_1 are sufficiently close to 0, we can thus make $E(t) = E(0)$ as small as desired.

The problem is that we cannot quite conclude from the smallness of E that u is itself small, because V could turn quite negative away from the origin. However, such a scenario can only occur when u is large. Thus we need to assume that u is small in order to prove that u is small. This may seem circular, but fortunately the bootstrap principle allows one to justify this argument.

Let $\varepsilon > 0$ be a parameter to be chosen later, and let $\mathbf{H}(t)$ denote the statement

$$\|\partial_t u(t)\|^2_{\mathcal{D}} + \|u(t)\|^2_{\mathcal{D}} \le (2\varepsilon)^2$$

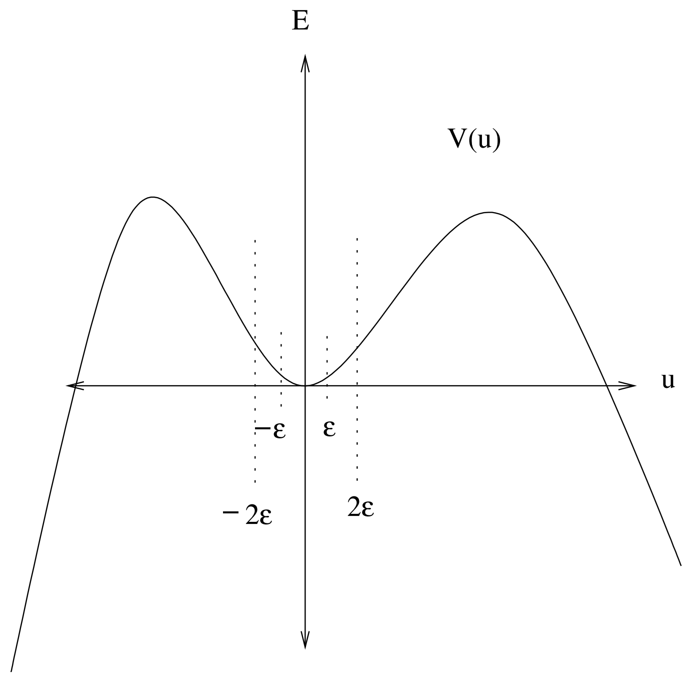

FIGURE 7. The potential well V in Proposition 1.24. As long as $\|\partial_t u(t)\|_{\mathcal{D}}^2 + \|u(t)\|_{\mathcal{D}}^2$ is known to be bounded by $(2\varepsilon)^2$, the Hamiltonian becomes coercive and energy conservation will trap a particle in the region $\|\partial_t u(t)\|_{\mathcal{D}}^2 + \|u(t)\|_{\mathcal{D}}^2 \leq \varepsilon^2$ provided that the initial energy is sufficiently small. The bootstrap hypothesis can be removed because the motion of the particle is continuous. Without that bootstrap hypothesis, it is conceivable that a particle could discontinuously "tunnel" through the potential well and escape, without violating conservation of energy.

and let $\mathbf{C}(t)$ denote the statement

$$\|\partial_t u(t)\|_{\mathcal{D}}^2 + \|u(t)\|_{\mathcal{D}}^2 \leq \varepsilon^2.$$

Since u is continuously twice differentiable, and blows up at any finite endpoint of I, we can easily verify properties (b) and (c) of the bootstrap principle, and if u_0 and u_1 are sufficiently close to 0 (depending on ε) we can also verify (d) at time $t = 0$. Now we verify (a), showing that the hypothesis $\mathbf{H}(t)$ can be "bootstrapped" into the stronger conclusion $\mathbf{C}(t)$. If $\mathbf{H}(t)$ is true, then $\|u(t)\|_{\mathcal{D}} = O(\varepsilon)$. We then see from the hypotheses on V and Taylor expansion that

$$V(u(t)) \geq c\|u(t)\|_{\mathcal{D}}^2 + O(\varepsilon^3)$$

for some $c > 0$. Inserting this into (1.24), we conclude

$$\frac{1}{2}\|\partial_t u(t)\|_{\mathcal{D}}^2 + c\|u(t)\|_{\mathcal{D}}^2 \leq E(0) + O(\varepsilon^3).$$

This is enough to imply the conclusion $\mathbf{C}(t)$ as long as ε is sufficiently small, and $E(0)$ is also sufficiently small. This closes the bootstrap, and allows us to conclude

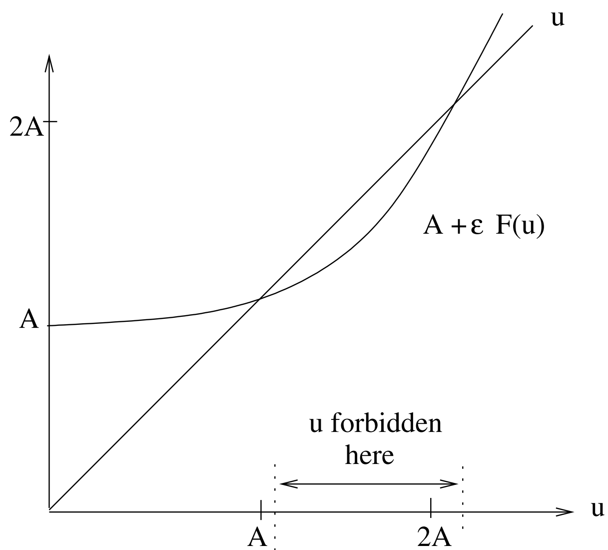

FIGURE 8. A depiction of the situation in Exercise 1.21. Note the impenetrable barrier in the middle of the u domain.

that $\mathbf{C}(t)$ is true for all $t \in I$. In particular, I must be infinite, since we know that $\|\partial_t u(t)\|_{\mathcal{D}}^2 + \|u(t)\|_{\mathcal{D}}^2$ would blow up at any finite endpoint of I, and we are done. $\square$

One can think of the bootstrap argument here as placing an "impenetrable barrier"

$$\varepsilon^2 < \|\partial_t u(t)\|_{\mathcal{D}}^2 + \|u(t)\|_{\mathcal{D}}^2 \leq (2\varepsilon)^2$$

in phase space. Property (a) asserts that the system cannot venture into this barrier. Properties (b), (c) ensure that this system cannot "jump" from one side of the barrier to the other instantaneously. Property (d) ensures that the system starts out on the "good" side of the barrier. We can then conclude that the system stays in the good side for all time; see Figure 7. Note also the division of labour in proving these properties. The properties (b), (c) are proven using the local existence theory (i.e. Theorem 1.17). The property (d) comes from the hypotheses on the initial datum. The property (a) requires some structural information on the equation, in this case the existence of a conserved energy $E(t)$ with enough locally "coercive" properties to contain the system within the desired barrier. This pattern of argument is very common in the analysis of nonlinear ODE and PDE, and we shall see several more examples of this later in this monograph.

EXERCISE 1.20. Show by example that Proposition 1.21 fails if any one of its four hypotheses are removed.

EXERCISE 1.21. Let I be a time interval, and $u \in C^0_{\mathrm{loc}}(I \to \mathbf{R}^+)$ be a non-negative function obeying the inequality

$$u(t) \leq A + \varepsilon F(u(t)) \tag{1.25}$$

for some $A, \varepsilon > 0$ and some function $F : \mathbf{R}^+ \to \mathbf{R}^+$ which is locally bounded. Suppose also that $u(t_0) \leq 2A$ for some $t_0 \in I$. If ε is sufficiently small depending on A and F, show that in fact $u(t) \leq 2A$ for all $t \in I$. Show that the conclusion can fail if u is not continuous or ε is not small. Note however that no assumption is made on the growth of F at infinity. Informally speaking, this means that if one ever obtains an estimate of the form $u \leq A + \varepsilon F(u)$, then one can drop the $\varepsilon F(u)$ term (at the cost of increasing the main term A by a factor of 2) provided that ε is suitably small, some initial condition is verified, and some continuity is available. This is particularly useful for showing that a nonlinear solution obeys almost the same estimates as a linear solution if the nonlinear effect is sufficiently weak. Compare this with Principle 1.23.

EXERCISE 1.22. Let I be a time interval, and let $u \in C^0_{\mathrm{loc}}(I \to \mathbf{R}^+)$ obey the inequality

$$u(t) \leq A + \varepsilon F(u(t)) + Bu(t)^\theta$$

for some $A, B, \varepsilon > 0$ and $0 < \theta < 1$, and some locally bounded function $F : \mathbf{R}^+ \to \mathbf{R}^+$. Suppose also that $u(t_0) \leq A'$ for some $t_0 \in I$ and $A' > 0$. Show that if ε is sufficiently small depending on A, A', B, θ, F, then we have $u(t) \lesssim_\theta A + B^{1/(1-\theta)}$ for all $t \in I$. Thus we can tolerate an additional u-dependent term on the right-hand side of (1.25) as long as it grows slower than linearly in u.

EXERCISE 1.23 (Compactness solutions). Let $t_0 \in \mathbf{R}$ be a time, let $u_0 \in \mathcal{D}$, and let $F : \mathcal{D} \to \mathcal{D}$ be a function which is continuous (and hence bounded) in a neighbourhood of u_0. Show that there exists an open time interval I containing t_0, and a classical solution $u \in C^1(I \to \mathcal{D})$ to the Cauchy problem (1.7). (Hint: approximate F by a sequence of Lipschitz functions F_m and apply Theorem 1.17 to obtain solutions u_m to the problem $\partial_t u_m = F_m(u_m)$ on some maximal interval $(T_{-,m}, T_{+,m})$. Use a bootstrap argument and Gronwall's inequality to show that for some fixed open interval I (independent of m) containing t_0, the solutions u_m will stay uniformly bounded and uniformly Lipschitz (hence equicontinuous) in this interval, and that this interval is contained inside all of the $(T_{-,m}, T_{+,m})$. Then apply the Arzela-Ascoli theorem to extract a uniformly convergent subsequence of the u_m on I, and see what happens to the integral equations $u_m(t) = u_m(t_0) + \int_{t_0}^t F_m(u_m(s))\, ds$ in the limit, using Lemma 1.3 if necessary.) This is a simple example of a *compactness method* to construct solutions to equations such as (1.13), for which uniqueness is not available.

EXERCISE 1.24 (Persistence of constraints, II). Let $u : [t_0, t_1] \to \mathcal{D}$ be a classical solution to the ODE $\partial_t u(t) = F(u(t))$ for some continuous $F : \mathcal{D} \to \mathcal{D}$, and let $H_1, \ldots, H_n \in C^1_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$ have the property that

$$\langle F(v), dH_j(v) \rangle \geq 0$$

whenever $1 \leq j \leq n$ and $v \in \mathcal{D}$ is such that $H_j(v) = 0$ and $H_i(v) \geq 0$ for all $1 \leq i \leq n$. Show that if the statement

$$H_i(u(t)) \geq 0 \text{ for all } 1 \leq i \leq n$$

is true at time $t = t_0$, then it is true for all times $t \in [t_0, t_1]$. Compare this result with Exercise 1.13.

EXERCISE 1.25 (Forced blowup). Let $k \geq 1$, and let $u : [0, T_*) \to \mathbf{R}$ be a classical solution to the equation $\partial_t^k u(t) = F(u(t))$, where $F : \mathbf{R} \to \mathbf{R}$ is continuous.

Suppose that $u(0) > 0$ and $\partial_t^j u(0) \geq 0$ for all $1 \leq j < k$, and suppose that one has the lower bound such that $F(v) \gtrsim v^p$ for all $v \geq u(0)$ and some $p > 1$. Conclude the upper bound $T_* \lesssim_{p,k} u(0)^{(1-p)/k}$ on the time of existence. (Hint: first establish that $u(t) \geq u(0)$ and $\partial_t^j u(t) \geq 0$ for all $1 \leq j < k$ and $0 \leq t < T_*$, for instance by using Exercise 1.24. Then bootstrap these bounds to obtain some estimate on the doubling time of u, in other words to obtain an upper bound on the first time t for which $u(t)$ reaches $2u(0)$.) This shows that equations of the form $\partial_t^k u(t) = F(u(t))$ can blow up if the initial datum is sufficiently large and positive.

Exercise 1.26. Use the continuity method to give another proof of Gronwall's inequality (Theorem 1.10). (Hint: for technical reasons it may be easier to first prove that $u(t) \leq (1+\varepsilon)A\exp(\int_{t_0}^t B(s)\ ds)$ for each $\varepsilon > 0$, as continuity arguments generally require "an epsilon of room".) This alternate proof of Gronwall's inequality is more robust, as it can handle additional nonlinear terms on the right-hand side provided that they are suitably small.

1.4. Noether's theorem

Now symmetry and consistency are convertible terms - thus Poetry and Truth are one. (Edgar Allen Poe, "Eureka: A Prose Poem")

A remarkable feature of many important differential equations, especially those arising from mathematical physics, is that their dynamics, while complex, still continue to maintain a certain amount of unexpected structure. One of the most important examples of such structures are *conservation laws* - certain scalar quantities of the system that remain constant throughout the evolution of the system; another important example are *symmetries* of the equation - that there often exists a rich and explicit group of transformations which necessarily take one solution of the equation to another. A remarkable result of Emmy Noether shows that these two structures are in fact very closely related, provided that the differential equation is *Hamiltonian*; as we shall see, many interesting nonlinear dispersive and wave equations will be of this type. Noether's theorem is one of the fundamental theorems of Hamiltonian mechanics, and has proven to be extremely fruitful in the analysis of such PDE. Of course, the field of Hamiltonian mechanics offers many more beautiful mathematical results than just Noether's theorem; it is of great interest to see how much else of this theory (which is still largely confined to ODE) can be extended to the PDE setting. See [**Kuk3**] for some further discussion.

Noether's theorem can be phrased *symplectically*, in the context of Hamiltonian mechanics, or *variationally*, in the context of Lagrangian mechanics. We shall opt to focus almost exclusively on the former; the variational perspective has certain strengths for the nonlinear PDE we shall analyse (most notably in elucidating the role of the stress-energy tensor, and of the distinguished role played by ground states) but we will not pursue it in detail here (though see Exercises 1.44, 1.45, 2.60). We shall content ourselves with describing only a very simple special case of this theorem; for a discussion of Noether's theorem in full generality, see [**Arn**].

Hamiltonian mechanics can be defined on any symplectic manifold, but for simplicity we shall restrict our attention to symplectic vector spaces.

Definition 1.26. A *symplectic vector space* $(\mathcal{D}, \omega)$ is a finite-dimensional real vector space $\mathcal{D}$, equipped with a *symplectic form* $\omega : \mathcal{D} \times \mathcal{D} \to \mathbf{R}$, which is bilinear and anti-symmetric, and also non-degenerate (so for each non-zero $u \in \mathcal{D}$ there exists a $v \in \mathcal{D}$ such that $\omega(u, v) \neq 0$). Given any $H \in C^1_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$, we define the *symplectic gradient* $\nabla_\omega H \in C^0_{\mathrm{loc}}(\mathcal{D} \to \mathcal{D})$ to be the unique function such that

$$\langle v, dH(u) \rangle = \frac{d}{d\varepsilon} H(u + \varepsilon v)|_{\varepsilon=0} = \omega(\nabla_\omega H(u), v); \tag{1.26}$$

this definition is well-defined thanks to the non-degeneracy of ω and the finite dimensionality of $\mathcal{D}$. Given two functions $H, E \in C^1_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$, we define the *Poisson bracket* $\{H, E\} : \mathcal{D} \to \mathbf{R}$ by the formula

$$\{H, E\}(u) := \omega(\nabla_\omega H(u), \nabla_\omega E(u)). \tag{1.27}$$

A *Hamiltonian function* on a phase space $(\mathcal{D}, \omega)$ is any function[12] $H \in C^2_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$; to each such Hamiltonian, we associate the corresponding *Hamiltonian flow*

$$\partial_t u(t) = \nabla_\omega H(u(t)). \tag{1.28}$$

Note that with this definition, Hamiltonian ODE (1.28) are automatically autonomous (time-translation-invariant). However it is possible to consider time-varying Hamiltonians also: see Exercise 1.42. Note that the Hamiltonian of an equation is only determined up to a constant, since replacing H by $H + C$ does not affect the symplectic gradient of H.

EXAMPLE 1.27. If

$$\mathcal{D} := \mathbf{R}^n \times \mathbf{R}^n = \{(q_1, \dots, q_n, p_1, \dots, p_n) : q_1, \dots, q_n, p_1, \dots, p_n \in \mathbf{R}\}$$

for some $n \geq 1$, and $\omega : \mathcal{D} \times \mathcal{D} \to \mathbf{R}$ is the bilinear form

$$\omega := \sum_{j=1}^n dq_j \wedge dp_j$$

or in other words

$$\omega((q_1, \dots, q_n, p_1, \dots, p_n), (q_1', \dots, q_n', p_1', \dots, p_n')) := \sum_{j=1}^n p_j' q_j - p_j q_j'$$

then $(\mathcal{D}, \omega)$ is symplectic, and for any $H, E \in C^1_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$ we have

$$\nabla_\omega H = (\frac{\partial H}{\partial p_1}, \dots, \frac{\partial H}{\partial p_n}, -\frac{\partial H}{\partial q_1}, \dots, -\frac{\partial H}{\partial q_n})$$

and

$$\{H, E\} = \sum_{j=1}^n \frac{\partial H}{\partial p_j} \frac{\partial E}{\partial q_j} - \frac{\partial H}{\partial q_j} \frac{\partial E}{\partial p_j}.$$

In particular, the Hamiltonian ODE associated to a Hamiltonian function $H \in C^2_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$ is given by *Hamilton's equations of motion*

$$\partial_t q_j(t) = \frac{\partial H}{\partial p_j}(q(t), p(t)); \quad \partial_t p_j(t) = -\frac{\partial H}{\partial q_j}(q(t), p(t)), \tag{1.29}$$

where we write

$$u(t) = (q(t), p(t)); \quad q(t) = (q_1(t), \dots, q_n(t)); \quad p(t) = (p_1(t), \dots, p_n(t)).$$

[12] One can weaken this hypothesis of continuous twice differentiability and still define a Hamiltonian flow, but the theory becomes more delicate and we will not address it here.

Thus, for instance, if H takes the form

$$H(q,p) = \frac{1}{2m}|p|^2 + V(q) \tag{1.30}$$

where $|p|^2 := p_1^2 + \ldots + p_n^2$, $m > 0$ is a constant, and $V \in C^2_{\mathrm{loc}}(\mathbf{R}^n \to \mathbf{R})$, then Hamilton's equations become *Newton's laws of motion*

$$\partial_t q(t) = \frac{1}{m}p(t); \quad \partial_t p(t) = -\nabla V(q(t)).$$

EXAMPLE 1.28. Let $\mathcal{D} = \mathbf{C}^n$ be endowed with the symplectic form

$$\omega := \sum_{j=1}^n \frac{1}{2}\mathrm{Im}(dz_j \wedge d\overline{z_j}) \tag{1.31}$$

or in other words

$$\omega((z_1,\ldots,z_n),(z_1',\ldots,z_n')) := \sum_{j=1}^n \mathrm{Im}(z_j\overline{z_j'}).$$

Then for any $H, E \in C^1_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$ we have

$$\nabla_\omega H = (2i\frac{\partial H}{\partial \overline{z_1}}, \ldots, 2i\frac{\partial H}{\partial \overline{z_n}})$$

and

$$\{H, E\} = 4\sum_{j=1}^n \mathrm{Im}(\frac{\partial H}{\partial \overline{z_j}}\frac{\partial E}{\partial z_j})$$

where $\frac{\partial H}{\partial z} := \frac{\partial H}{\partial x} - i\frac{\partial H}{\partial y}$ and $\frac{\partial H}{\partial \overline{z}} := \frac{\partial H}{\partial x} + i\frac{\partial H}{\partial y}$. Thus for instance if H is the simple harmonic oscillator

$$H(z) := \sum_{j=1}^n \frac{1}{2}\lambda_j|z_j|^2 \tag{1.32}$$

for some $\lambda_1, \ldots, \lambda_n \in \mathbf{R}$, then the equations of motion are

$$\partial_t z_j(t) = i\lambda_j z_j(t).$$

This is in fact the canonical form for any quadratic Hamiltonian; see Exercise 1.41.

Hamiltonian equations enjoy a number of good properties. Since $H \in C^2_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$, the function $\nabla_\omega H \in C^{0,1}_{\mathrm{loc}}(\mathcal{D} \to \mathcal{D})$ is locally Lipschitz, and so the Picard existence theorem applies; in particular, for any bounded set $\Omega \subset \mathcal{D}$, there is a positive time $T > 0$ for which we have Lipschitz flow maps $S(t) = S_0(t) \in \dot{C}^{0,1}(\Omega \to \mathcal{D})$ defined for $|t| \le T$. In the *quadratic growth* case when $\nabla^2 H$ is bounded, then $\nabla_\omega H$ is globally Lipschitz, and one thus has global flow maps $S(t) \in \dot{C}^{0,1}_{\mathrm{loc}}(\mathcal{D} \to \mathcal{D})$ defined for all times $t \in \mathbf{R}$ (by Exercise 1.10). These maps obey the group laws $S(t + t') = S(t)S(t')$, $S(0) = \mathrm{id}$. Furthemore the $S(t)$ are diffeomorphisms and symplectomorphisms; see Exercise 1.4 and Exercise 1.32.

Let $H, E \in C^2_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$ be two Hamiltonians on a symplectic phase space $(\mathcal{D}, \omega)$, and let $u \in C^1_{\mathrm{loc}}(I \to \mathcal{D})$ be a classical solution to the Hamiltonian ODE

(1.28). From the chain rule and (1.26), (1.27), we thus have the *Poisson equation*[13]

$$\frac{d}{dt}E(u(t)) = \{H, E\}(u(t)). \tag{1.33}$$

Let us say that a quantity E is *conserved* by (1.28) if $E(u(t))$ is constant for any (classical) solution $u : I \to \mathcal{D}$ of (1.28). From (1.33) and the anti-symmetry $\{H, E\} = -\{E, H\}$ of the Poisson bracket, we conclude

THEOREM 1.29 (Noether's theorem). *Let H and E be two Hamiltonians on a symplectic phase space $(\mathcal{D}, \omega)$. Then the following are equivalent.*

(a) $\{H, E\} = 0$.
(b) *The quantity E is conserved by the Hamiltonian flow of H.*
(c) *The quantity H is conserved by the Hamiltonian flow of E.*

If any of the above three properties hold, we say that H and E *Poisson commute*. As stated, Noether's theorem is symmetric in H and E. However, this theorem is often interpreted in a somewhat asymmetric way. Assume for sake of argument that the flow maps $S_E(t)$ of E are globally defined (this is the case, for instance, if E is quadratic growth). We view the flow maps $S_E(t)$ as a one-dimensional *group action* on the phase space $\mathcal{D}$. Noether's theorem then asserts that E is a conserved quantity for the equation (1.28) if and only if H is *symmetric* (i.e. invariant) with respect to the group actions $S_E(t)$; for a generalisation to higher-dimensional group actions, see Exercise 1.34. Thus this theorem gives a very satisfactory link between the symmetries of the Hamiltonian H to the conserved quantities of the flow (1.28). The larger the group of symmetries, the more conserved quantities one obtains[14].

For instance, since H clearly Poisson commutes with itself, we see that H itself is a conserved quantity, thus $H(u(t_0)) = H(u_0)$ for any classical solution $u \in C^1_{\mathrm{loc}}(I \to \mathcal{D})$ to the Cauchy problem

$$\partial_t u(t) = \nabla_\omega H(u(t)); \quad u(t_0) = u_0. \tag{1.34}$$

As another example, if $(\mathcal{D}, \omega)$ is the complex phase space given in Example 1.28, and the Hamiltonian is invariant under phase rotations, thus

$$H(e^{i\theta} z_1, \ldots, e^{i\theta} z_n) = H(z_1, \ldots, z_n) \text{ for all } z_1, \ldots, z_n \in \mathbf{C}, \theta \in \mathbf{R}$$

then the total charge $\sum_{j=1}^n |z_j|^2$ is conserved by the flow. Indeed, the phase rotation is (up to a factor of two) nothing more than the Hamiltonian flow associated to the total charge.

Another class of important examples concerns the phase space $(\mathbf{R}^d \times \mathbf{R}^d)^N$ of N particles in $\mathbf{R}^d$, parameterised by N position variables $q_1, \ldots, q_N \in \mathbf{R}^d$ and N

---

[13] This equation is unrelated to the PDE $\Delta u = \rho$, which is sometimes also referred to as Poisson's equation.

[14] Provided, of course, that the symmetries themselves come from Hamiltonian flows. Certain symmetries, notably scaling symmetries, are thus difficult to place in this framework, as they typically violate Louville's theorem and thus cannot be Hamiltonian flows, though they do tend to generate *almost conserved quantities*, such as that arising in the virial identity. Also, discrete symmetries such as time reversal symmetry or permutation symmetry also are not in the range of applicability for Noether's theorem.

Table 1. Some common symmetry groups and their associated conservation laws (or approximate conservation laws, in the case of the virial identity). Not all of these follow directly from Noether's theorem as stated, and are best viewed instead using the "Lagrangian" approach to this theorem. In some cases, the interpretation of the conserved quantity depends on the equation; for instance spatial translation corresponds to momentum for wave and Schrödinger equations, but corresponds instead to mass for KdV type equations.

Symmetry	Conserved quantity
time translation	energy / Hamiltonian
spatial translation	momentum / mass
spatial rotation	angular momentum
Galilean transformation	(renormalised) centre-of-mass
Lorentz transformation	(renormalised) centre-of-energy
scaling	(virial identity)
base space diffeomorphism	stress-energy
phase rotation	mass / probability / charge
gauge transform	charge

momentum variables $p_1, \ldots, p_N \in \mathbf{R}^d$, with the symplectic form

$$\omega := \sum_{j=1}^N dq_j \wedge dp_j = \sum_{j=1}^N \sum_{i=1}^d dq_{j,i} \wedge dp_{j,i}.$$

If a Hamiltonian $H(q_1, \ldots, q_N, p_1, \ldots, p_N)$ is invariant under spatial translations, thus

$$H(q_1 - x, \ldots, q_N - x, p_1, \ldots, p_N) = H(q_1, \ldots, q_N, p_1, \ldots, p_N)$$

for all $x, p_1, \ldots, p_N, q_1, \ldots, q_N \in \mathbf{R}^d$, then Noether's theorem implies that the total momentum $p = \sum_{j=1}^N p_j$ is conserved by the flow. If the Hamiltonian takes the form

$$H(q_1, \ldots, q_N, p_1, \ldots, p_N) = \sum_{j=1}^N \frac{1}{2} m_i |p_i|^2 + V(q_1, \ldots, q_N) \tag{1.35}$$

for some (translation invariant) potential $V \in C^2_{\mathrm{loc}}((\mathbf{R}^d)^N \to \mathbf{R})$, then the total momentum takes the familiar form

$$p = \sum_{j=1}^N m_j \frac{dq_j}{dt}.$$

Similarly, if the Hamiltonian is invariant under angular rotations $U : \mathbf{R}^d \to \mathbf{R}^d$, thus

$$H(Uq_1, \ldots, Uq_N, Up_1, \ldots, Up_N) = H(q_1, \ldots, q_N, p_1, \ldots, p_N)$$

for all $p_1, \ldots, p_N, q_1, \ldots, q_N \in \mathbf{R}^d$ and $U \in SO(d)$, then Noether's theorem (or more precisely the generalisation in Exercise 1.34) implies that the angular momentum $L := \sum_{j=1}^N q_j \wedge p_j \in \bigwedge^2 \mathbf{R}^d$ is also preserved by the flow.

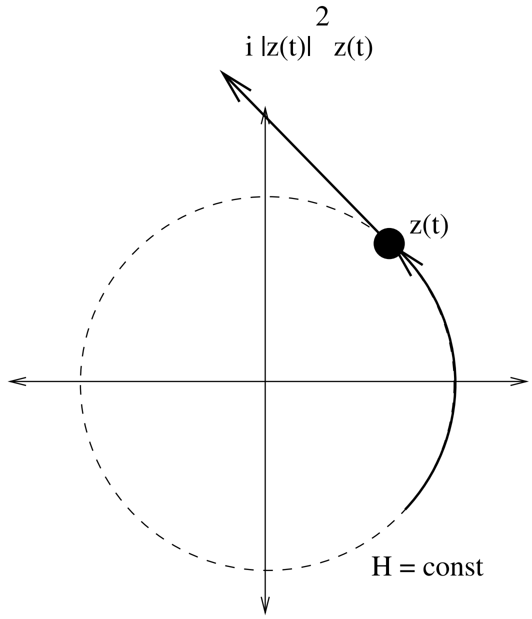

FIGURE 9. The complex scalar ODE $\dot{z} = i|z|^2 z$ is a Hamiltonian ODE with the conserved Hamiltonian $H(z) := \frac{1}{4}|z|^4$. This conservation law coerces the solution z to stay inside a bounded domain, and hence blowup does not occur. This is in contrast with the similar-looking ODE $\dot{z} = +|z|^2 z$, which blows up in finite time from any non-zero initial datum. Note also the rotation symmetry of this equation, which by Noether's theorem implies conservation of $|z|^2$.

REMARK 1.30. Noether's theorem connects exact (Hamiltonian) symmetries with exact conservation laws. There are a number of generalisations (both rigorous and informal) to this theorem. In particular, we expect *approximate* or *non-Hamiltonian* symmetries to be related to *approximate* conservation laws. One important instance of this heuristic involves *conformal Killing vector fields*, which can be viewed as approximate symmetries of the underlying geometry; see Section 2.5.

EXERCISE 1.27. Let $\mathcal{D}$ be a real Hilbert space, and let $J \in \mathrm{End}(\mathcal{D})$ be a linear map such that $J^2 = -\mathrm{id}$. Show that the bilinear form $\omega : \mathcal{D} \times \mathcal{D} \to \mathbf{R}$ defined by $\omega(u, v) := \langle u, Jv \rangle$ is a symplectic form, and that $\nabla_\omega H = -J \nabla H$ (where ∇ is the gradient with respect to the Hilbert space structure). This is the constant-coefficient version of a more general fact, that a symplectic form can be combined with an *almost complex structure* J to produce a Riemannian metric; this fact is fundamental to the theory of symplectic topology, which is far beyond the scope of this text (though see Section 4.3).

EXERCISE 1.28 (Linear Darboux theorem). Show that any symplectic space $(\mathcal{D}, \omega)$ is equivalent, after a linear change of variables, to the standard symplectic space in Example 1.27; in particular symplectic spaces are always even-dimensional.

(Hint: induct on the dimension of $\mathcal{D}$. If the dimension is non-zero, use the non-degeneracy of ω to locate two linearly independent vectors $u, v \in \mathcal{D}$ such that $\omega(u, v) \neq 0$. Then restrict to the symplectic complement $\{w \in \mathcal{D} : \omega(u, w) = \omega(v, w) = 0\}$ and use the induction hypothesis.) Note that this change of variables will usually not be unique. Conclude in particular that every symplectic phase space has an even number of dimensions.

EXERCISE 1.29. Show that if $H \in C^2_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$ is a Hamiltonian which has a non-degenerate local minimum at some $u_0 \in \mathcal{D}$ (thus $\nabla H(u_0) = 0$ and $\nabla^2 H(u_0)$ is strictly positive definite), then one has global solutions to the associated Hamiltonian equation as soon as the initial datum $\tilde{u}_0$ is sufficiently close to u_0. Note that this generalises Proposition 1.24; indeed, one can proceed by a modification of the proof of that proposition. Similarly, show that if H is a Hamiltonian which is *globally coercive* in the sense that $\lim_{v \to \infty} |H(v)| = \infty$, then one has global solutions to the associated Hamiltonian equation for arbitrary initial data.

EXERCISE 1.30. Show that if one applies the time reversal change of variable $t \mapsto -t$ to a Hamiltonian equation, one obtains another Hamiltonian equation; what is the new Hamiltonian?

EXERCISE 1.31. Let I be a time interval, and let $(\mathcal{D}, \omega)$, $(\mathcal{D}', \omega')$ be symplectic vector spaces. Let $u \in C^1_{\mathrm{loc}}(I \to \mathcal{D})$ solve a Hamiltonian equation $\partial_t u(t) = \nabla_\omega H(u(t))$ for some Hamiltonian $H \in C^2_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$, and let $u' \in C^1_{\mathrm{loc}}(I \to \mathcal{D}')$ solve a Hamiltonian equation $\partial_t u'(t) = \nabla_{\omega'} H'(u'(t))$ for some Hamiltonian $H' \in C^2_{\mathrm{loc}}(\mathcal{D}' \to \mathbf{R})$. Show that the combined system $(u, u') \in C^1_{\mathrm{loc}}(I \to \mathcal{D} \times \mathcal{D}')$ solves a Hamiltonian equation on $\mathcal{D} \times \mathcal{D}'$, with an appropriate symplectic form $\omega \oplus \omega'$ and a Hamiltonian $H \oplus H'$. This shows that a system of many non-interacting particles is automatically Hamiltonian if each component particle evolves in a Hamiltonian manner.

EXERCISE 1.32 (Preservation of symplectic form). Let $(\mathcal{D}, \omega)$ be a symplectic space, let $H \in C^2_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$ be a Hamiltonian, and let $u \in C^2_{\mathrm{loc}}(\mathbf{R} \times \mathbf{R} \times \mathbf{R} \to \mathcal{D})$ be such that for each $x, y \in \mathbf{R}$, the function $t \mapsto u(t, x, y)$ solves the Hamiltonian equation $\partial_t u(t, x, y) = \nabla_\omega H(u(t, x, y))$. Show that for each $x, y \in \mathbf{R}$, the quantity $\omega(\partial_x u(t, x, y), \partial_y u(t, x, y))$ is conserved in time. Conclude in the quadratic growth case (with $\nabla^2 H$ bounded) that the solution maps $S(t)$ are symplectomorphisms (they preserve the symplectic form ω). There is a local converse, namely that any smooth one-parameter group of symplectomorphisms must locally arise from a Hamiltonian, but we will not detail this here.

EXERCISE 1.33 (Liouville's theorem). Let $(\mathcal{D}, \omega)$ be a symplectic space, and let dm be a Haar measure on $\mathcal{D}$. (One can define a canonical Haar measure, namely *Louville measure*, by setting $m := \omega^{\dim(D)/2}$.) Let $H \in C^2_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$ be a Hamiltonian, and let Ω be any open bounded set in $\mathcal{D}$, thus we have a solution map $S(t) \in C^0_{\mathrm{loc}}(\Omega \to \mathcal{D})$ for any sufficiently small t. Show that $S(t)(\Omega)$ has the same m-measure as Ω. (Hint: use Exercise 1.32.) More generally, show that the previous claim is true if we replace dm by the (non-normalised) *Gibbs measure* $d\mu_\beta := e^{-\beta H} dm$ for any $\beta \in \mathbf{R}$. This constructs for us a small family of invariant measures for the Hamiltonian flow; a major (and rather difficult) problem in the field is to construct similar invariant measures for Hamiltonian PDE, and to investigate to what extent these are the only invariant measures available. See for instance [**Kuk3**], [**Bou4**].

EXERCISE 1.34 (Moment maps). Let G be a finite-dimensional Lie group acting (on the left) on a symplectic phase space $(\mathcal{D}, \omega)$, let $\mathfrak{g}$ be the Lie algebra and let $\mathfrak{g}^*$ be the dual Lie algebra. We identify each Lie algebra element $x \in \mathfrak{g}$ with a vector field X_x on $\mathcal{D}$ in the obvious manner. Suppose we have a *moment map* $\Phi \in C^2_{\mathrm{loc}}(\mathcal{D} \to \mathfrak{g}^*)$, in other words a map with the property that

$$X_x(u) = \nabla_\omega \langle x, \Phi(u) \rangle \text{ for all } u \in \mathcal{D}, x \in \mathfrak{g}.$$

(For instance, if G is the additive real line $\mathbf{R}$, then the group action is simply the Hamiltonian flow maps $S(t)$ associated to the Hamiltonian Φ.) Show that if $H \in C^2_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$ is a Hamiltonian which is G-invariant (thus $H(gu) = H(u)$ for all $u \in \mathcal{D}, g \in G$), then Φ is conserved by the Hamiltonian flow of H. Show that the converse is also true if G is connected. Use this generalisation of Noether's theorem to verify the claims concerning conservation of momentum and angular momentum made above.

EXERCISE 1.35. If $H_1, H_2, H_3 \in C^2_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$ are three Hamiltonians, verify the *Jacobi identity* $\{H_1, \{H_2, H_3\}\} + \{H_2, \{H_3, H_1\}\} + \{H_3, \{H_1, H_2\}\} = 0$ and the *Leibnitz rule*

$$\{H_1, H_2 H_3\} = \{H_1, H_2\} H_3 + H_2 \{H_1, H_3\} \tag{1.36}$$

EXERCISE 1.36 (Poisson bracket vs. Lie bracket). If $H \in C^1_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$, let D_H denote the differential operator defined by $D_H E := \{H, E\}$ for $E \in C^\infty_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$. Thus for instance, (1.33) asserts that $\frac{d}{dt} E(u) = D_H E(u)$ for all $E \in C^\infty_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$ and all solutions u to (1.28). Show that for any Hamiltonians $H_1, H_2 \in C^2_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$, that $[D_{H_1}, D_{H_2}] = D_{\{H_1, H_2\}}$, where $[A, B] := AB - BA$ denotes the *Lie bracket* of A and B. This is the *classical* relationship between the Poisson and Lie brackets; see Exercise 2.15 for some discussion of the *quantum* relationship between these brackets also.

EXERCISE 1.37. A function $E \in C^1_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$ is said to be an *integral of motion* of an ODE $\partial_t u(t) = F(u(t))$ if there is a function $G : C^0_{\mathrm{loc}}(\mathcal{D} \to \mathcal{D}^*)$ assigning a linear functional $G(u) \in \mathrm{End}(\mathcal{D} \to \mathbf{R})$ to each $u \in \mathcal{D}$, such that we have the identity

$$E(u(t_1)) - E(u(t_0)) = \int_{t_0}^{t_1} G(u)(\partial_t u(t) - F(u(t)))\, dt$$

for all time intervals $[t_0, t_1]$ and all functions $u \in C^1_{\mathrm{loc}}([t_0, t_1] \to \mathcal{D})$ (which may or may not solve the ODE). Show that a Hamiltonian function E is an integral of motion for a Hamiltonian ODE $\partial_t u(t) = \nabla_\omega H(u(t))$ if and only if E Poisson commutes with H.

EXERCISE 1.38. Let $H \in C^2_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$ be a Hamiltonian. Show that the space of all Hamiltonians E which Poisson commute with H form an algebra (thus the space is a vector space and is also closed under pointwise multiplication), and is also closed under all change of variable maps $E \mapsto \Phi \circ E$ for any $\Phi \in C^2_{\mathrm{loc}}(\mathbf{R} \to \mathbf{R})$. (In fact, these claims are valid for the space of integrals of motion for any first-order ODE, not just the Hamiltonian ones.)

EXERCISE 1.39. Let $H, E \in C^2_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$ be two quadratic growth Hamiltonian functions (so $\nabla^2 H$, $\nabla^2 E$ are bounded), and let $S_H(t)$ and $S_E(s)$ be the associated flow maps for $t, s \in \mathbf{R}$. Show that H, E Poisson commute if and only if $S_H(t)$ and $S_E(s)$ commute for all $t, s \in \mathbf{R}$. (Hint: use Exercise 1.14.)

EXERCISE 1.40. Let $H \in C^2_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$. Show that the flow maps $S_H(t) : \mathcal{D} \to \mathcal{D}$ are linear for all times t if and only if H is a quadratic form, plus a constant.

EXERCISE 1.41 (Symplectic normal forms). Let $(\mathcal{D}, \omega)$ be a $2n$-dimensional symplectic vector space, and let $H : \mathcal{D} \to \mathbf{R}^+$ be a positive definite quadratic form on $\mathcal{D}$. Show that there exists real numbers $\lambda_1 \geq \ldots \geq \lambda_n > 0$ and linear coordinate functions $z_1, \ldots, z_n \in \mathrm{End}(\mathcal{D} \to \mathbf{C})$ such that ω takes the form (1.31) and H takes the form (1.32). (Hint: choose a real coordinate system on $\mathcal{D}$ (identifying it with $\mathbf{R}^{2n}$) so that H is just the standard Euclidean form $H(x) = |x|^2$. Then the symplectic form is given by $\omega(x, y) = x \cdot Jy$ for some anti-symmetric non-degenerate real-valued $2n \times 2n$ matrix J. Analyze the eigenspaces and eigenvalues of J and use this to construct the complex coordinates $z_1, \ldots, z_d$.) Conclude in particular that the ellipsoid $\{z \in \mathcal{D} : H(z) = 1\}$ contains n periodic orbits for the Hamiltonian flow with periods $2\pi/\lambda_1, \ldots, 2\pi/\lambda_n$ respectively. We refer to $\lambda_1, \ldots, \lambda_n$ as the *frequencies* of the Hamiltonian H. One can devise analogues of this transformation for more general Hamiltonians (which contain higher order terms in addition to a quadratic component), leading to the theory of *Birkhoff normal forms*, which we will not discuss here.

EXERCISE 1.42. Let $(\mathcal{D}, \omega)$ be a symplectic space, let $H \in C^1_{\mathrm{loc}}(\mathbf{R} \times \mathcal{D} \to \mathbf{R})$, and consider the time-varying Hamiltonian equation

$$\partial_t u(t) = \nabla_\omega H(t, u(t)).$$

Show that it is possible to convert this time-varying Hamiltonian equation into a time-independent equation on a symplectic vector space $\mathbf{R}^2 \times \mathcal{D}$, by a trick similar to that mentioned in Section 1.1. Does this equation preserve the Hamiltonian $H(t, u(t))$? If not, is there some substitute quantity which is preserved?

EXERCISE 1.43. Let $\mathcal{D} = (\mathbf{R}^d \times \mathbf{R}^d)^N$ be the phase space of N particles in d dimensions. Suppose that a Hamiltonian equation is invariant under the Galilean symmetry

$$(q_1, \ldots, q_N, p_1, \ldots, p_N) \mapsto (q_1 - vt, \ldots, q_N - vt, p_1 - m_1 v, \ldots, p_N - m_N v)$$

for any $v \in \mathbf{R}^d$ and some fixed $m_1, \ldots, m_N > 0$, in the sense that whenever the function

$$t \mapsto (q_1(t), \ldots, q_N(t), p_1(t), \ldots, p_N(t))$$

solves the Hamiltonian ODE, then so does the transformed function

$$t \mapsto (q_1 - vt, \ldots, q_N - vt, p_1 - m_1 v, \ldots, p_N - m_N v).$$

Conclude that the normalised centre of mass

$$\sum_{j=1}^N m_j q_j - t \sum_{j=1}^N p_j$$

is an invariant of the flow. (Hint: convert t into another phase space variable as in Exercise 1.42, so that Noether's theorem can be applied.)

EXERCISE 1.44 (Connection between Hamiltonian and Lagrangian mechanics, I). Let $(\mathcal{D}, \omega)$ be the standard symplectic phase space in Example 1.27, and let

$L \in C^\infty_{\mathrm{loc}}(\mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R})$; we use $q_1, \ldots, q_n, \dot q_1, \ldots, \dot q_n$ to denote the variables of L. Define the momentum coordinates $p_1, \ldots, p_n$ by

$$p_j := \frac{\partial L}{\partial \dot q_j}(q, \dot q) \tag{1.37}$$

and assume that the coordinates $(q_1, \ldots, q_n, \dot q_1, \ldots, \dot q_n)$ and $(q_1, \ldots, q_n, p_1, \ldots, p_n)$ are diffeomorphic to each other. We then define the Hamiltonian $H \in C^\infty_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$ by

$$H(q, p) := \dot q \cdot p - L(q, \dot q)$$

where $\dot q$ is defined implicitly by (1.37). Show that if I is a bounded interval and $q \in C^\infty(I \to \mathbf{R}^n)$, then q is a formal critical point for the Lagrangian

$$S(q) := \int_I L(q(t), \partial_t q(t))\ dt$$

with endpoints held fixed, if and only if $(q(t), p(t))$ solves the Hamiltonian ODE (1.29). (You may wish to first work with the $n = 1$ case and with the example (1.30) to get some intuition.)

Exercise 1.45 (Connection between Hamiltonian and Lagrangian mechanics, II). Let $(\mathcal{D}, \omega)$ be the standard symplectic phase space in Example 1.27, and let $H \in C^\infty_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$ be a Hamiltonian phase function. Let I be a bounded time interval. Show that if $q, p \in C^\infty(I \to \mathbf{R}^n)$ obey the constraint

$$\partial_t q_j(t) = \frac{\partial H}{\partial p_j}(q(t), p(t)) \tag{1.38}$$

(which can be viewed as an implicit definition of the momentum $p(t)$ in terms of the position $q(t)$ and the velocity $\partial_t q(t)$, at least if H is sufficiently non-degenerate), then q and p obey the Hamiltonian ODE (1.29), if and only if q and p are formal critical points of the Lagrangian

$$S(q, p) := \int_I (\partial_t q(t)) \cdot p(t) - H(q(t), p(t))\ dt$$

subject to the constraint (1.38) and also fixing the values of $q(t)$ and $p(t)$ at the endpoints. Explain why this connection is essentially the inverse of that in the preceding exercise.

1.5. Monotonicity formulae

If something cannot go on forever, it will stop. (Herbert Stein)

As we have already seen, conservation laws (such as conservation of the Hamiltonian) can be very useful for obtaining long-time existence and bounds for solutions to ODE. A very useful variant of a conservation law is that of a *monotonicity formula* - a quantity $G(u(t), t)$ depending on the solution $u(t)$, and perhaps on the time t, which is always monotone increasing in time t, or perhaps monotone decreasing in time t. These monotone quantities can be used to obtain long-time control of a solution in several ways. For instance, if a quantity $G(u(t), t)$ is large at some initial time t_0 and is monotone increasing, then clearly it will stay large for all later times $t > t_0$; conversely, if $G(u(t), t)$ is bounded at time t_0, is monotone decreasing, and is manifestly non-negative, then it will stay bounded for all later times $t > t_0$. If G is monotone increasing, and is itself the time derivative of another quantity $K(t)$, then we also learn that $K(t)$ is convex in time, which can be useful in a number of ways. Finally, if one knows that $G(u(t), t)$ is bounded uniformly in time (e.g. by using conservation laws), and is monotone, then we conclude from the fundamental theorem of calculus that the derivative $\partial_t G(u(t), t)$ is absolutely integrable in time, and thus decays to zero as $t \to \pm\infty$, at least in some averaged sense. This type of long-time decay is especially useful for understanding the asymptotic behaviour of the solution.

We will be most interested in monotonicity formulae in the setting of PDE. However, we can present some simple ODE analogues of some of the more common monotonicity formulae here, which may help motivate the otherwise miraculous-seeming formulae which we will encounter in later chapters.

Unlike conservation laws, which can be systematically generated from symmetries via Noether's theorem, we do not have a fully automated way for producing monotone or convex quantities other than trial and error (for instance by starting with a conserved quantity such as energy or momentum and perturbing it somehow to be monotone instead of conserved), although certain tactics (e.g. exploiting conformal Killing fields, see Section 2.5) have proven to be generally quite fruitful. Thus we shall content ourselves in this section by presenting some illustrative examples of monotonicity formulae for ODE, each of which has an extension to a PDE such as the nonlinear Schrödinger equation.

EXAMPLE 1.31 (Virial identity). Let $V \in C^2_{\mathrm{loc}}(\mathbf{R}^d \to \mathbf{R})$ be a twice continuously differentiable potential, and consider a classical solution $x : I \to \mathbf{R}^d$ to Newton's equations of motion

$$\partial_t^2 x(t) = -\nabla V(x(t)). \tag{1.39}$$

Then we observe the *virial identity*

$$\begin{aligned} \partial_t^2(|x(t)|^2) &= 2\partial_t(x(t) \cdot \partial_t x(t)) \\ &= 2|\partial_t x(t)|^2 + 2x(t) \cdot \partial_t^2 x(t) \\ &= 2|\partial_t x(t)|^2 - 2x(t) \cdot \nabla V(x(t)). \end{aligned}$$

This has a number of consequences. If V is radially decreasing, so that $x \cdot \nabla V(x) \leq 0$ for all $x \in \mathbf{R}^d$, then we thus conclude that $|x(t)|^2$ is convex. If instead we have a bound of the form

$$x \cdot \nabla V(x) \leq -CV(x)$$

for some $C \geq 2$, then we can obtain the lower bound

$$\partial_t^2(|x(t)|^2) \geq 2CE$$

where E is the conserved energy

$$E = E(t) = \frac{1}{2}|\partial_t x(t)|^2 + V(x(t)). \tag{1.40}$$

Thus $|x(t)|^2$ is now strictly convex when the energy is positive. At a heuristic level, we thus see that positive energy tends to repel solutions from the origin, whereas negative energy tends to focus solutions towards the origin. For another application, see Exercise 1.48. For the linear and nonlinear Schrödinger analogues of these estimates, see (2.38), (3.72).

Example 1.32 (Morawetz identity). We continue the previous example. A variant of the virial identity is the *Morawetz identity*

$$\begin{aligned}\partial_t^2|x(t)| &= \partial_t\Big(\frac{x(t)}{|x(t)|}\cdot\partial_t x(t)\Big)\\ &= \frac{|\partial_t x(t)|^2}{|x(t)|} - \frac{(x(t)\cdot\partial_t x(t))^2}{|x(t)|^3} + \frac{x(t)}{|x(t)|}\cdot\partial_t^2 x(t)\\ &= \frac{|\pi_{x(t)}(\partial_t x(t))|^2}{|x(t)|} - \frac{x(t)\cdot\nabla V(x(t))}{|x(t)|}\end{aligned}$$

whenever $x(t)\neq 0$, where $\pi_x(v) := v - \frac{x}{|x|}\langle\frac{x}{|x|}, v\rangle$ is the projection of a vector v to the orthogonal complement of x. Now suppose that V is radially decreasing and non-negative, then the above identity shows that the quantity $\frac{x(t)}{|x(t)|}\cdot\partial_t x(t)$, which measures the radial component of the velocity, is monotone increasing (and that $|x(t)|$ is convex). This is intuitively plausible; particles that move towards the origin must eventually move away from the origin, but not vice versa, if the potential is repulsive. On the other hand, we have

$$\Big|\frac{x(t)}{|x(t)|}\cdot\partial_t x(t)\Big| \leq |\partial_t x(t)| \leq \sqrt{2E}$$

where the energy E is defined in (1.40). From the fundamental theorem of calculus, we thus conclude that

$$\int_I \frac{|\pi_{x(t)}(\partial_t x(t))|^2}{|x(t)|}\,dt + \int_I \frac{-x(t)\cdot\nabla V(x(t))}{|x(t)|}\,dt \leq 2\sqrt{2E}, \tag{1.41}$$

provided that x does not pass through the origin in the time interval I. (This latter hypothesis can be removed by limiting arguments; see Exercise 1.46.) If $I = \mathbf{R}$, this estimate is asserting in particular a certain decay for the angular component $\pi_{x(t)}(\partial_t x(t))$ of the velocity; that particles following this law of motion must eventually move in mostly radial directions. For the linear and nonlinear Schrödinger analogues of these estimates, see (2.40), (3.37).

Example 1.33 (Local smoothing). Again continuing the previous example, one can obtain a smoother analogue of the Morawetz inequality by replacing the non-differentiable function $|x|$ by the smoother function $\langle x\rangle := (1+|x|^2)^{1/2}$. One then has[15]

$$\begin{aligned}\partial_t^2\langle x(t)\rangle &= \partial_t\Big(\frac{x(t)}{\langle x(t)\rangle}\cdot\partial_t x(t)\Big)\\ &= \frac{|\partial_t x(t)|^2}{\langle x(t)\rangle} - \frac{(x(t)\cdot\partial_t x(t))^2}{\langle x(t)\rangle^3} + \frac{x(t)}{\langle x(t)\rangle}\cdot\partial_t^2 x(t)\\ &= \frac{|\partial_t x(t)|^2}{\langle x(t)\rangle^3} + \frac{|x(t)|^2|\partial_t x(t)|^2 - (x(t)\cdot\partial_t x(t))^2}{\langle x(t)\rangle^3} - \frac{x(t)\cdot\nabla V(x(t))}{\langle x(t)\rangle}.\end{aligned}$$

This time there is no need to exclude the case when $x(t) = 0$. In particular, if V is radially decreasing and non-negative, we conclude that

$$\partial_t\Big(\frac{x(t)}{\langle x(t)\rangle}\cdot\partial_t x(t)\Big) \geq \frac{|\partial_t x(t)|^2}{\langle x(t)\rangle^3}$$

[15]One can in fact deduce this new identity from the previous one by adding an extra dimension to the state space $\mathbf{R}^d$, and replacing x by $(x,1)$; we omit the details.

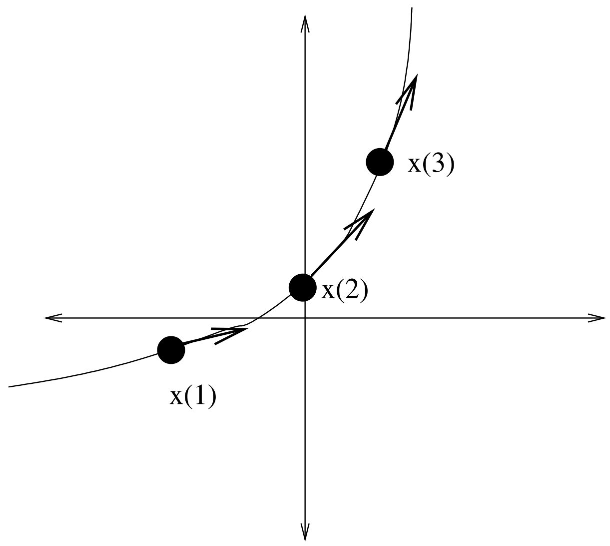

FIGURE 10. A particle passing by the origin, encountering a repulsive force, will convert its ingoing momentum to outgoing momentum. Since there is no way to convert outgoing momentum back to ingoing momentum, we conclude that if the total energy (and hence momentum) is bounded, then the particle cannot move past the origin for extended periods of time. Note that this diagram is slightly different from the one in Figure 1 because the equation is second-order rather than first-order in time; the position controls the acceleration rather than the velocity.

and hence by using the fundamental theorem of calculus we obtain the *local smoothing estimate*

$$\int_I \frac{|\partial_t x(t)|^2}{\langle x(t)\rangle^3}\, dt \leq CE^{1/2}$$

for some absolute constant $C > 0$. This result is perhaps a little surprising, since $E^{1/2}$ only seems to control the speed $|\partial_t x(t)|$, as opposed to the square of the speed $|\partial_t x(t)|^2$. Intuitively, the reason for this is the localisation factor $\frac{1}{\langle x(t)\rangle^3}$, combined with the integration in time. When the particle $x(t)$ is travelling at very high speeds, then $|\partial_t x(t)|^2$ is much larger than $|\partial_t x(t)|$, but to compensate for this, the particle only lives near the origin (where the localisation factor $\frac{1}{\langle x(t)\rangle^3}$ is large) for a brief time. In Section 2.4, we shall quantise this estimate to Schrödinger equations; the ability to upgrade the speed to the square of the speed will become a smoothing effect of half a derivative, which may help explain the terminology "local smoothing".

EXAMPLE 1.34 (Interaction Morawetz). Consider an N-particle system of non-interacting particles with masses $m_1, \ldots, m_N > 0$, with the classical solution $x : I \to (\mathbf{R}^d)^N$ given by Newton first law

$$m_i \partial_t^2 x_i(t) = 0 \text{ for } i = 1, \ldots, N.$$

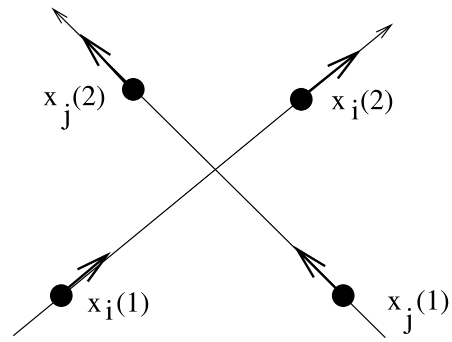

FIGURE 11. When two particles "collide" (i.e. pass through each other), their mutual ingoing momentum is converted to mutual outgoing momentum. As there is no mechanism to convert mutual outgoing momentum back into mutual ingoing momentum, we thus see that the total number of collisions (weighted by their mass and relative velocity) is controlled by the total momentum.

It is easily verified that this system has a conserved energy

$$E := \sum_i \frac{1}{2} m_i |\partial_t x_i(t)|^2$$

and one trivially also has a conserved mass $M := \sum_i m_i$. Let define a *collision* to be a triplet (i, j, t) where t is a time and $1 \leq i < j \leq N$ are indices such that $x_i(t) = x_j(t)$. Let us make the assumption that only finitely many collisions occur in the time interval I. If t is not one of the times where a collision occurs, we can define the *interaction momentum*

$$P(t) := \sum \sum_{1 \leq i < j \leq N} m_i m_j \frac{x_i(t) - x_j(t)}{|x_i(t) - x_j(t)|} \cdot \partial_t (x_i(t) - x_j(t));$$

roughly speaking, this measures how much the particles are receding from each other. A computation shows that

$$\partial_t P(t) = \sum \sum_{1 \leq i < j \leq N} m_i m_j \frac{|\pi_{x_i(t) - x_j(t)} \partial_t (x_i(t) - x_j(t))|^2}{|x_i(t) - x_j(t)|} \geq 0$$

when t is not a collision time. Each collision (i, j, t) causes a jump in $P(t)$ by $2 m_i m_j |\partial_t (x_i(t) - x_j(t))|$, thus P is monotone increasing. Using the crude bound

$$|P(t)| \leq \sum_i m_i \sum_j m_j |\partial_t x_j(t)| \leq M\sqrt{2ME}$$

from Cauchy-Schwarz, we thus conclude the *interaction Morawetz inequality*

$$\sum_{(i,j,t)} 2 m_i m_j |\partial_t (x_i(t) - x_j(t))| \leq 2M\sqrt{2ME}$$

where (i, j, t) runs over all collisions. There is a related (though not completely analogous) inequality for the nonlinear Schrödinger equation; see (3.42).

EXERCISE 1.46. Let $V : \mathbf{R}^d \to \mathbf{R}$ be twice continuously differentiable, radially decreasing (so in particular $\nabla V(0) = 0$), and non-negative, and let $u_0 \in \mathbf{R}^d$. Show that there is a unique global solution $u : \mathbf{R} \to \mathbf{R}^d$ to (1.39) with initial datum $u(0) = u_0$. Also, show that if $x(t)$ is not identically zero, then $x(t)$ can equal zero for at most one time $t_0 \in \mathbf{R}$, and in such a case we can refine (1.41) to

$$\int_{\mathbf{R}} \frac{|\pi_{x(t)}(\partial_t x(t))|^2}{|x(t)|}\ dt + \int_{\mathbf{R}} \frac{-x(t) \cdot \nabla V(x(t))}{|x(t)|}\ dt + 2|\partial_t x(t_0)| \leq 2\sqrt{2E}.$$

EXERCISE 1.47. With the same hypotheses as Exercise 1.46, show that for each $\varepsilon > 0$ we have the estimate

$$\int_{\mathbf{R}} \frac{|\partial_t x(t)|^2}{\langle x(t)\rangle^{1+\varepsilon}}\ dt \lesssim_\varepsilon E^{1/2}.$$

This improves upon Example 1.33, which dealt with the case $\varepsilon = 2$. (Hint: use the monotonicity formulae already established, as well as some new formulae obtained by considering derivatives of expressions such as $\langle x(t)\rangle^{-1-\varepsilon}(x(t) \cdot \partial_t x(t))$.) Show that the estimate fails at the endpoint $\varepsilon = 0$, even when $V = 0$ (compare this with (1.41)).

EXERCISE 1.48 (Virial identity). Suppose that $x_1, \ldots, x_N \in C^2_{\mathrm{loc}}(\mathbf{R} \to \mathbf{R}^d)$ are solutions to the system of ODE

$$m_j \partial_t^2 x_j(t) = -\sum_{i \neq j} G \frac{m_i m_j}{|x_i(t) - x_j(t)|^2}$$

where the masses $m_1, \ldots, m_N$ are positive, and $G > 0$ is an absolute constant; this models the behaviour of N particles under Newtonian gravity. Assume that the x_j are all uniformly bounded in time and that $|x_i - x_j|$ never vanishes for any $i \neq j$. Suppose also that the average kinetic and potential energies

$$\langle T\rangle := \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \sum_{j=1}^{N} \frac{1}{2} m_j |\partial_t x_j(t)|^2\ dt;$$

$$\langle V\rangle := \lim_{T\to\infty} -\frac{1}{2T} \int_{-T}^{T} \sum_{1\leq i<j\leq N} \frac{G m_i m_j}{|x_i(t) - x_j(t)|}\ dt$$

exist. Conclude the *virial identity* $\langle T\rangle = -\frac{1}{2}\langle V\rangle$. (Hint: look at Exercise 1.31.) This identity is of importance in astrophysics, as it allows one to infer the potential energy (and hence the possible existence of dark matter) from measurements of the kinetic energy.

1.6. Linear and semilinear equations

> *Mathematics would certainly have not come into existence if one had known from the beginning that there was in nature no exactly straight line, no actual circle, no absolute magnitude.* (Friedrich Nietzsche, "Human, All Too Human")

Let us now return to the general Cauchy problem (1.7). We shall assume that we have a special solution, the *vacuum solution* $u(t) \equiv 0$; in other words, we assume that $F(0) = 0$. If F is continuously differentiable, we can then perform a Taylor expansion

$$F(u) = Lu + N(u)$$

where $L \in \mathrm{End}(\mathcal{D})$ is a linear operator, and $N : \mathcal{D} \to \mathcal{D}$ vanishes faster than linearly at 0, thus

$$\lim_{\|u\|_{\mathcal{D}} \to 0} \frac{N(\|u\|_{\mathcal{D}})}{\|u\|_{\mathcal{D}}} = 0. \tag{1.42}$$

We shall refer to L as the *linear component* of F, and N as the *genuinely nonlinear component*. Thus we can write our quasilinear equation as a *semilinear equation*[16]

$$\partial_t u(t) - Lu(t) = N(u(t)). \tag{1.43}$$

If $N = 0$, we say that the equation is *linear*, otherwise it is *nonlinear*. In general, linear equations are much better understood than nonlinear equations, as a vast array of tools such as linear algebra, spectral theory, Fourier analysis, special functions (explicit solutions), and the principle of superposition can be now invoked to analyze the equation. A very profitable idea in solving equations such as (1.43) is to treat the genuine nonlinearity $N(u)$ as negligible, and thus to view the equation (1.43) as a *perturbation* of the linear equation

$$\partial_t u(t) - Lu(t) = 0. \tag{1.44}$$

This perturbation strategy is reasonable if u is small (so that $N(u)$, which vanishes to at least second order, will be very small compared to Lu) or if one is only solving the equation for very short times (so that the nonlinearity does not have much of a cumulative influence on the solution). However, when considering large solutions for long periods of time, the perturbation approach usually needs to be abandoned in favour more "global" or "non-perturbative" techniques such as energy methods or monotonicity formula methods, although if one establishes a sufficient amount of decay on the solution in time, then one can often re-instate the perturbation analysis in the asymptotic limit $t \to \infty$, which can be used for instance to obtain a scattering theory.

In accordance to this perturbation philosophy, let us study (1.43) by first considering the linear equation (1.44), say with initial datum $u(0) = u_0 \in \mathcal{D}$. (We know from time translation invariance that the choice of initial time t_0 is not particularly relevant.) Then there is a unique global solution to (1.44), given by

$$u(t) = e^{tL} u_0 = \sum_{n=0}^{\infty} \frac{t^n}{n!} L^n u_0;$$

the finite dimensionality of $\mathcal{D}$ ensures that L is bounded, and hence this sum is absolutely convergent. We refer to the linear operators $e^{tL} \in \mathrm{End}(\mathcal{D})$ as the *linear propagators*; observe that we have the group law $e^{tL} e^{sL} = e^{(t+s)L}$ with $e^{0L} = \mathrm{id}$. In particular, if u_0 is an eigenvector of L, thus $Lu_0 = \lambda u_0$ for some $\lambda \in \mathbf{C}$, then the unique global solution to (1.44) is given by $u(t) = e^{t\lambda} u_0$.

It is thus clear that the eigenvalues of L will play an important role in the evolution of the equation (1.44). If L has eigenvalues with negative real part, then the equation is *stable* or *dissipative* in the corresponding eigenspaces, displaying exponential decay in time as $t \to +\infty$ (but exponential growth as $t \to -\infty$). Conversely, if L has eigenvalues with positive real part, then the equation is *unstable*

[16]For ODE, there is little distinction between a quasilinear equation and a semilinear one. For PDE, one usually requires in a semilinear equation that the genuinely nonlinear part $N(u)$ of the equation is lower order (i.e. involves fewer spatial derivatives) than the linear part Lu; some authors require that $N(u)$ contain no derivatives whatsoever.

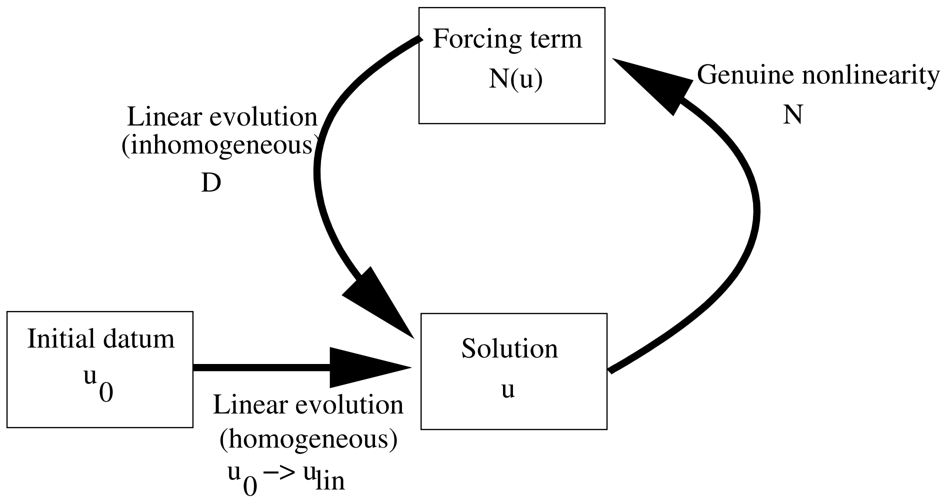

FIGURE 12. The Duhamel formulation of a semilinear ODE, relating the initial datum u_0, the solution $u(t)$, and the nonlinearity $N(u)$. Again, compare with Figure 2.

or *anti-dissipative* in the corresponding eigenspaces, exhibiting exponential growth as $t \to +\infty$. We will be concerned primarily with the *dispersive* case, in between the stable and unstable modes, in which the eigenvalues are all purely imaginary; in particular, we will usually consider the case when L is skew-adjoint with respect to a suitable Hilbert space structure on $\mathcal{D}$. In such cases, we see from the spectral theorem that there exists an orthogonal *Fourier basis* $(e_\xi)_{\xi\in\Xi}$ of $\mathcal{D}$, with each e_ξ being an eigenvector of L with some imaginary eigenvalue $ih(\xi)$:

$$Le_\xi = ih(\xi)e_\xi.$$

The function $h : \Xi \to \mathbf{R}$ will be referred to as the *dispersion relation* of L. If we then define the *Fourier coefficients*

$$\hat{f}(\xi) := \langle f, e_\xi \rangle \tag{1.45}$$

for any $f \in \mathcal{D}$ and $\xi \in \Xi$, then the solution to (1.44) can be given on the Fourier side as

$$\widehat{u(t)}(\xi) = e^{ith(\xi)}\hat{u}_0(\xi). \tag{1.46}$$

Thus each Fourier mode in (1.44) oscillates independently in time, with the time oscillation frequency given by the dispersion relation $\xi \mapsto h(\xi)$. The magnitude $|\widehat{u(t)}(\xi)|$ is conserved by the flow, so each Fourier coefficient simply moves in a circle.

In order to perturb the linear equation (1.44) to the nonlinear equation (1.43), we need the fundamental identity

PROPOSITION 1.35 (Duhamel's formula). *Let I be a time interval, let t_0 be a time in I, and let $L \in \mathrm{End}(\mathcal{D})$, $u \in C^1(I \to \mathcal{D})$, $f \in C^0(I \to \mathcal{D})$. Then we have*

$$\partial_t u(t) - Lu(t) = f(t) \textit{ for all } t \in I \tag{1.47}$$

if and only if

$$u(t) = e^{(t-t_0)L}u(t_0) + \int_{t_0}^{t} e^{(t-s)L} f(s)\ ds \text{ for all } t \in I, \tag{1.48}$$

where we adopt the convention that $\int_{t_0}^{t} = -\int_{t}^{t_0}$ *if* $t < t_0$.

REMARK 1.36. The case $L = 0$ is just the fundamental theorem of calculus. Indeed one can view Duhamel's formula as the fundamental theorem of calculus, twisted (i.e. conjugated) by the linear propagator e^{tL}; this helps explain the similarity between Figure 2 and Figure 12.

PROOF. If we make the ansatz[17] $u(t) = e^{tL}v(t)$ for some $v : I \to \mathcal{D}$ in (1.47), then (1.47) is equivalent to

$$\partial_t v(t) = e^{-tL} f(t),$$

which by the fundamental theorem of calculus is equivalent to

$$v(t) = v(t_0) + \int_{t_0}^{t} e^{-sL} f(s)\ ds.$$

The claim then follows by multiplying both sides by e^{tL} and using the group law. □

In view of this proposition, we see that if $N : \mathcal{D} \to \mathcal{D}$ is continuous and u is assumed to be continuous, then the Cauchy problem

$$\partial_t u - Lu = N(u); \quad u(0) = u_0,$$

is equivalent to the integral equation

$$u(t) = e^{tL}u_0 + \int_0^t e^{(t-s)L} N(u(s))\ ds. \tag{1.49}$$

This should be compared with the solution $u(t) = e^{tL}u_0$ of the corresponding linear problem; thus if we think of N as being small, (1.49) is a quantitative formulation of the assertion that the nonlinear solution resembles the linear solution. One can view (1.49) as the strong solution concept for u, adapted to the flow e^{tL} of the linear operator L.

The equation (1.49) is a variant of (1.8), but is a little "smarter" in that it uses the more accurate approximation $e^{tL}u_0$ to the nonlinear solution u, as opposed to the somewhat cruder approximation u_0. As a consequence, the error term in (1.49) tends to be somewhat smaller than that in (1.8), as it involves just the genuinely nonlinear component N of the nonlinearity. Just as (1.8) can be iterated using the contraction mapping theorem to obtain the Picard existence theorem, the variant (1.49) can also be iterated to obtain a variant of the Picard existence theorem, which can exploit the special properties of the linear propagator e^{tL} to give better bounds on the time of existence. To describe this iteration scheme, let us first work abstractly, viewing (1.49) as instance of the more general equation

$$u = u_{\text{lin}} + DN(u) \tag{1.50}$$

[17] This is of course the technique of integrating factors, which is a special case of the method of *variation of parameters*. The choice of ansatz $u(t) = e^{tL}v(t)$ is inspired by the fact that one solves the linear equation (1.44) if and only if v is constant.

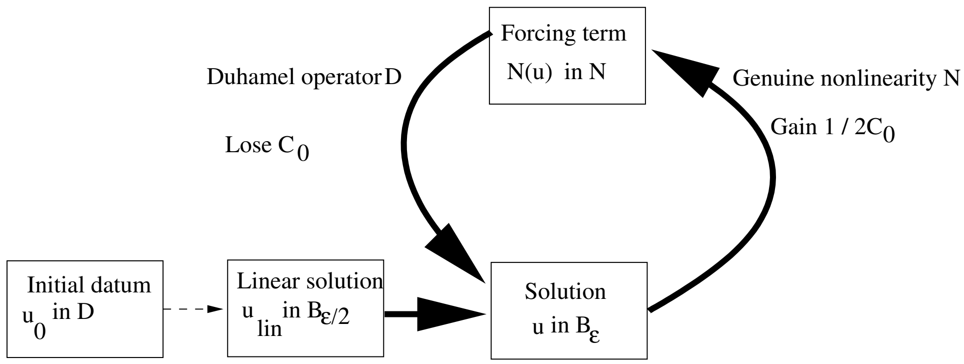

FIGURE 13. The iteration scheme for Proposition 1.38. In practice, the object $u_{\rm lin}$ arises as the linear evolution of some initial datum u_0, as in Figure 12, though we do not use this in the statement and proof of the Proposition.

where $u_{\rm lin}(t) := e^{tL}u_0$ is the linear solution, and D is the Duhamel operator

$$DF(t) := \int_0^t e^{(t-s)L}F(s)\ ds.$$

A useful heuristic principle in trying to solve equations of this general abstract type is

PRINCIPLE 1.37 (Perturbation principle). *If one is working on a time interval* $[0,T]$ *such that* $DN(u) \ll u_{\rm lin}$*, then* u *should evolve on* $[0,T]$ *as if it were linear (in particular, the solution should exist and obey the same type of estimates that* $u_{\rm lin}$ *does). If one is working instead on a time interval where* $DN(u) \gg u_{\rm lin}$*, one should expect* u *to exhibit nonlinear behaviour (which could range from blowup or excessive growth on one hand, to additional decay on the other, or something in between such as nontrivial nonlinear oscillation).*

This is of course a very vague principle, since terms such as "$\ll$", "$\gg$", or "nonlinear behaviour" are not well defined. In practice, $DN(u)$ will tend to be small compared to $u_{\rm lin}$ if the initial datum u_0 is suitably small, or if the time t is close to 0, so for small data or small times one expects linear-type behaviour. For large data or large times, perturbation theory does not predict linear behaviour, and one could now have nonlinear effects such as blowup[18]. To control solutions in this regime one generally needs to augment the perturbation theory with other tools such as conservation laws.

Let us now give a rigorous formulation of the first half of this principle, by using the following variant of the contraction mapping theorem to construct solutions.

[18]Note however that it is possible for the nonlinear term to dominate the linear term but still be able to construct and control solutions. This for instance occurs if there is an "energy cancellation" that shows that the nonlinear term, while nominally stronger than the linear term, is somehow "almost orthogonal" to the solution in the sense that it does not significantly increase certain energies; we shall see several examples of this in the text. Thus the solutions will not stay close to the linear solution but will still be bounded in various norms. In certain defocusing dissipative settings it is even possible for the nonlinearity to always act to *reduce* the energy, thus giving a *better* behaved solution than the linear equation.

PROPOSITION 1.38 (Abstract iteration argument). *Let $\mathcal{N}$, $\mathcal{S}$ be two Banach spaces. Suppose we are given a linear operator $D : \mathcal{N} \to \mathcal{S}$ with the bound*

$$\|DF\|_{\mathcal{S}} \leq C_0 \|F\|_{\mathcal{N}} \tag{1.51}$$

for all $F \in \mathcal{N}$ and some $C_0 > 0$, and suppose that we are given a nonlinear operator $N : \mathcal{S} \to \mathcal{N}$ with $N(0) = 0$, which obeys the Lipschitz bounds

$$\|N(u) - N(v)\|_{\mathcal{N}} \leq \frac{1}{2C_0} \|u - v\|_{\mathcal{S}} \tag{1.52}$$

for all u, v in the ball $B_\varepsilon := \{u \in \mathcal{S} : \|u\|_{\mathcal{S}} \leq \varepsilon\}$, for some $\varepsilon > 0$. (In other words, $\|N\|_{\dot{C}^{0,1}(B_\varepsilon \to \mathcal{N})} \leq \frac{1}{2C_0}$.) Then for all $u_{\mathrm{lin}} \in B_{\varepsilon/2}$ there exists a unique solution $u \in B_\varepsilon$ to the equation (1.50), *with the map $u_{\mathrm{lin}} \mapsto u$ Lipschitz with constant at most* 2. *In particular we have*

$$\|u\|_{\mathcal{S}} \leq 2 \|u_{\mathrm{lin}}\|_{\mathcal{S}}. \tag{1.53}$$

This proposition is established by the arguments used to prove the contraction mapping principle, and is left as an exercise. The idea of using this type of abstract Duhamel iteration to tackle nonlinear PDE dates back to [**Seg**].

REMARKS 1.39. Note that we have considerable freedom in selecting the spaces $\mathcal{S}$ and $\mathcal{N}$; this freedom becomes very important when considering the low-regularity local wellposedness theory of PDE. The Picard existence argument in Theorem 1.7 corresponds, roughly speaking, to the choice $\mathcal{S} = \mathcal{N} = C^0(I \to \mathcal{D})$, with (1.8) taking the place of (1.49). There are a number of variations of this iteration scheme; for instance, instead of measuring the solution u in a single norm $\mathcal{S}$, one sometimes is in a situation where u is measured both in a "smooth" norm $\mathcal{S}$ and a "rough" norm $\mathcal{S}_0$; the solution may be large in the smooth norm but small in the rough norm. In such cases it can still be possible to close an iteration argument using estimates that combine both norms together; this becomes important in the large data theory and in the persistence of regularity theory. While it is possible to build an abstract framework for such schemes, the formulation becomes rather complicated, and so when these types of situations arise (see for instance Proposition 3.11) we shall simply perform the iteration by hand.

REMARK 1.40. As with Remark 1.5, the proof of the above theorem provides an explicit iteration scheme to construct the desired solution, starting with the *linear iterate* $u^{(0)} := u_{\mathrm{lin}}$ and then constructing successive *Duhamel iterates* $u^{(n)} := u_{\mathrm{lin}} + DN(u^{(n-1)})$. This scheme often converges better than the one in Remark 1.5, though it is far from the most rapidly convergent scheme (and is usually not used directly in numerical computations).

We illustrate the iteration method by establishing global existence for linearly stable nonlinear equations from small data.

PROPOSITION 1.41 (Linear stability implies nonlinear stability). *Let $\mathcal{D}$ be a finite-dimensional real Hilbert space, and let $L \in \mathrm{End}(\mathcal{D})$ be a linear operator which is* linearly stable *in the sense that*[19] *there exists $\sigma > 0$ such that $\langle Lu, u \rangle \leq -\sigma \|u\|_{\mathcal{D}}^2$*

[19] In the finite-dimensional case, linear stability is equivalent to the spectrum of L being contained entirely in the interior of the left half-plane. In the infinite-dimensional case, the relationship between stability and spectrum is more delicate, especially if L fails to be normal or self-adjoint. Indeed, for PDE, nonlinear stability is often significantly more difficult to establish than linear stability.

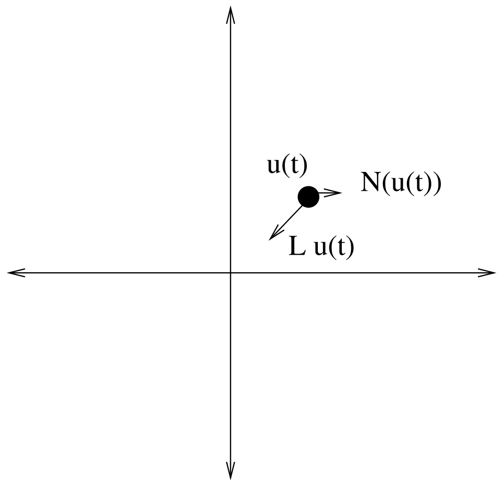

Figure 14. Proposition 1.41 from the vector field perspective of Figure 1. If u is sufficiently small, then the dissipative effect of the linear term Lu will dominate the effect of the nonlinearity $N(u)$, regardless of the orientation of $N(u)$, causing u to decay exponentially towards the origin.

for all $u \in \mathcal{D}$. *Let* $N \in C^2_{\mathrm{loc}}(\mathcal{D} \to \mathcal{D})$ *be a function which vanishes to more than first order at the origin in the sense of* (1.42) *(in fact, the* C^2 *hypothesis shows that* N *will vanish to second order). If* $u_0 \in \mathcal{D}$ *is sufficiently close to the origin, there exists a unique classical solution* $u : [0, +\infty) \to \mathcal{D}$ *to* (1.43) *with initial datum* $u(0) = u_0$*, and furthermore there is an estimate of the form*

$$\|u(t)\|_{\mathcal{D}} \leq 2e^{-\sigma t}\|u_0\|_{\mathcal{D}}. \tag{1.54}$$

Proof. The uniqueness of u follows from the Picard uniqueness theorem, so it suffices to establish existence, as well as the estimate (1.54). A simple Gronwall argument (see Exercise 1.54) gives the dissipative estimate

$$\|e^{tL}u_0\|_{\mathcal{D}} \leq e^{-\sigma t}\|u_0\|_{\mathcal{D}} \tag{1.55}$$

for all $u_0 \in \mathcal{D}$ and $t \geq 0$. Let us now define the spaces $\mathcal{S}$ to be the space of all functions $u \in C^0([0, +\infty) \to \mathcal{D})$ whose norm

$$\|u\|_{\mathcal{S}} := \sup_{t \geq 0} e^{\sigma t}\|u(t)\|_{\mathcal{D}}$$

is finite, and let $\mathcal{N}$ be the space of all functions $F \in C^0([0, +\infty) \to \mathcal{D})$ then

$$\|u\|_{\mathcal{N}} := \sup_{t \geq 0} e^{\sigma t}\|u(t)\|_{\mathcal{D}}$$

is finite. Thus if we set $u_{\mathrm{lin}}(t) := e^{tL}u_0$ then $\|u_{\mathrm{lin}}\|_{\mathcal{S}} \leq \|u_0\|_{\mathcal{D}}$. Next, observe from (1.42) that $\nabla N(0) = 0$, and so by Taylor expansion and the hypothesis that N is C^2, we have $\nabla N(u) = O_N(\|u\|_{\mathcal{D}})$ for $\|u\|_{\mathcal{D}}$ sufficiently small (the implied constant depends of course on N). From the fundamental theorem of calculus this implies that

$$\|N(u) - N(v)\|_{\mathcal{D}} \ll_N \|u - v\|_{\mathcal{D}}(\|u\|_{\mathcal{D}} + \|v\|_{\mathcal{D}})$$

whenever $\|u\|_{\mathcal{D}}, \|v\|_{\mathcal{D}}$ are sufficiently small. In particular, if $\|u\|_{\mathcal{S}}, \|v\|_{\mathcal{S}} \leq \varepsilon$ for some sufficiently small ε, then we have

$$\|N(u(t)) - N(v(t))\|_{\mathcal{D}} \ll_N e^{-\sigma t}\|u - v\|_{\mathcal{S}}\varepsilon e^{-\sigma t}$$

and hence (if ε is sufficiently small)

$$\|N(u) - N(v)\|_{\mathcal{N}} \leq \frac{1}{2}\sigma\|u - v\|_{\mathcal{S}}.$$

whenever $\|u\|_{\mathcal{S}}, \|v\|_{\mathcal{S}} \leq \varepsilon$. Also, from the triangle inequality and (1.55) we have

$$\|DF\|_{\mathcal{S}} \leq \frac{1}{\sigma}\|F\|_{\mathcal{N}}$$

where D is the Duhamel operator. From Proposition 1.38 we thus see that if $\|u_0\|_D \leq \varepsilon/2$, then can thus construct a solution u to (1.47) with $\|u\|_{\mathcal{S}} \leq 2\|u_{\text{lin}}\|_{\mathcal{S}} \leq 2\|u_0\|_{\mathcal{D}}$, and the claim follows. $\square$

Remark 1.42. If one merely assumes N to be Lipschitz rather than C^2_{loc}, then one can obtain some weaker decay estimates, for instance given any $\sigma' < \sigma$ one can obtain an estimate $\|u(t)\|_{\mathcal{D}} \leq 2e^{-\sigma' t}\|u_0\|_{\mathcal{D}}$ if $\|u_0\|_{\mathcal{D}}$ is sufficiently small depending on N and σ, σ'. This can be achieved for instance by modifying the Gronwall inequality argument in Exercise 1.54. We leave the details to the reader as an exercise. Thus for dissipative equations one often has the luxury of conceding an exponential factor in the time variable t, which is very helpful for keeping the iteration under control. When we turn to nonlinear dispersive equations, no such exponential factors are available for us to concede (although in some cases the dispersion will give us polynomial type decay in t, which can be quite useful), and we must proceed with a bit more care.

An important special case of the general equation (1.43) occurs when the genuinely nonlinear component N is *k-linear* for some $k \geq 2$, in the sense that

$$N(u) = N_k(u, u, \ldots, u)$$

where $N_k : X^k \to X$ is a function which is (real-)linear in each of the k variables. In the $k = 2$ case we call N *bilinear* or *quadratic*, in the $k = 3$ case we call N *trilinear* or *cubic*, and so forth. The condition $k \geq 2$ is essentially forced upon us by the condition (1.42). In these cases, the hypothesis (1.52) will hold for ε small provided that $\mathcal{N}$ is bounded from $\mathcal{S}$ to $\mathcal{N}$; see Exercise 1.51.

When the nonlinearity is k-linear and the linear term L is skew-adjoint, one can view the evolution (1.43) in terms of frequency interactions (using the "Fourier transform" (1.45)). We illustrate this in the $k = 2$ case $N(u) = N_2(u, u)$; to simplify the exposition, we will also assume that N_2 is not only real linear but is in fact complex linear. The situation for antilinear nonlinearities and for higher orders $k > 2$ requires some simple modifications to the discussion below which are left to the reader. If we take Fourier transforms of (1.49) (using (1.46)), we obtain

$$\widehat{u(t)}(\xi) = e^{ith(\xi)}\hat{u}_0(\xi) + \int_0^t e^{i(t-s)h(\xi)} \sum_{\xi_1, \xi_2} c_\xi^{\xi_1, \xi_2} \widehat{u(s)}(\xi_1)\widehat{u(s)}(\xi_2)\ ds$$

where $c_\xi^{\xi_1, \xi_2}$ is the *structure constant*

$$c_\xi^{\xi_1, \xi_2} := \langle N_2(e_{\xi_1}, e_{\xi_2}), e_\xi \rangle.$$

Typically, the structure constants will usually be zero; given any mode ξ, only a few pairs of modes ξ_1, ξ_2 can interact to excite that mode; a typical constraint in order for $c_\xi^{\xi_1,\xi_2}$ to be non-zero is of the form $\xi = \xi_1 + \xi_2$ (where one places some group structure on the space Ξ of frequencies). Making the renormalisation $\widehat{u(t)}(\xi) := e^{ith(\xi)} a_\xi(t)$, which is suggested to us by the variation of parameters method, we obtain an integral system of equations for the functions $(a_\xi(t))_{\xi \in \Xi}$:

$$a_\xi(t) = \hat{u}_0(\xi) + \int_0^t e^{is(h(\xi_1)+h(\xi_2)-h(\xi))} c_\xi^{\xi_1,\xi_2} a_{\xi_1}(s) a_{\xi_2}(s)\, ds. \tag{1.56}$$

Thus each $a_\xi(t)$ is initially set equal to $\hat{u}_0(\xi)$, but as time evolves, the ξ-modes $a_\xi(t)$ is influenced by the bilinear interactions of the pairs of modes $a_{\xi_1}(t)$, $a_{\xi_2}(t)$ that can excite the ξ-mode. The *resonance function* $h(\xi_1) + h(\xi_2) - h(\xi)$ plays a key role in the analysis. If this quantity is large in magnitude, then the integral in (1.56) will be highly oscillatory, and thus likely to be rather small; in this case, we say that the interaction between the modes ξ_1, ξ_2, ξ is *non-resonant*. The dominant contribution to (1.56) typically consists instead of the *resonant interactions*, in which the resonance function is zero or small. In order to obtain an iterative scheme for solving this equation (using for instance Proposition 1.38), especially at low regularities, one often has to spend some effort to control the resonant portions of interaction, either by showing that the resonant interactions are fairly rare, or by extracting some "null structure" from the structure coefficients $c_\xi^{\xi_1,\xi_2}$, which causes them to vanish whenever the resonance function vanishes. We will see some examples of this in later sections.

Exercise 1.49. Prove Proposition 1.38. (Hint: review the proof of Theorem 1.4, Theorem 1.7, and Exercise 1.2.)

Exercise 1.50 (Stability). Let the notation and hypotheses be as in Proposition 1.38. Suppose that $u_{\mathrm{lin}} \in B_{\varepsilon/2}$, and we have an approximate solution $\tilde{u} \in B_\varepsilon$ to the equation (1.50), in the sense that $\tilde{u} = u_{\mathrm{lin}} + DN(\tilde{u}) + e$ for some $e \in \mathcal{S}$. Let $u \in B_\varepsilon$ be the actual solution to (1.50) given by the above Proposition. Show that $\|\tilde{u} - u\|_{\mathcal{S}} \leq 2\|e\|_{\mathcal{S}}$. Note that this generalises the Lipschitz claim in Proposition 1.38.

Exercise 1.51. Let $\mathcal{N}, \mathcal{S}$ be Banach spaces as in Proposition 1.38, and suppose that one is given a k-linear nonlinearity $N(u) = N_k(u, \ldots, u)$, which maps $\mathcal{S}$ to $\mathcal{N}$ with the k-linear estimate

$$\|N(u_1, \ldots, u_k)\|_{\mathcal{N}} \leq C_1 \|u_1\|_{\mathcal{S}} \ldots \|u_k\|_{\mathcal{S}}$$

for all $u_1, \ldots, u_k \in \mathcal{S}$ and some constant $C_1 > 0$. Show that the hypothesis (1.52) holds for $u, v \in B_\varepsilon$ with $\varepsilon := \frac{1}{2kC_0C_1}$.

Exercise 1.52 (Second order Duhamel). Let $L \in \mathrm{End}(\mathcal{D})$. Show that the solution to the homogeneous linear second-order ODE

$$\partial_{tt} u - Lu = 0$$

with initial datum $u(0) = u_0$, $\partial_t u(0) = u_1$ is given by $u(t) = U_0(t)u_0 + U_1(t)u_1$ for some operators $U_0 : \mathbf{R} \times \mathcal{D} \to \mathcal{D}$, $U_1 : \mathbf{R} \times \mathcal{D} \to \mathcal{D}$. Show that the unique classical solution $u \in C^2_{\mathrm{loc}}(\mathcal{D} \to \mathbf{R})$ to the inhomogeneous linear second-order ODE

$$\partial_{tt} u - Lu = f$$

with initial datum $u(t_0) = u_0$, $\partial_t u(t_0) = u_1$, where $f \in C^0_{\mathrm{loc}}(\mathcal{D} \to \mathcal{D})$ and $t_0 \in \mathbf{R}$, is given by the Duhamel formula

$$u(t) = U_0(t - t_0)u_0 + U_1(t - t_0)u_1 + \int_{t_0}^{t} U_1(t - s)f(s)\ ds.$$

(Hint: convert the second-order equation to a first order one, then use Proposition 1.35.)

Exercise 1.53 (Duhamel vs. resolvents). Let $L, V, L_0 \in \mathrm{End}(\mathcal{D})$ be such that $L = L_0 + V$. Use Duhamel's formula to show that

$$e^{tL} = e^{tL_0} + \int_0^t e^{(t-s)L_0} V e^{sL}\ ds = e^{tL_0} + \int_0^t e^{(t-s)L} V e^{sL_0}\ ds.$$

If λ is a scalar such that the resolvent operators $R(\lambda) := (L - \lambda)^{-1}$ and $R_0(\lambda) := (L_0 - \lambda)^{-1}$ exist, establish the *resolvent identity*

$$R(\lambda) = R_0(\lambda) - R_0(\lambda)VR(\lambda) = R_0(\lambda) - R(\lambda)VR_0(\lambda)$$

and discuss the relationship between the above identities using the Fourier duality between t and λ.

Exercise 1.54. Let L be as in Proposition 1.41, and let u solve the equation (1.44). Use Gronwall's inequality to establish the bound $\|u(t)\|_{\mathcal{D}} \leq e^{-\sigma t}\|u(0)\|_{\mathcal{D}}$ for all $t \geq 0$. (Hint: establish a monotonicity formula for $\|u(t)\|^2_{\mathcal{D}}$, and then use the version of Gronwall's inequality mentioned in Exercise 1.7.)

Exercise 1.55 (Stable manifold). Let $\mathcal{D}$ be a finite-dimensional real Hilbert space, and let $L \in \mathrm{End}(\mathcal{D})$ be weakly linearly stable in the sense that $\langle Lu, u\rangle \leq 0$ for all $u \in \mathcal{D}$. Let $N \in \dot{C}^{0,1}_{\mathrm{loc}}(\mathcal{D} \to \mathcal{D})$ have the property that $\langle Nu, u\rangle \leq 0$ for all $u \in \mathcal{D}$. Show that for any $u_0 \in \mathcal{D}$, there exists a unique classical solution $u : [0, +\infty) \to \mathcal{D}$ to (1.43) with initial datum $u(0) = u_0$, which is uniformly bounded and obeys the estimate

$$\int_0^\infty |\langle Lu(t), u(t)\rangle|\ dt \leq \|u_0\|^2_{\mathcal{D}}.$$

Conclude in particular that if V is the subspace $V := \{u \in D : \langle Lu, u\rangle = 0\}$, that $\mathrm{dist}(u(t), V) \to 0$ as $t \to +\infty$.

1.7. Completely integrable systems

Tyger! Tyger! burning bright
In the forests of the night,
What immortal hand or eye
Could frame thy fearful symmetry?
(William Blake, "The Tyger")

We have already encountered Hamiltonian ODE in Section 1.4, which enjoy at least one conserved integral of motion, namely the Hamiltonian H itself. This constrains the Hamiltonian flow to a codimension one subspace of the symplectic phase space $\mathcal{D}$. Additional conserved integrals of motion can constrain the flow further. It turns out that the largest number of independent conserved integrals that one can have in a Hamiltonian system is half the dimension of the phase space (see

Exercise 1.56). When this occurs, we say that the system is *completely integrable*[20]; the phase space splits completely into the conserved quantities $E_1, \ldots, E_N$ (also called *action variables*), together with the dynamic variables (also called *angle variables*) induced by the N flows corresponding to $E_1, \ldots, E_N$.

EXAMPLE 1.43 (Simple harmonic oscillator). Let $\mathcal{D} = \mathbf{C}^n$ be the phase space in Example 1.28, and let H be the Hamiltonian (1.32). Then there are n independent conserved quantities

$$E_1 := |z_1|^2; \quad \ldots \quad ; E_n := |z_n|^2$$

and n angle variables $\theta_1, \ldots, \theta_n \in \mathbf{T}$, defined for most points in phase space by polar coordinates

$$z_1 = |z_1|e^{i\theta_1}; \quad \ldots \quad ; z_n = |z_n|e^{i\theta_n}.$$

Then the Hamiltonian flow in these action-angle coordinates becomes linear:

$$\partial_j E_j(t) = 0; \quad \partial_j \theta_j(t) = \lambda_j.$$

Also, observe that the Hamiltonian H is just a linear combination of the basic conserved quantities $E_1, \ldots, E_n$, which is of course consistent with the fact that H is itself conserved. More generally, any linear system (1.44) in which L is skew-adjoint will lead to a completely integrable system.

There are many ways to determine if a system is completely integrable. We shall discuss only one, the method of *Lax pairs*.

DEFINITION 1.44. Consider an ODE

$$\partial_t u(t) = F(u(t)), \tag{1.57}$$

where $F \in \dot{C}^{0,1}_{\mathrm{loc}}(\mathcal{D} \to \mathcal{D})$ and $\mathcal{D}$ is a finite-dimensional phase space. Let H be a finite-dimensional complex Hilbert space, and let $\mathrm{End}(H)$ be the space of linear maps from H to itself (for instance, if $H = \mathbf{C}^n$, then $\mathrm{End}(H)$ is essentially the ring of $n \times n$ complex matrices). A *Lax pair* for the ODE (1.57) is any pair $L, P \in C^1_{\mathrm{loc}}(\mathcal{D} \to \mathrm{End}(H))$ of functions such that we have the identity

$$\partial_t L(u(t)) = [L(u(t)), P(u(t))] \tag{1.58}$$

for all classical solutions $u : I \to \mathcal{D}$ to the ODE (1.57), or equivalently if

$$(F(u) \cdot \nabla)L(u) = [L(u), P(u)] \text{ for all } u \in \mathcal{D}.$$

Here $[A, B] := AB - BA$ denotes the usual Lie bracket of the matrices A and B.

REMARK 1.45. Geometrically, (1.58) asserts that the matrix $L(u(t))$ evolves via "infinitesimal rotations" that are "orthogonal" to $L(u(t))$. In many cases, P will take values in the Lie algebra $\mathfrak{g}$ of some Lie group G in $\mathrm{End}(H)$, and L will take values either in the Lie algebra $\mathfrak{g}$ or the Lie group G; note that the equation (1.58) is consistent with this assumption, since Lie algebras are always closed under the Lie bracket (see also Exercise 1.15).

[20]This definition unfortunately does not rigorously extend to the infinite dimensional phase spaces one encounters in PDE. Indeed, we do not yet have a fully satisfactory definition of what it means for a PDE to be completely integrable, though we can certainly identify certain very suggestive "symptoms" of complete integrability of a PDE, such as the presence of infinitely many conserved quantities, a Lax pair formulation, or the existence of explicit multisoliton solutions.

A trivial example of a Lax pair is when $L : \mathcal{D} \to M_n(\mathbf{C})$ is constant, and P is chosen to commute with L; we shall be more interested in non-trivial examples when L, and more precisely the spectrum (eigenvalues) of L, admit some genuine variation across the phase space $\mathcal{D}$. A simple example is provided by the one-dimensional harmonic oscillator

$$\partial_t u = i\omega u$$

in the phase space $\mathcal{D} = \mathbf{C}$, with $H = \mathbf{C}^2$ and Lax pair

$$L(u) := \begin{pmatrix} i(|u|^2 - \lambda) & iu^2 \\ i\bar{u}^2 & i(\lambda - |u|^2) \end{pmatrix}; \quad P(u) := \begin{pmatrix} -i\omega & 0 \\ 0 & i\omega \end{pmatrix} \tag{1.59}$$

where the *spectral parameter* λ is an arbitrary complex number. Here, L and P take values in the Lie algebra $\mathfrak{su}_2(\mathbf{C})$ of $SU_2(\mathbf{C})$, the group of 2×2 unitary matrices. The higher-dimensional harmonic oscillator in Example 1.28 can also be given a Lax pair by taking direct sums of the above example; we omit the details.

Now we show how Lax pairs lead to conserved quantities.

Proposition 1.46. *Suppose that an ODE* (1.57) *is endowed with a Lax pair* $L : \mathcal{D} \to \mathrm{End}(H)$, $P : \mathcal{D} \to \mathrm{End}(H)$. *Then for any non-negative integer k, the moment* $\mathrm{tr}(L^k)$ *is preserved by the flow* (1.57)*, as is the spectrum* $\sigma(L) := \{\lambda \in \mathbf{C} : L - \lambda$ *not invertible*$\}$.

Proof. We begin with the moments $\mathrm{tr}(L^k)$. Let $u : I \to \mathcal{D}$ solve (1.57). From the Leibnitz rule and the first trace identity

$$\mathrm{tr}(AB) = \mathrm{tr}(BA) \tag{1.60}$$

we have

$$\partial_t \mathrm{tr}(L(u(t))^k) = k\mathrm{tr}(L(u(t))^{k-1}\partial_t L(u(t))) = k\mathrm{tr}(L(u(t))^{k-1}[L(u(t)), P(u(t))]).$$

But from the second trace identity

$$\mathrm{tr}(A[B,C]) = \mathrm{tr}(B[C,A]) = \mathrm{tr}(C[A,B]) \tag{1.61}$$

(which follows easily from the first trace identity), and the obvious fact that L^{k-1} commutes with L, we obtain $\partial_t \mathrm{tr}(L(u(t))^k) = 0$ as desired.

One can conclude conservation of the spectrum $\sigma(L)$ from that of the moments by using the characteristic polynomial of L. For a more direct approach (which does not rely as much on the finite dimensionality of L), see Exercise 1.57. $\square$

The quantities $\mathrm{tr}(L), \mathrm{tr}(L^2), \ldots$ may seem like an infinite number of conserved quantities, but they are of course not all independent. For instance in the example (1.59), all the quantities $\mathrm{tr}(L^k)$ are functions of a single conserved quantity $|z|^2$. This makes the number of conserved quantities equal to the half the (real) dimension of the phase space $\mathbf{C}$, and so this equation is completely integrable.

One special case of solutions to a completely integrable system arises when the spectrum $\sigma(L)$ of the Lax operator is unexpectedly simple, for instance if L is a rank one operator. This often leads to very algebraically structured solutions such as solitary waves (solitons). For instance, in Example 1.43, the case when L is rank one corresponds to that of a single excited mode, when only one of the z_j is non-zero, which can be viewed as a rather trivial instance of a solitary wave. The more general task of reconstructing the solution given the spectral information on L (and certain supplemental "scattering data" associated to the initial datum u_0) is known as *inverse scattering* and is a very rich subject involving some beautiful analysis, algebra, and geometry. It is well outside the scope of this monograph; we refer the reader to [**HSW**] for an introduction.

We now give some non-trivial examples of completely integrable systems. The first is the *periodic Toda lattice*

$$\partial_t a_n = a_n(b_{n+1} - b_n); \quad \partial_t b_n = 2(a_n^2 - a_{n-1}^2) \tag{1.62}$$

where n ranges over a cyclic group $\mathbf{Z}/N\mathbf{Z}$, and $a_n : \mathbf{R} \to \mathbf{R}$, $b_n : \mathbf{R} \to \mathbf{R}$ are real-valued functions of time; this can be viewed as a discrete version of the periodic Korteweg-de Vries (KdV) equation. To place this lattice in Lax pair form, we let H be an N-dimensional real Hilbert space with orthonormal basis $\{e_n : n \in \mathbf{Z}/N\mathbf{Z}\}$, and for any given state $u = ((a_n, b_n))_{n\in\mathbf{Z}/N\mathbf{Z}}$ we define $L = L(u) : H \to H$ and $P = P(u) : H \to H$ on basis vectors by

$$\begin{aligned} Le_n &:= a_n e_{n+1} + b_n e_n + a_{n-1} e_{n-1} \\ Pe_n &:= a_n e_{n+1} - a_{n-1} e_{n-1}. \end{aligned}$$

One can easily verify the Lax pair equation (1.58) by testing it on basis vectors. Note that L is self-adjoint and P is skew-adjoint, which is of course consistent with (1.58). The Toda lattice enjoys N independent conserved quantities arising from L, including the trace

$$\mathrm{tr}(L) = \sum_{n\in\mathbf{Z}/N\mathbf{Z}} b_n$$

and the second moment

$$\mathrm{tr}(L^2) = \sum_{n\in\mathbf{Z}/N\mathbf{Z}} b_n^2 + 2a_n^2;$$

one may verify by hand that these quantities are indeed preserved by (1.62). The equation (1.62) is not a Hamiltonian flow using the standard symplectic form on the state space, but can be transformed into a Hamiltonian flow (with Hamiltonian $2\mathrm{tr}(L^2)$) after a change of variables, see Exercise 1.59. One can create higher order Toda flows by using higher moments of L as Hamiltonians, but we will not pursue this here.

Another example of a completely integrable system is the *periodic Ablowitz-Ladik system*

$$\partial_t F_n = i(1 - |F_n|^2)(F_{n-1} + F_{n+1}), \tag{1.63}$$

where n ranges over a cyclic group $\mathbf{Z}/N\mathbf{Z}$, and $F_n : \mathbf{R} \to \mathbf{C}$ are complex-valued functions of time with $|F_n| < 1$ for all n (this property is preserved by the flow). This is a discrete analogue of the cubic defocusing periodic nonlinear Schrödinger equation. To define a Lax pair (L, P) for this equation, we take H to be a complex Hilbert space spanned by $2N$ orthonormal basis vectors $\{v_n, w_n : n \in \mathbf{Z}/N\mathbf{Z}\}$. The Lax operator $L = L(F) : H \to H$ is then defined on basis elements by

$$\begin{aligned} Lv_n &:= \sqrt{1 - |F_n|^2}\, v_{n+1} + F_n w_n \\ Lw_{n+1} &:= -\overline{F_n} v_{n+1} + \sqrt{1 - |F_n|^2}\, w_{n+1}; \end{aligned}$$

note that L is in fact a unitary operator (a discrete analogue of a Dirac operator), with adjoint $L^* = L^{-1}$ given by

$$L^* v_{n+1} := \sqrt{1-|F_n|^2} v_n - F_n w_{n+1}$$
$$L^* w_n := \overline{F_n} v_n + \sqrt{1-|F_n|^2} w_{n+1};$$

The P operator is a little trickier to define. We first define the reflection operator $J : H \to H$ as

$$J v_n := v_n; \quad J w_n := -w_n$$

and then the diagonal operator $D = \frac{J}{8}([L,J]^2 + [L^*,J]^2)$ by

$$D v_n := \frac{\overline{F_{n-1}} F_n + F_{n-1}\overline{F_n}}{2} v_n; \quad D w_n := -\frac{\overline{F_{n-1}} F_n + F_{n-1}\overline{F_n}}{2} w_n$$

and then define P by

$$P := i\left(\frac{LJL + L^* J L^*}{2} - D\right).$$

The verification of (1.58) is rather tedious but straightforward. Note that P is skew-adjoint, which is consistent with (1.58) and the unitarity of L.

A completely integrable system contains some quite low-dimensional invariant sets; in many cases (as with the harmonic oscillator), these invariant sets take the form of torii. A very interesting question concerns the stability of such invariant surfaces; if one makes a perturbation to the Hamiltonian (destroying the complete integrability), does the invariant surface similarly perturb? The answer can be surprisingly subtle, involving the theory of *Kolmogorov-Arnold-Moser torii*, *Nekhoroshev stability*, and *Arnold diffusion*, among other things. We will not attempt to describe this theory here, but refer the reader to [**Kuk3**] for a discussion of these topics in the context of Hamiltonian PDE.

Exercise 1.56 (Lagrangian submanifolds). Call a linear subspace V of a symplectic phase space $(\mathcal{D}, \omega)$ *null* if $\omega(v, v') = 0$ for all $v, v' \in V$. Show that if V is null, then the dimension of V cannot exceed half the dimension of $\mathcal{D}$. (Hint: look at the symplectic complement $V^\perp := \{u \in \mathcal{D} : \omega(v,u) = 0 \text{ for all } v \in V\}$.) Conclude that if $E_1, \ldots, E_k$ are functions which Poisson commute with a given Hamiltonian H, then for each $u \in \mathcal{D}$ the gradients $\nabla_\omega E_1(u), \ldots, \nabla_\omega E_k(u)$ span a space of dimension at most half the dimension of $\mathcal{D}$.

Exercise 1.57 (Conservation of spectrum). Let the notation and hypotheses be as in Proposition 1.46. Suppose that for some time $t_0 \in I$ and some $\lambda \in L$ we have $\lambda \in \sigma(L)$, thus there exists a non-zero eigenvector $\phi_0 \in H$ such that $L(u(t_0))\phi_0 - \lambda\phi_0 = 0$. Now let $\phi : I \to H$ solve the Cauchy problem

$$\partial_t \phi(t) = P(u(t))\phi(t); \quad \phi(t_0) = \phi_0.$$

Show that such a solution ϕ exists, and furthermore we have

$$L(u(t))\phi(t) - \lambda\phi(t) = 0$$

for all $t \in I$. (Hint: use Exercise 1.13). Conclude that the spectrum $\sigma(L)$ is an invariant of the flow.

Exercise 1.58 (Lax pairs vs. Hamiltonian mechanics). uppose that a symplectic phase space $(\mathcal{D}, \omega)$ is endowed with maps $L \in C^1_{\mathrm{loc}}(\mathcal{D} \to \mathrm{End}(H))$ and $R \in C^1_{\mathrm{loc}}(\mathcal{D} \to \mathrm{End}(\mathrm{End}(H)))$. Suppose we also have the *R-matrix identity*

$$\{\mathrm{tr}(AL), \mathrm{tr}(BL)\} = \mathrm{tr}(BR([L,A]) - AR([L,B]))$$

for all $A, B \in M_n(\mathbf{C})$, where $\{,\}$ denotes the Poisson bracket. Conclude the Poisson commutation relations

$$\{\mathrm{tr}(AL), \mathrm{tr}(L^k)\} = -k\mathrm{tr}(A[L, R^t(L^{k-1})])$$

and

$$\{\mathrm{tr}(L^m), \mathrm{tr}(L^k)\} = 0$$

for all $m, k \geq 0$ and $A \in \mathrm{End}(H)$, where $R^t : \mathcal{D} \to \mathrm{End}(\mathrm{End}(H))$ is the transpose of R, thus $\mathrm{tr}(AR(B)) = \mathrm{tr}(BR(A))$ for all $A, B \in \mathrm{End}(H)$. (Hint: take advantage of the trace identities (1.60), (1.61) and the Leibnitz rule (1.36)). Conclude that the Hamiltonian flows given by the Poisson-commuting Hamiltonians $\mathrm{tr}(L^k)$ each have a Lax pair (L, P_k) with $P_k := -kR^t(L^{k-1})$.

Exercise 1.59 (Hamiltonian formulation of Toda). Let $\mathcal{D} = \mathbf{R}^N \times \mathbf{R}^N$ be the phase space in Exercise 1.27, where we shall abuse notation and write the phase space variables as p_n, q_n where n ranges over the cyclic group $\mathbf{Z}/N\mathbf{Z}$. Consider the Hamiltonian

$$H(q,p) = \sum_{n \in \mathbf{Z}/N\mathbf{Z}} \frac{1}{2}p_n^2 + V(q_{n+1} - q_n)$$

where $V : \mathbf{R} \to \mathbf{R}$ is the *Toda potential* $V(x) := e^{-x} + x - 1$. Show that the associated Hamiltonian flow is equivalent to the Toda equations (1.62) after making the *Flaschka change of variables*

$$a_n := \frac{1}{2}e^{-(q_{n+1}-q_n)/2}; \quad b_n := -\frac{1}{2}p_n.$$

Furthermore, show that $H = 2\mathrm{tr}(L^2)$.

Exercise 1.60. Suppose we are given initial data $F_n(0)$ for $n \in \mathbf{Z}/N\mathbf{Z}$ with $|F_n(0)| < 1$ for all $n \in \mathbf{Z}/N\mathbf{Z}$. Show that there is a unique global classical solution to (1.63) with this initial data, and that we have $|F_n(t)| < 1$ for all $n \in \mathbf{Z}/N\mathbf{Z}$ and $t \in \mathbf{R}$.

CHAPTER 2

Constant coefficient linear dispersive equations

God runs electromagnetics by wave theory on Monday, Wednesday, and Friday, and the Devil runs them by quantum theory on Tuesday, Thursday, and Saturday. (Sir William Bragg)

Having concluded our discussion of ODE, we begin the analysis of dispersive[1] PDE. In this chapter, we shall begin with the study of constant-coefficient linear dispersive PDE, which are the simplest example of a dispersive equation. Furthermore, much of the theory of nonlinear PDE, especially for short times or small data, is obtained by perturbation of the linear theory; thus it is essential to have a satisfactory theory of the linear equation before proceeding to the nonlinear one.

To simplify the discussion[2], our partial differential equations shall always take as their spatial domain either a Euclidean space $\mathbf{R}^d$, or the standard torus $\mathbf{T}^d = (\mathbf{R}/2\pi\mathbf{Z})^d$; functions on the latter domain can of course be viewed as periodic functions on the former domain, and so we shall give our definitions for $\mathbf{R}^d$ only, as the generalisation to $\mathbf{T}^d$ will be clear. Also, we shall begin by focusing on PDE which are first-order in time. A constant-coefficient linear dispersive PDE then takes the form

$$\partial_t u(t,x) = Lu(t,x); \quad u(0,x) = u_0(x) \tag{2.1}$$

where the *field*[3] $u : \mathbf{R} \times \mathbf{R}^d \to V$ takes values in a finite-dimensional Hilbert space V, and L is a skew-adjoint constant coefficient differential operator in space, thus taking the form

$$Lu(x) := \sum_{|\alpha| \leq k} c_\alpha \partial_x^\alpha u(x),$$

where $k \geq 1$ is an integer (the *order* of the differential operator), $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbf{Z}_+^n$ ranges over all multi-indices with $|\alpha| := \alpha_1 + \ldots + \alpha_d$ less than or equal to k,

[1]Informally, "dispersion" will refer to the fact that different frequencies in this equation will tend to propagate at different velocities, thus dispersing the solution over time. This is in contrast to transport equations such as (2.2), which move all frequencies with the same velocity (and is thus a degenerate case of a dispersive equation), or dissipative equations such as the heat equation $\partial_t u = \Delta u$, in which frequencies do not propagate but instead simply attenuate to zero. The wave equation (2.9) is partly dispersive - the frequency of a wave determines the direction of propagation, but not the speed; see Principle 2.1.

[2]The study of linear dispersive equations in the presence of potentials, obstacles or other boundary conditions, or on curved manifolds or in variable coefficient situations, is of great importance in PDE, with applications to spectral theory, geometry, and even number theory; however, we will not attempt to investigate these topics here.

[3]We shall say that the field is *real* if V is a real vector space, and *complex* if V is a complex vector space. We say that the field is *scalar* if V is one-dimensional, *vector* if V is viewed as a vector space, *tensor* if V is viewed as a tensor space, etc. For instance, a field taking values in $\mathbf{C}^d$ would be a complex vector field. We will not use the term "field" in the algebraic sense in this text.

∂_x^α is the partial derivative

$$\partial_x^\alpha := (\frac{\partial}{\partial x_1})^{\alpha_1} \dots (\frac{\partial}{\partial x_d})^{\alpha_d},$$

and $c_\alpha \in \mathrm{End}(V)$ are coefficients that do not depend on x. This operator is classically only defined on k-times continuously differentiable functions, but we may extend it to distributions or functions in other function spaces in the usual manner; thus we can talk about both classical and weak (distributional) solutions to (2.1). We can also write $L = ih(D)$, where D is the frequency operator

$$D := \frac{1}{i}\nabla = (\frac{1}{i}\partial_{x_1}, \dots, \frac{1}{i}\partial_{x_d})$$

and $h : \mathbf{R}^d \to \mathrm{End}(V)$ is the polynomial

$$h(\xi_1, \dots, \xi_d) = \sum_{|\alpha| \leq k} i^{|\alpha|-1} c_\alpha \xi_1^{\alpha_1} \dots \xi_d^{\alpha_d}.$$

We assume that L is skew-adjoint, thus

$$\int \langle Lu(x), v(x) \rangle_V \; dx = -\int \langle u(x), Lv(x) \rangle_V \; dx$$

for all test functions u, v; this is equivalent to requiring that coefficients of the polynomial h be self-adjoint, so in the scalar case we require h to be real-valued. Note that we do not restrict the time variable to an interval I; this is because the solutions we shall construct to (2.1) will automatically exist globally in time. We refer to the polynomial h as the *dispersion relation* of the equation (2.1).

A somewhat degenerate example of an equation of the form (2.1) is the phase rotation equation

$$\partial_t u(t,x) = i\omega u(t,x); \quad u(0,x) = u_0(x)$$

where u is a complex field and $\omega \in \mathbf{R}$; this has the explicit solution $u(t,x) = e^{i\omega t} u_0(x)$, and the dispersion relation is $h(\xi) = \omega$. Another degenerate example is the transport equation

$$\partial_t u(t,x) = -v \cdot \nabla_x u(t,x); \quad u(0,x) = u_0(x) \tag{2.2}$$

for some constant vector $v \in \mathbf{R}^d$; this has the explicit solution $u(t,x) = u_0(x - vt)$, and the dispersion relation is $h(\xi) = -v \cdot \xi$. More interesting examples (many of which arise from physics) can be constructed if one either raises the order of L, or makes u vector-valued instead of scalar. Examples of the former include the *free Schrödinger equation*

$$i\partial_t u + \frac{\hbar}{2m}\Delta u = 0, \tag{2.3}$$

where $u : \mathbf{R} \times \mathbf{R}^d \to V$ is a complex field and $\Delta = \sum_{j=1}^d \frac{\partial^2}{\partial x_j^2}$ is the Laplacian and *Planck's constant* $\hbar > 0$ and the mass $m > 0$ are fixed scalars, as well as the one-dimensional *Airy equation*

$$\partial_t u + \partial_{xxx} u = 0 \tag{2.4}$$

where $u : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ is a real scalar field. The dispersion relations here are $h(\xi) = -\frac{\hbar}{2m}|\xi|^2$ and $h(\xi) = \xi^3$ respectively. Examples of the latter include *vacuum Maxwell's equations*

$$\partial_t E = c^2 \nabla_x \times B; \quad \partial_t B = -\nabla_x \times E; \quad \nabla_x \cdot E = \nabla_x \cdot B = 0 \tag{2.5}$$

in three dimensions $d = 3$, where $E, B : \mathbf{R}^{1+3} \times \mathbf{R}^3$ are real vector[4] fields and the *speed of light* $c > 0$ is constant; the constraints that E and B be divergence-free are not of the dynamical form (2.1), but nevertheless they end up being compatible with the flow (Exercise 2.16). The Maxwell equations are a special case of the *abelian Yang-Mills equations*

$$\partial_\alpha F^{\alpha\beta} = 0; \quad \partial_\alpha F_{\beta\gamma} + \partial_\beta F_{\gamma\alpha} + \partial_\gamma F_{\alpha\beta} = 0 \tag{2.6}$$

where $F : \mathbf{R}^{1+d} \to \bigwedge^2 \mathbf{R}^{1+d}$ is an real anti-symmetric two-form field, and $\mathbf{R}^{1+d} = \mathbf{R} \times \mathbf{R}^d$ is endowed[5] with the standard Minkowski metric $g^{\alpha\beta}$, defined using the spacetime interval $dg^2 = -c^2 dt^2 + dx_1^2 + dx_2^2 + dx_3^2$ (with the convention $x_0 = t$), and which is used to raise and lower indices in the usual manner.

Another example from physics is the *Dirac equation*

$$i\gamma^\alpha \partial_\alpha u = \frac{mc}{\hbar} u, \tag{2.7}$$

where $\gamma^0, \ldots, \gamma^3 \in \mathrm{End}(V)$ are the *gamma matrices*, acting on a four-dimensional complex vector space V, known as *spinor space*, via the commutation relations

$$\gamma^\alpha \gamma^\beta + \gamma^\beta \gamma^\alpha = -2g^{\alpha\beta} \mathrm{id}_V \tag{2.8}$$

where $g^{\alpha\beta}$ is the Minkowski metric, the mass $m \geq 0$ is non-negative, and $u : \mathbf{R}^{1+3} \to V$ is a spinor field; see Exercise 2.1 for one construction of spinor space.

It is also of interest to consider dispersive equations which are second-order in time. We will not give a systematic description of such equations here, but instead only mention the two most important examples, namely the *wave equation*

$$\Box u = 0; \quad u(0, x) = u_0(x); \quad \partial_t u(0, x) = u_1(x) \tag{2.9}$$

where $u : \mathbf{R}^{1+d} \to V$ is a field, and $\Box$ is the *d'Alembertian operator*

$$\Box = \partial^\alpha \partial_\alpha = -\frac{1}{c^2}\partial_t^2 + \Delta,$$

and the slightly more general *Klein-Gordon equation*

$$\Box u = \frac{m^2 c^2}{\hbar^2} u; \quad u(0, x) = u_0(x); \quad \partial_t u(0, x) = u_1(x) \tag{2.10}$$

where the mass $m \geq 0$ is fixed.

Equations which involve c are referred to as *relativistic*, while equations involving $\hbar$ are *quantum*. Of course, one can select units of space and time so that $c = \hbar = 1$, and one can also normalise $m = 1$ without much difficulty; these constants need to be retained however if one wants to analyze the *non-relativistic limit* $c \to \infty$ of a relativistic equation, the *classical limit* $\hbar \to 0$ of a quantum equation, or the *massless limit* $m \to 0$ of a massive equation.

[4]A more geometrically covariant statement would be that E and B combine to form a rank two tensor field $F_{\alpha\beta}$, as in the example of the Yang-Mills equations below. Indeed, $F_{\alpha\beta}$ should be interpreted as the curvature of a connection; cf. Section 6.2.

[5]We shall use $\mathbf{R}^{1+d}$ to denote Minkowski space with the Minkowski metric, and $\mathbf{R} \times \mathbf{R}^d$ to denote a spacetime without any Minkowski metric. Relativistic equations such as the Maxwell, Yang-Mills, Dirac, Klein-Gordon, and wave equations live on Minkowski space, and thus interact well with symmetries of this space such as the Lorentz transformations, whereas non-relativistic equations such as the Airy and Schrödinger equation have no relation to the Minkowski metric or with any related structures such as Lorentz transformations, the light cone, or the raising and lowering conventions associated with the metric.

The constant-coefficient dispersive equations have a number of symmetries. All are invariant under time translation $u(t,x) \mapsto u(t-t_0,x)$ and spatial translation $u(t,x) \mapsto u(t,x-x_0)$. Several also enjoy a scaling symmetry (Exercise 2.9). There is also usually a time reversal symmetry, though the precise nature of the symmetry varies. For instance, for the Schrödinger equation one takes $u(t,x) \mapsto \overline{u(-t,-x)}$, for the Airy equation one takes $u(t,x) \mapsto u(-t,-x)$, and for the wave and Klein-Gordon equations one takes $u(t,x) \mapsto u(-t,x)$. For tensor equations such as Dirac, Maxwell, and Yang-Mills, one has to apply the time reversal to the tensor space as well as to the spacetime domain. The equations in $\mathbf{R}^d$ typically enjoy a rotation and reflection invariance, which for scalar equations is simply $u(t,x) \mapsto u(t,Ux)$ for all orthogonal matrices $U \in O(d)$; in particular, this implies (once one has a reasonable uniqueness theory) that radially symmetric data leads to radially symmetric solutions. Again, for tensor equations one has to rotate the tensor as well as the domain. The Schrödinger equation also has a very useful Galilean invariance (Exercise 2.5), while the relativistic equations have a similarly useful Lorentz invariance (Exercise 2.6). Finally, the Schrödinger equation also enjoys the pseudo-conformal symmetry (Exercise 2.28), while the wave equation similarly enjoys the conformal symmetry (Exercise 2.14).

The above equations are all connected to each other in various ways; some of them are given in the exercises.

Exercise 2.1 (Spinor space). Let $V = \mathbf{C}^4$ be endowed with the sesquilinear form

$$\{(z_1,z_2,z_3,z_4),(w_1,w_2,w_3,w_4)\} := z_1\overline{w_1} + z_2\overline{w_2} - z_3\overline{w_3} - z_4\overline{w_4}$$

and let $\gamma^0,\ldots,\gamma^3 \in \mathrm{End}(V)$ be given as

$$\begin{aligned}
\gamma^0(z_1,z_2,z_3,z_4) &= \frac{1}{c}(z_1,z_2,-z_3,-z_4)\\
\gamma^1(z_1,z_2,z_3,z_4) &= (z_4,z_3,-z_2,-z_1)\\
\gamma^2(z_1,z_2,z_3,z_4) &= (-iz_4,iz_3,iz_2,-iz_1)\\
\gamma^3(z_1,z_2,z_3,z_4) &= (z_3,-z_4,-z_1,z_2).
\end{aligned}$$

Show that we have the commutation relations (2.8). Furthermore, for all $u,v \in V$, we have the symmetry $\{\gamma^\alpha u, v\} = \{u, \gamma^\alpha v\}$, and the 4-vector $\{u,\gamma^\alpha u\}$ is positive time-like in the sense that

$$\{u,\gamma^0 u\} \geq 0 \text{ and } -\{u,\gamma^\alpha u\}\{u,\gamma_\alpha u\} \geq \{u,u\}^2 \geq 0.$$

Exercise 2.2 (Maxwell vs. Yang-Mills; Dirac vs. Klein-Gordon). Show that any $C^2_{t,x,\mathrm{loc}}$ solution to Maxwell's equation (2.5) or the abelian Yang-Mills equation (2.6), also solves the wave equation (2.9). Conversely, if $A = (A_\alpha)_{\alpha=0,\ldots,d} \in C^2_{t,x,\mathrm{loc}}(\mathbf{R}^{1+d} \to \mathbf{R}^{1+d})$ solves the wave equation, show that the curvature $F_{\alpha\beta} := \partial_\alpha A_\beta - \partial_\beta A_\alpha$ solves the abelian Yang-Mills equation. Also explain why Maxwell's equation is a special case of the abelian Yang-Mills equation. In a similar spirit, show that any continuously twice differentiable solution to Dirac's equation (2.7) also solves the Klein-Gordon equation (2.10), and conversely if $\phi \in C^2_{t,x,\mathrm{loc}}(\mathbf{R}^{1+3} \to V)$ solves the Klein-Gordon equation then $u := \gamma^\alpha\partial_\alpha\phi + m\phi$ solves the Dirac equation.

Exercise 2.3 (Airy vs. Schrödinger). Let $u \in C^\infty_{t,x,\mathrm{loc}}(\mathbf{R}\times\mathbf{R} \to \mathbf{C})$ solve the Schrödinger equation $i\partial_t u + \partial_x^2 u = 0$, with all derivatives uniformly bounded. Let

$N > 1$ be a large number, and let $v : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ be the real scalar field

$$v(t,x) := \mathrm{Re}\left(e^{iNx+iN^3t}u(t, \frac{x+3N^2t}{\sqrt{3N}})\right).$$

Show that v is an approximate solution to the Airy equation (2.4), in the sense that

$$\partial_t v + \partial_x^3 v = O_u(N^{-3/2}).$$

This suggests that solutions to the Airy equation can behave, in certain circumstances, like suitably rescaled and modulated solutions to the Schrödinger equation. See [**BC**], [**CCT**], [**Schn**] for some nonlinear developments of this idea.

EXERCISE 2.4 (Group velocity of waves). Let $h : \mathbf{R}^d \to \mathbf{R}$ be a polynomial with real coefficients, and let $L := iP(D)$. Show that if $\phi \in C^\infty_{t,x,\mathrm{loc}}(\mathbf{R}^d \to \mathbf{C})$ has all derivatives bounded, and $\varepsilon > 0$, then the complex scalar field $u \in C^\infty_{t,x,\mathrm{loc}}(\mathbf{R}\times\mathbf{R}^d \to \mathbf{C})$ defined by

$$u(t,x) := e^{ix\cdot\xi_0+ith(\xi_0)}\phi(\varepsilon(x+\nabla h(\xi_0)t))$$

is an approximate solution to (2.1) in the sense that

$$\partial_t u = Lu + O_\phi(\varepsilon^2).$$

This suggests that (sufficiently broad) solutions to (2.1) which oscillate in space with frequency ξ_0, should travel at group velocity $-\nabla h(\xi_0)$, and oscillate in time with frequency $h(\xi_0)$; see also Principle 2.1. In the case of the Schrödinger equation (2.3), conclude (on a heuristic level) *de Broglie's law* $mv = \hbar\xi$, where v denotes the group velocity. (Note the *phase velocity* in this case will be twice the group velocity, $2v$. More generally, an inspection of the phase $x \cdot \xi_0 + th(\xi_0)$ shows that the phase velocity is $-\frac{\xi_0}{|\xi_0|}\frac{h(\xi_0)}{|\xi_0|}$.)

EXERCISE 2.5 (Galilean invariance). Let $u \in C^2_{t,x,\mathrm{loc}}(\mathbf{R}\times\mathbf{R}^d \to V)$ be a complex field, let $v \in \mathbf{R}^d$, and let $\tilde u \in C^2_{t,x,\mathrm{loc}}(\mathbf{R}\times\mathbf{R}^d \to \mathbf{C})$ be the field

$$\tilde u(t,x) := e^{imx\cdot v/\hbar}e^{imt|v|^2/2\hbar}u(t,x-vt).$$

show that $\tilde u$ solves the Schrödinger equation (2.3) if and only if u does.

EXERCISE 2.6 (Lorentz invariance). Let $u \in C^2_{t,x,\mathrm{loc}}(\mathbf{R}^{1+d} \to V)$ be a field, let $v \in \mathbf{R}^d$ be such that $|v| < c < \infty$, and let $u_v \in C^2_{t,x,\mathrm{loc}}(\mathbf{R}^{1+d} \to \mathbf{C})$ be the field

$$u_v(t,x) := u(\frac{t - v\cdot x/c^2}{\sqrt{1-|v|^2/c^2}}, x - x_v + \frac{x_v - vt}{\sqrt{1-|v|^2/c^2}})$$

where $x_v := (x \cdot \frac{v}{|v|})\frac{v}{|v|}$ is the projection of x onto the line parallel to v (with the convention that $x_v = 0$ when $v = 0$). Thus for instance, if $c = 1$, $v = v_1e_1$ for some $-1 < v_1 < 1$ and $\underline{x} := (x_2, \ldots, x_d)$, then

$$u_v(t,x_1,\underline{x}) := u(\frac{t-v_1x_1}{\sqrt{1-v_1^2}}, \frac{x_1-v_1t}{\sqrt{1-v_1^2}}, \underline{x}).$$

Show that u_v solves the wave equation (2.9) if and only if u does, and similarly for the Klein-Gordon equation (2.10). (Hint: show that the Minkowski metric is preserved by the Lorentz transformation $(t,x) \mapsto (\frac{t-v\cdot x/c^2}{\sqrt{1-|v|^2/c^2}}, x - x_v + \frac{x_v-vt}{\sqrt{1-|v|^2/c^2}})$.) What is the analogous symmetry for the Dirac, Maxwell, and Yang-Mills equations?

EXERCISE 2.7 (Schrödinger vs. Klein-Gordon). Let $u \in C^\infty_{t,x,\mathrm{loc}}(\mathbf{R}^{1+d} \to V)$ be a complex field solving the Klein-Gordon equation (2.10). Show that if one applies the change of variables $u = e^{-itmc^2/\hbar} v$ then one obtains

$$i\partial_t v + \frac{\hbar}{2m}\Delta v = \frac{\hbar}{2mc^2}\partial_t^2 v.$$

This suggests that the Klein-Gordon equation can converge to the Schrödinger equation in the non-relativistic limit $c \to \infty$, though one has to be extremely careful with this heuristic due to the double time derivative on the right-hand side. (A more robust approximation is given in the next exercise.)

EXERCISE 2.8 (Schrödinger vs. Dirac). Let $u : \mathbf{R} \times \mathbf{R}^3 \to V$ be a spinor field solving the Schrödinger equation

$$ic\gamma^0\partial_t u - \frac{\hbar}{2m}\Delta u = 0$$

with all derivatives uniformly bounded. Let $v : \mathbf{R} \times \mathbf{R}^3 \to V$ be the spinor field

$$v := e^{-imc^2\gamma^0 t/\hbar} u - \frac{\hbar}{2imc} e^{imc^2\gamma^0 t/\hbar}\gamma^j \partial_{x_j} u$$

where the j index is summed over $1, 2, 3$. Show that v is an approximate solution to the Dirac equation (2.7) (and hence the Klein-Gordon equation) in the sense that

$$i\gamma^\alpha\partial_\alpha v = \frac{mc}{\hbar} v + O_{\hbar,m,u}(\frac{1}{c^2})$$

Thus in the non-relativistic limit $c \to \infty$, certain solutions of the Dirac and Klein-Gordon equations resemble suitably rescaled and modulated solutions of the Schrödinger equation. See [**MNO**], [**MNO2**] for some nonlinear developments of this idea. By using this correspondence between Schrödinger and Klein-Gordon, one can also establish in a certain sense that the Lorentz invariance degenerates to the Galilean invariance in the non-relativistic limit $c \to \infty$; we omit the details.

EXERCISE 2.9 (Scaling symmetry). Show that if $P : \mathbf{R}^d \to \mathbf{C}$ is a homogeneous polynomial of degree k, and $L = P(\nabla)$, then the equation (2.1) is invariant under the scaling $u(t, x) \mapsto u(\frac{t}{\lambda^k}, \frac{x}{\lambda})$ for any $\lambda > 0$. Thus for instance, with the Schrödinger equation the time variable has "twice the dimension" of the space variable, whereas for the Airy equation the time variable has three times the dimension of the space variable. For relativistic equations such as the wave equation, space and time have the same dimension.

EXERCISE 2.10 (Wave vs. Klein-Gordon). Let $u \in C^2_{t,x,\mathrm{loc}}(\mathbf{R} \times \mathbf{R}^d \to V)$ be a complex field, and define $v \in C^2_{t,x,\mathrm{loc}}(\mathbf{R} \times \mathbf{R}^{d+1} \to \mathbf{C})$ by

$$v(t, x_1, \ldots, x_n, x_{d+1}) = e^{imcx_{d+1}/\hbar} u(t, x_1, \ldots, x_d).$$

Show that v solves the $d+1$-dimensional wave equation (2.9) if and only if u solves the d-dimensional Klein-Gordon equation (2.10). This allows one to use the *method of descent* (analyzing a lower-dimensional PDE by a higher-dimensional PDE) to obtain information about the Klein-Gordon equation from information about the wave equation. As the name implies, the method of descent is largely one-way; it identifies general solutions to lower-dimensional PDE with *special* solutions to higher-dimensional PDE, but does not yield much information on *general* solutions to the higher-dimensional PDE.

EXERCISE 2.11 (Wave vs. Schrödinger). Let $u \in C^2_{t,x,\mathrm{loc}}(\mathbf{R} \times \mathbf{R}^d \to V)$ be a complex field, and define $v \in C^2_{t,x,\mathrm{loc}}(\mathbf{R} \times \mathbf{R}^{d+1} \to V)$ by

$$v(t, x_1, \dots, x_d, x_{d+1}) = e^{-i(t+x_{d+1})} u(\frac{t - x_{d+1}}{2}, x_1, \dots, x_d).$$

Set $\hbar = c = m = 1$. Show that v solves the $d+1$-dimensional wave equation (2.9) if and only if u solves the d-dimensional Schrödinger equation (2.3). See also Exercise 3.2.

EXERCISE 2.12 (Wave vs. wave). Suppose that the field $u \in C^2_{t,x,\mathrm{loc}}(\mathbf{R} \times (\mathbf{R}^d \backslash \{0\}) \to V)$ solves the d-dimensional wave equation $\Box u = 0$, thus by abuse of notation we can write $u(t, x) = u(t, |x|)$ and consider $u = u(t, r)$ now as a function from $\mathbf{R} \times (0, \infty) \to \mathbf{C}$. Conclude the radial field $v \in C^2_{t,x,\mathrm{loc}}(\mathbf{R} \times (\mathbf{R}^{d+2} \backslash \{0\}) \to V)$ defined (again abusing notation) by $v(t, r) := \frac{1}{r} \partial_r u(t, r)$ solves the $d+2$-dimensional wave equation. Thus in the radial case at least, it is possible to construct solutions to the $d + 2$-dimensional equation out of the d-dimensional equation.

EXERCISE 2.13 ($1 + 1$ waves). Show that if the field $u \in C^\infty_{t,x,\mathrm{loc}}(\mathbf{R} \times \mathbf{R} \to V)$ solves the one-dimensional wave equation $\partial_t^2 u - \partial_x^2 u = 0$ with initial data $u(0, x) = u_0(x), \partial_t u(0, x) = u_1(x)$, then

$$u(t, x) := \frac{u_0(x + t) + u_0(x - t)}{2} + \frac{1}{2} \int_{x-t}^{x+t} u_1(y)\ dy$$

for all $t, x \in \mathbf{R}$. This is a rare example of a PDE which can be solved entirely by elementary means such as the fundamental theorem of calculus.

EXERCISE 2.14 (Conformal invariance). Let $\Gamma_+ \subset \mathbf{R}^{1+d}$ be the forward light cone $\{(\tau, \xi) \in \mathbf{R}^{1+d} : \tau > c|\xi|\}$, and let $u \in C^2_{t,x,\mathrm{loc}}(\Gamma_+ \to V)$ be a field. Let $\tilde{u} \in C^2_{t,x,\mathrm{loc}}(\Gamma_+ \to V)$ be the conformal inversion of u, defined by

$$\tilde{u}(t, x) := (t^2 - |x|^2)^{-(d-1)/2} u(\frac{t}{t^2 - |x|^2}, \frac{x}{t^2 - |x|^2}).$$

Establish the identity

$$\Box \tilde{u}(t, x) = (t^2 - |x|^2)^{-\frac{d-1}{2} - 1} \Box u(\frac{t}{t^2 - |x|^2}, \frac{x}{t^2 - |x|^2}).$$

In particular u solves the wave equation (2.9) with $c = 1$ if and only if $\tilde{u}$ does. (One can use in fact hyperbolic polar coordinates to recast the wave equation as a wave equation on hyperbolic space H^d, in such a way that conformal inversion amounts simply to time reversal. Another approach is to observe that *Kelvin inversion* $(t, x) \mapsto \frac{1}{t^2 - |x|^2}(t, x)$ is a conformal transformation of Minkowski space.)

EXERCISE 2.15 (Quantum vs. classical dynamics). Let $\vec{p} = (p_1, \dots, p_d)$ denote the *momentum operator* $\vec{p} := -i\hbar\nabla$; show that when applied to the approximate solutions to the Schrödinger equation in Exercise 2.4 that we have *de Broglie's law* $\vec{p} \approx mv = \hbar\xi$. Let $H := \frac{|\vec{p}|^2}{2m} = \frac{p_1^2 + \ldots + p_d^2}{2m}$ denote the *Hamiltonian* (this is the analogue of the classical Hamiltonian $H = \frac{|\vec{p}|^2}{2m} = \frac{1}{2} m |v|^2$ in the absence of a potential). If $u \in C_t^\infty \mathcal{S}_x(\mathbf{R} \times \mathbf{R}^d)$ is a classical solution to the Schrödinger equation (2.3) and $E : \mathcal{S}_x(\mathbf{R}^d) \to \mathcal{S}_x(\mathbf{R}^d)$ is any (time-independent) continuous linear operator, establish the *Heisenberg equation*

$$\frac{d}{dt}\langle Eu(t), u(t)\rangle_{L^2_x(\mathbf{R}^d)} = \langle \frac{i}{\hbar}[H,E]u(t), u(t)\rangle_{L^2_x(\mathbf{R}^d)}$$

where $[H,E] := HE - EH$ is the *Lie bracket* and $\langle f, g\rangle_{L^2_x(\mathbf{R}^d)} := \int_{\mathbf{R}^d} f(x)\overline{g(x)}$. Compare this with (1.33) and with Exercise 1.36. The precise connection between the Lie bracket and the Poisson bracket involves the subjects of *semiclassical analysis*, *microlocal analysis*, and *geometric quantization*, which we will not discuss here.

EXERCISE 2.16 (Consistency of Maxwell's equations). Suppose that $E, B \in C^\infty_t \mathcal{S}_x(\mathbf{R}^{1+3} \to \mathbf{R}^3)$ solve the "dynamic" component $\partial_t E = c^2 \nabla_x \times B; \quad \partial_t B = -\nabla_x \times E$ of Maxwell's equations (2.5). Suppose also that the "static" components of the equation hold at time $t = 0$, thus $\nabla_x \cdot E(0,x) = \nabla_x \cdot B(0,x) = 0$. Show that Maxwell's equations in fact hold for all time. (Hint: compute the time derivative of the quantity $\int_{\mathbf{R}^3} |\nabla \cdot E(t,x)|^2 + |\nabla \cdot B(t,x)|^2\ dx$ using integration by parts.)

2.1. The Fourier transform

> *His life oscillates, as everyone's does, not merely between two poles, such as the body and the spirit, the saint and the sinner, but between thousands, between innumerable poles.* (Herman Hesse, "Steppenwolf")

The spatial Fourier transform, and the closely related spacetime Fourier transform, is an exceptionally well-suited[6] tool to analyze constant coefficient linear dispersive equations such as (2.1). This is ultimately due to the invariance of these equations under translations in either space or time. A brief summary of the properties of the Fourier transform that we shall need (as well as other notation appearing here, such as the Sobolev spaces $H^s_x(\mathbf{R}^d)$) can be found in Appendix A.

One hint that the Fourier transform will be useful for solving equations such as (2.1) comes from the simple observation that given any frequency $\xi_0 \in \mathbf{R}^d$ and any $P : \mathbf{R}^d \to \mathbf{R}$, the plane wave $e^{ix\cdot\xi_0 + ith(\xi_0)}$ solves the equation (2.1) with $L = ih(D)$ (see Exercise 2.4). From the principle of superposition for linear equations, we thus see that we can construct solutions to (2.1) as superpositions of plane waves.

In order to obtain an adequate wellposedness theory for dispersive equations, it is often necessary to restrict attention to solutions which not only have some smoothness, but also some decay. To get some idea of the problems involved, consider the complex scalar field $u : (-\infty, 0) \times \mathbf{R}^d \to \mathbf{C}$ defined by

$$u(t,x) := \frac{1}{|t|^{d/2}} e^{im|x|^2/2\hbar t}.$$

This field can be verified to be a smooth solution to the Schrödinger equation (2.3) for all negative times, but becomes singular at time $t = 0$. The problem is that while this solution is smooth, it does not decay at all as $x \to \infty$. For an even worse example of bad behaviour of the Schrödinger equation - namely breakdown of uniqueness even for smooth solutions - see Exercise 2.24.

[6]In some ways, it is *too* well suited; there are a number of results in this field which are so easily proven using the Fourier transform, that non-Fourier-based alternative proofs have not been adequately explored, and as such one encounters difficulty extending those results to variable-coefficient, curved space, or nonlinear settings in which the Fourier transform is less useful.

To avoid these issues, we begin by restricting attention to the Schwartz space $\mathcal{S}_x(\mathbf{R}^d)$. To simplify the discussion, let us now only consider scalar equations, so that the dispersion relation $h : \mathbf{R}^d \to \mathbf{R}$ is now real-valued; the vector-valued case introduces a number of interesting new technical issues which we will not discuss here. If $u \in C^1_{t,\mathrm{loc}}\mathcal{S}_x(\mathbf{R} \times \mathbf{R}^d)$ is a classical solution to (2.1), then by taking Fourier transforms of (2.1) we conclude

$$\partial_t \widehat{u(t)}(\xi) = ih(\xi)\widehat{u(t)}(\xi)$$

which has the unique solution

$$\widehat{u(t)}(\xi) = e^{ith(\xi)}\widehat{u_0}(\xi). \tag{2.11}$$

Note that as $h(\xi)$ is real and $\widehat{u_0}$ is Schwartz, the function $e^{ith(\xi)}\widehat{u_0}(\xi)$ is then also Schwartz for any $t \in \mathbf{R}$. We may thus apply the Fourier inversion formula and obtain the solution

$$u(t,x) = \int_{\mathbf{R}^d} e^{ith(\xi)+ix\cdot\xi}\widehat{u_0}(\xi)\ d\xi; \tag{2.12}$$

because of this, we shall let $e^{tL} = e^{ith(D)}$ denote the linear propagator

$$e^{tL}u_0(x) := \int_{\mathbf{R}^d} e^{ith(\xi)+ix\cdot\xi}\widehat{u_0}(\xi)\ d\xi.$$

This propagator is defined initially for Schwartz functions, but can be extended by standard density arguments to other spaces. For instance, Plancherel's theorem allows one to extend e^{tL} to be defined on the Lebesgue space $L^2_x(\mathbf{R}^d)$, or more generally to the inhomogeneous Sobolev space[7] $H^s_x(\mathbf{R}^d)$ for any $s \in \mathbf{R}$, as well as the homogeneous Sobolev spaces $\dot{H}^s_x(\mathbf{R}^d)$ (see Appendix A). It is clear that e^{tL} is a unitary operator on these spaces, and in particular on $L^2_x(\mathbf{R}^d)$ (which by Plancherel's identity is equivalent to $H^0_x(\mathbf{R}^d) = \dot{H}^0_x(\mathbf{R}^d)$, except for an inessential factor of $(2\pi)^{d/2}$):

$$\|e^{tL}f\|_{H^s_x(\mathbf{R}^d)} = \|f\|_{H^s_x(\mathbf{R}^d)}; \|e^{tL}f\|_{\dot{H}^s_x(\mathbf{R}^d)} = \|f\|_{\dot{H}^s_x(\mathbf{R}^d)} \quad \|e^{tL}f\|_{L^2_x(\mathbf{R}^d)} = \|f\|_{L^2_x(\mathbf{R}^d)}.$$

One can of course also extend these propagator to tempered distributions by duality.

Propagators are examples of Fourier multipliers and as such, they commute with all other Fourier multipliers, including constant coefficient differential operators, translations, and other propagators. In particular they commute with the fractional differentiation and integration operators $\langle\nabla\rangle^s$ for any $s \in \mathbf{R}$.

The Fourier transform can also be defined[8] on the torus $\mathbf{T}^d$. If $f \in C^\infty_x(\mathbf{T}^d \to \mathbf{C})$ is smooth, the Fourier transform $\hat{f} : \mathbf{Z}^d \to \mathbf{C}$ is defined by

$$\hat{f}(k) := \frac{1}{(2\pi)^d}\int_{\mathbf{T}^d} f(x)e^{-ik\cdot x}\ dx.$$

[7] In the notation of the next chapter, the function $u(t) = e^{tL}u_0$ is the unique *strong* $H^s_x(\mathbf{R}^d)$ *solution* to the Cauchy problem (2.1). As a rule of thumb, as long as one restricts attention to strong solutions in a space such as H^s_x, the linear evolution is completely non-pathological.

[8] Of course, the Fourier transform can in fact be defined on any reasonable abelian group, and even (with some modifications) on most non-abelian groups; but we will not pursue these issues here. Also, completely integrable PDE often come with a "scattering transform" which can be viewed as a nonlinear version of the Fourier transform, but again we will not discuss this here.

One can show that $\hat{f}(k)$ is a rapidly decreasing function of k, and the inversion formula is given by

$$f(x) = \sum_{k \in \mathbf{Z}^d} \hat{f}(k) e^{ik \cdot x}.$$

Much of the preceding discussion extends to the periodic setting, with some minor changes; we leave the details to the reader. One advantage of the periodic setting is that the individual Fourier modes $e^{ik \cdot x}$ are now themselves square-integrable (and more generally lie in all the Sobolev spaces $H^s_x(\mathbf{T}^d)$, thus for instance $\|e^{ik \cdot x}\|_{H^s_x(\mathbf{T}^d)}$ is equal to $\langle k \rangle^s$). This makes it easier to talk about the evolution of individual Fourier modes, as compared to the non-periodic case in which the Fourier modes lie in a continuum.

The spatial Fourier transform $f(x) \mapsto \hat{f}(\xi)$ brings into view the oscillation of a function in space. In the analysis of dispersive PDE, it is also important to analyze the oscillation in *time*, which leads to the introduction of the spacetime Fourier transform. If $u : \mathbf{R} \times \mathbf{R}^d \to \mathbf{C}$ is a complex scalar field, we can define its *spacetime Fourier transform* $\tilde{u} : \mathbf{R} \times \mathbf{R}^d \to \mathbf{C}$ formally as

$$\tilde{u}(\tau, \xi) := \int_{\mathbf{R}} \int_{\mathbf{R}^d} u(t,x) e^{-i(t\tau + x \cdot \xi)}\ dtdx.$$

To begin with, this definition is only sensible for sufficiently nice functions such as those which are Schwartz in both space and time, but one can then extend it to much more general functions and to tempered distributions by density arguments or duality. Of course, in such case one does not always expect $\tilde{u}$ to be well-behaved, for instance it could be a measure or even a tempered distribution. Formally at least, we have the inversion formula

$$u(t,x) = \frac{1}{(2\pi)^{d+1}} \int_{\mathbf{R}} \int_{\mathbf{R}^d} e^{i(t\tau + x \cdot \xi)} \tilde{u}(\tau, \xi)\ d\tau d\xi.$$

The advantage of performing this transform is that it not only diagonalises the linear operator L in (2.1), but also diagonalises the time derivative ∂_t. Indeed, if u is any tempered distributional solution to (2.1) (which will in particular include classical solutions which grow at most polynomially in space and time) with $L = P(\nabla) = ih(\nabla/i)$ as before, then on taking the spacetime Fourier transform we obtain

$$i\tau \tilde{u}(\tau, \xi) = ih(\xi) \tilde{u}(\tau, \xi)$$

and thus

$$(\tau - h(\xi)) \tilde{u}(\tau, \xi) = 0.$$

The theory of distributions then shows that $\tilde{u}(\tau, \xi)$ is supported in the *characteristic hypersurface* $\{(\tau, \xi) : \tau = h(\xi)\}$ of the spacetime frequency space $\mathbf{R} \times \mathbf{R}^d$, and in fact takes the form

$$\tilde{u}(\tau, \xi) = \delta(\tau - h(\xi)) a(\xi)$$

for some spatial tempered distribution a, where δ is the Dirac delta. In the case of the Schwartz solution (2.12), we have $a = \hat{u}_0$, thus

$$\tilde{u}(\tau, \xi) = \delta(\tau - h(\xi)) \hat{u}_0(\xi).$$

For comparison, if we consider u_0 as a function of spacetime via the trivial extension $u_0(t,x) := u_0(x)$, then we have

$$\tilde{u}_0(\tau, \xi) = \delta(\tau) \hat{u}_0(\xi).$$

Thus in the spacetime frequency space, one can think of the linear solution u to (2.1) as the time-independent field u_0, twisted by the transformation $(\tau, \xi) \mapsto (\tau + h(\xi), \xi)$.

In applications to nonlinear PDE, it is often not feasible to use the spacetime Fourier transform to the solution u directly, because often u is only defined on a spacetime slab $I \times \mathbf{R}^d$ instead of the entire spacetime $\mathbf{R}^{1+d}$. This necessitates some sort of artificial extension[9] of the solution u from I to $\mathbf{R}$ in order to take advantage of the features of the spacetime Fourier transform. Nevertheless, the spacetime Fourier transform (and in particular the Sobolev spaces $X^{s,b} = X^{s,b}_{\tau=h(\xi)}$ adapted to the characteristic hypersurface) has been proven to be a very useful tool in both the study of linear and nonlinear dispersive equations; see Section 2.6.

Another approach to analyzing these PDE proceeds by taking the Fourier transform in the time variable only, keeping the spatial variable untouched. This approach is well suited for settings in which the operator L has variable-coefficients, or has a domain which is curved or has a boundary, and leads to *spectral theory*, which analyzes the behaviour of propagators such as e^{tL} in terms of resolvents $(L - z)^{-1}$, as well as the closely related spectral measure of L. This perspective has proven to be immensely useful for the linear autonomous setting, but has had less success when dealing with non-autonomous or nonlinear systems, and we will not pursue it here.

We summarise some of the above discussion in Table 1, as well as in the following principle.

PRINCIPLE 2.1 (Propagation of waves). *Suppose that a solution u solves a scalar dispersive equation* (2.1) *with initial datum u_0. If u_0 has spatial frequency roughly ξ_0 (in other words, the Fourier transform $\hat{u}_0$ is concentrated near ξ_0), then $u(t)$ will have spatial frequency roughly ξ_0 for all times, and u will also oscillate in time with frequency roughly $h(\xi_0)$. In physical space, u will travel with velocity roughly $-\nabla h(\xi_0)$. These heuristics are only valid to accuracies consistent with the spatial and frequency uncertainty of u; the wave is initially coherent, but as time progresses, the frequency uncertainty (and hence velocity uncertainty) overwhelms the spatial uncertainty, leading to dispersion. (In particular, the uncertainty principle itself is a mechanism for dispersion over time.)*

Thus for instance, fields of frequency ξ_0 will propagate at velocities $\hbar\xi/v$ under the Schrödinger evolution, $-3\xi^2$ for the Airy evolution, and $c\xi/|\xi|$ for the wave evolution, subject to limitations given by the uncertainty principle. See Exercise 2.4 for a partial verification of this principle. For the wave equation, thi principle suggests that waves propagate at the speed of light c; we shall expand upon this *finite speed of propagation* property in Section 2.5. One can also use techniques from oscillatory integrals, in particular the method of stationary phase, to make rigorous formulations of the above principle, but we will prefer to leave it as an informal heuristic. The situation for systems (as opposed to scalar equations) can be more complicated; one usually has to decompose $h(\xi)$ (which is now an operator) into eigenspaces, and thus decompose the wave of frequency ξ into *polarised* components, each of which can propagate in a different direction. We will not discuss this here.

[9]One could also work with various truncated variants of the Fourier transform, such as cosine bases or the Laplace transform. However, the advantages of such tailored bases are minor,

TABLE 1. Some different perspectives to analyzing an evolution equation, and the coordinates used in that analysis. This list is not intended to be exhaustive, and there is of course overlap between the various perspectives.

Approach	Sample tools
Spatial variable (x)	Elliptic theory, gauge fixing, coercivity estimates
Causality (t)	Gronwall, bootstraps, Duhamel, time subdivision
Physical space (x, t)	Integration by parts, substitutions, fund. solution
Spacetime geom. (x^α)	Vector fields, null surfaces, conformal maps
Frequency space (ξ, t)	Littlewood-Paley theory, multipliers, paraproducts
Spectral theory (x, τ)	Spectral measure, resolvents, eigenfunctions
Spacetime freq. (ξ, τ)	$X^{s,b}$ spaces, dispersion relation, null structure
Phase space (x, ξ, t)	Bicharacteristics, pseudodifferential operators, FIOs
Geom. optics (x, ξ, t, τ)	Eikonal and Hamilton-Jacobi equations, WKB
Hamiltonian ($\mathcal{D}$, t)	Noether's theorem, nonsqueezing, normal forms
Lagrangian (u)	Stress-energy tensor, symmetries, variational methods

EXERCISE 2.17 (Translation operators). Show that for any $x_0 \in \mathbf{R}^d$, the propagator $\exp(-x_0 \cdot \nabla)$ is the operation of translation by x_0, thus $\exp(-x_0 \cdot \nabla) f(x) = f(x - x_0)$. Compare this with Taylor's formula in the case that f is real analytic.

EXERCISE 2.18 (Wave propagators). Using the spatial Fourier transform, show that if $u \in C^2_{t,\mathrm{loc}} \mathcal{S}_x(\mathbf{R} \times \mathbf{R}^d)$ is a field obeying the wave equation (2.9) with $c = 1$, then

$$\widehat{u(t)}(\xi) = \cos(t|\xi|)\hat{u}_0(\xi) + \frac{\sin(t|\xi|)}{|\xi|}\hat{u}_1(\xi)$$

for all $t \in \mathbf{R}$ and $\xi \in \mathbf{R}^d$; one can also write this as

$$u(t) = \cos(t\sqrt{-\Delta})u_0 + \frac{\sin(t\sqrt{-\Delta})}{\sqrt{-\Delta}}u_1$$

or using the spacetime Fourier transform as

$$\tilde{u}(\tau, \xi) = \delta(|\tau| - |\xi|)(\frac{1}{2}\hat{u}_0(\xi) + \frac{\mathrm{sgn}(\tau)}{2i|\xi|}\hat{u}_1(\xi))$$

(note that some care is required to ensure that the product of $\delta(\tau - |\xi|)$ and $\mathrm{sgn}(\tau)$ actually makes sense). Thus the characteristic hypersurface for the wave equation is the light cone $\{(\tau, \xi) : |\tau| = |\xi|\}$. As these formulae make sense for any distributions u_0, u_1, we shall refer to the function or distribution $u(t)$ generated by these formulae as *the* solution to the wave equation with this specified data. Show that if $(u_0, u_1) \in H^s_x(\mathbf{R}^d) \times H^{s-1}_x(\mathbf{R}^d)$ for some $s \in \mathbf{R}$, then we have $u \in C^0_t H^s_x(\mathbf{R} \times \mathbf{R}^d) \cap C^1_t H^{s-1}_x(\mathbf{R} \times \mathbf{R}^d)$, and in fact we have the bounds

$$\|\nabla u(t)\|_{H^{s-1}_x(\mathbf{R}^d)} + \|\partial_t u(t)\|_{H^{s-1}_x(\mathbf{R}^d)} \lesssim_{d,s} \|u_0\|_{H^s_x(\mathbf{R}^d)} + \|u_1\|_{H^{s-1}_x(\mathbf{R}^d)})$$

and

$$\|u(t)\|_{H^s_x(\mathbf{R}^d)} \lesssim_{d,s} \langle t \rangle (\|u_0\|_{H^s_x(\mathbf{R}^d)} + \|u_1\|_{H^{s-1}_x(\mathbf{R}^d)})$$

and are usually outweighed by the loss of algebraic structure and established theory incurred by abandoning the Fourier transform.

for all times $t \in \mathbf{R}$. Thus the solution u stays bounded to top order, but lower order norms of u can grow linearly in time.

EXERCISE 2.19 (Klein-Gordon propagators). If $u \in C^2_{t,\mathrm{loc}} \mathcal{S}_x(\mathbf{R} \times \mathbf{R}^d)$ is a classical solution to the Klein-Gordon equation (2.10) with $c = m = 1$, then

$$\widehat{u(t)}(\xi) = \cos(t\langle\xi\rangle)\hat{u}_0(\xi) + \frac{\sin(t\langle\xi\rangle)}{\langle\xi\rangle}\hat{u}_1(\xi)$$

for all $t \in \mathbf{R}$ and $\xi \in \mathbf{R}^d$; one can also write this as

$$u(t) = \cos(t\sqrt{1-\Delta})u_0 + \frac{\sin(t\sqrt{1-\Delta})}{\sqrt{1-\Delta}}u_1$$

or

$$\tilde{u}(\tau, \xi) = \delta(|\tau| - \langle\xi\rangle)(\frac{1}{2}\hat{u}_0(\xi) + \frac{\mathrm{sgn}(\tau)}{2i\langle\xi\rangle}\hat{u}_1(\xi)).$$

Thus the characteristic hypersurface here is the two-sheeted hyperboloid $\{(\tau, \xi) : |\tau|^2 - |\xi|2 = 1\}$. Again we extend this formula to distributions to define the notion of a distributional solution to the Klein-Gordon equation. Show that if $(u_0, u_1) \in H^s_x(\mathbf{R}^d) \times H^{s-1}_x(\mathbf{R}^d)$, then we have $u \in C^0_t(\mathbf{R} \to H^s(\mathbf{R}^d)) \cap C^1_t(\mathbf{R} \to H^{s-1}_x(\mathbf{R}^d))$, and in fact we have the bounds

$$\|u(t)\|_{H^s_x(\mathbf{R}^d)} + \|\partial_t u(t)\|_{H^{s-1}_x(\mathbf{R}^d)} \lesssim_{d,s} \|u_0\|_{H^s_x(\mathbf{R}^d)} + \|u_1\|_{H^{s-1}_x(\mathbf{R}^d)}$$

for all times $t \in \mathbf{R}$. Thus the Klein-Gordon equation has slightly better regularity behaviour in (inhomogeneous) Sobolev spaces than the wave equation.

EXERCISE 2.20 (Geometry of characteristic surfaces). Interpret Exercises 2.3-2.11 using either the spatial Fourier transform or the spacetime Fourier transform, viewing each of these exercises as an assertion that one dispersion relation can be approximated by or transformed into another. Several of the exercises should correspond to a specific geometric approximation or transformation (e.g. approximating a cubic by a quadratic or a polynomial by a tangent, or a hyperboloid by two paraboloids; or by relating a cone to its conic sections). The exercises involving vector equations such as Dirac's equation are a little trickier to interpret this way, as one also needs to analyze the eigenspaces of the dispersion relation $h(\xi)$.

EXERCISE 2.21 (Duhamel formula). Let I be a time interval. Suppose that $u \in C^1_t \mathcal{S}_x(I \times \mathbf{R}^d)$ and $F \in C^0_t \mathcal{S}_x(I \times \mathbf{R}^d)$ solve the equation $\partial_t u = Lu + F$, where $L = ih(D)$ is skew-adjoint. Establish the *Duhamel formula*

$$u(t) = e^{(t-t_0)L}u(t_0) + \int_{t_0}^t e^{(t-s)L}F(s)\, ds \tag{2.13}$$

for all $t_0, t \in I$ (compare with (1.48)).

EXERCISE 2.22 (Wave Duhamel formula). Let I be a time interval. Suppose that $u \in C^2_t \mathcal{S}_x(I \times \mathbf{R}^d)$ and $F \in C^0_t \mathcal{S}_x(I \times \mathbf{R}^d)$ are fields such that $\Box u = F$. Establish the Duhamel formula

$$u(t) = \cos((t-t_0)\sqrt{-\Delta})u(t_0) + \frac{\sin((t-t_0)\sqrt{-\Delta})}{\sqrt{-\Delta}}\partial_t u(t_0) - \int_{t_0}^t \frac{\sin((t-s)\sqrt{-\Delta})}{\sqrt{-\Delta}}F(s)\, ds \tag{2.14}$$

forall $t_0, t \in I$ (compare with Exercise 1.52).

EXERCISE 2.23 (Invariant energy). Show that if $u \in C^\infty_{t,\mathrm{loc}}\mathcal{S}_x(\mathbf{R}^{1+d})$ is a classical solution to the wave equation with $c = 1$, then the *invariant energy* $C(t) := \|u(t)\|^2_{\dot H^{1/2}_x(\mathbf{R}^d)} + \|\partial_t u(t)\|^2_{\dot H^{-1/2}_x(\mathbf{R}^d)}$ is independent of the choice of time t. (Hint: use the spatial Fourier transform, and either the explicit solution for u or differentiate in time. One can also use the method sketched below.) Furthermore, it is also invariant under the Lorentz transformation in Exercise 2.6. (This is rather tricky. One way is to show that u can be expressed in the form

$$u(t,x) = \sum_\pm \int_{\mathbf{R}^d} f_\pm(\xi) e^{ix\cdot\xi \pm it|\xi|}\ \frac{d\xi}{|\xi|}$$

for some locally integrable functions $f_+(\xi), f_-(\xi)$, and that for any such representation we have

$$C(t) = \frac{1}{(2\pi)^d} \sum_\pm \int_{\mathbf{R}^d} |f_\pm(\xi)|^2 \frac{d\xi}{|\xi|}.$$

Then apply the Lorentz transform to u and see what that does to $f_\pm$. The measure $\frac{d\xi}{|\xi|}$ on the light cone $\tau^2 - |\xi|^2 = 0$ can also be interpreted more invariantly as the Dirac measure $2\delta(\tau^2 - |\xi|^2)$.)

EXERCISE 2.24 (Illposedness of Schrödinger in $C^\infty_{x,\mathrm{loc}}$). Give an example of a smooth solution $u \in C^\infty_{t,x,\mathrm{loc}}(\mathbf{R}\times\mathbf{R} \to \mathbf{C})$ to the Schrödinger equation $i\partial_t u + \partial_x^2 u = 0$ which vanishes on the upper half-plane $t \geq 0$ but is not identically zero. (Hint: Find a function $f(z)$ which is analytic[10] on the first quadrant $\{\mathrm{Re}(z), \mathrm{Im}(z) \geq 0\}$, which decays fast enough on the boundary γ of this quadrant (which is a contour consisting of the positive real axis and upper imaginary axis) so that $\int_\gamma f(z) e^{ixz+itz^2}\ dz$ converges nicely for any $x \in \mathbf{R}$ and $t \in \mathbf{R}$, is equal to zero for $t \geq 0$ by Cauchy's theorem, but is such that $\int_\gamma f(z) e^{itz^2}\ dz$ is not identically zero. For the latter fact, you can use one of the uniqueness theorems for the Fourier transform.) This shows that one no longer has a satisfactory uniqueness theory once we allow our solutions to grow arbitrarily large (so that the theory of tempered distributions no longer applies); thus infinite speed of propagation, combined with extremely large reserves of "energy" at infinity, can conspire to destroy uniqueness even in the smooth category. Show that such an example does not exist if we replace $C^\infty_{t,x,\mathrm{loc}}$ by $C^\infty_{t,x}$ (so that all derivatives of u are uniformly bounded).

EXERCISE 2.25. Let $u \in C^0_t H^1_x(\mathbf{R}^{1+d}) \cap C^1_t L^2_x(\mathbf{R}^{1+d})$ be an energy class solution to the wave equation with $c = 1$. Show that for any bounded time interval I we have the bound

$$\|\int_I u(t,x)\ dt\|_{\dot H^2_x(\mathbf{R}^d)} \lesssim_d \|u(0)\|_{\dot H^1_x(\mathbf{R}^d)} + \|\partial_t u(0)\|_{L^2_x(\mathbf{R}^d)}.$$

(This can be done either by direct integration by parts in physical space, or by the spatial Fourier transform.) Thus integrating a solution in time leads to a gain of regularity in space. This phenomenon is a consequence of the oscillation of u in time, and fails if one places absolute values inside the time integral on the left-hand side; see however the Strichartz estimates discussed in Section 2.3, which are also

[10]The author thanks Jared Wunsch, John Garnett, and Dimitri Shlyakhtenko for discussions which led to this example. A key point is that the functions involved grow so fast that they are not tempered distributions (despite being smooth), and thus beyond the reach of the distributional Fourier transform.

a kind of smoothing effect formed by averaging in time. (Thanks to Markus Keel for this problem.)

2.2. Fundamental solution

The wide wings flap but once to lift him up. A single ripple starts from where he stood. (Theodore Roethke, "The Heron")

In the previous section, we have given the solution $u(t)$ to the linear dispersive equation (2.1) as a spatial Fourier multiplier applied to the initial datum u_0, see (2.11). For simplicity let us take u to be a complex scalar field and u_0 to be Schwartz. Since multiplication on the Fourier transform side intertwines with convolution on the spatial side, we thus have

$$u(t,x) = u_0 * K_t(x) = \int_{\mathbf{R}^d} u_0(x-y) K_t(y)\, dy$$

where the *fundamental solution* (or *Riemann function*) K_t is the (distributional) inverse Fourier transform of the multiplier $e^{ith(\xi)}$:

$$K_t(x) := \frac{1}{(2\pi)^d} \int_{\mathbf{R}^d} e^{i(x\cdot\xi + th(\xi))}\, d\xi. \tag{2.15}$$

One can also think of K_t as the propagator e^{tL} applied to the Dirac delta function δ.

The integral in (2.15) is not absolutely convergent, and thus does not make sense classically. However in most cases one can solve this problem by a limiting procedure, for instance writing

$$K_t(x) = \lim_{\varepsilon\to 0} \frac{1}{(2\pi)^d} \int_{\mathbf{R}^d} e^{i(x\cdot\xi + th(\xi))} e^{-\varepsilon|\xi|^2}\, d\xi. \tag{2.16}$$

The integrals on the right-hand side are absolutely convergent, and the limit often also exists in the sense of (tempered) distributions. For instance, in the case of the Schrödinger equation (2.3), with $h(\xi) = -\frac{\hbar}{2m}|\xi|^2$, we have

$$K_t(x) = e^{it\hbar\Delta/2m}\delta = \frac{1}{(2\pi i\hbar t/m)^{d/2}} e^{im|x|^2/(2\hbar t)} \tag{2.17}$$

for all $t \neq 0$, where one takes the standard branch of the complex square root with branch cut on the negative real axis (see Exercise 2.26). Thus we have the formula

$$u(t,x) = \frac{1}{(2\pi i\hbar t/m)^{d/2}} \int_{\mathbf{R}^d} e^{im|x-y|^2/(2\hbar t)} u_0(y)\, dy \tag{2.18}$$

for $t \neq 0$ and all Schwartz solutions to the free Schrödinger equation (2.3) with initial datum $u(0,x) = u_0(x)$. This simple formula is even more remarkable when one observes that it is not obvious[11] at all that $u(t,x)$ actually does converge to $u_0(x)$ in the limit $t \to 0$. It also has the consequence that the Schrödinger evolution is *instantaneously smoothing* for localised data; if u_0 is so localised as to be absolutely integrable, but is not smooth, then from (2.18) shows that at all other times $t \neq 0$, the solution $u(t)$ is smooth (but not localised). Thus the Schrödinger evolution can instantaneously trade localisation for regularity (or vice versa, since the equation is time reversible). This effect is related to the *local smoothing phenomenon*, which we

[11]Indeed, the problem of establishing pointwise convergence back to u_0 when u_0 is only assumed to lie in a Sobolev space $H^s_x(\mathbf{R}^d)$ is still a partially unsolved problem, and is considered to be quite difficult. See for instance [**Sjo**], [**Veg**].

discuss later in this chapter. It can also be explained using the heuristic of Heisenberg's law $mv = \hbar\xi$ (from Principle 2.1); the high frequencies of u travel very fast and radiate quickly away from the origin where they are initially localised, leaving only the low frequencies, which are always smooth, to remain near the origin.

REMARK 2.2. The instantaneous smoothing effect is also closely related to the *infinite speed of propagation* for the Schrödinger equation; a solution which is compactly supported at time $t = 0$ will instantly cease to be compactly supported at any later time, again thanks to (2.18). Indeed, one can heuristically argue that any equation which is both time reversible and enjoys finite speed of propagation (as well as some sort of uniqueness for the evolution), such as the wave and Klein-Gordon equations, cannot enjoy any sort of fixed-time smoothing effect (though other nontrivial fixed-time estimates may well be available).

Next we consider the fundamental solution $K_t(x)$ for the Airy equation (2.4). Here we have $h(\xi) = \xi^3$, and thus we have

$$K_t(x) = \frac{1}{2\pi}\int_{\mathbf{R}} e^{i(x\xi + t\xi^3)}\ d\xi.$$

A simple rescaling argument then shows

$$K_t(x) = t^{-1/3}K_1(x/t^{-1/3})$$

where we adopt the convention that $(-t)^{1/3} = -(t^{1/3})$. The function K_1 is essentially the *Airy function*, and is known to be bounded (see Exercise 2.30). Thus we see that $K_t = O(t^{-1/3})$. This should be compared with the decay of $O(t^{-1/2})$ which arises from the one-dimensional Schrödinger equation. The difference in decay is explained by the different dispersion relations of the two equations ($h(\xi) = \xi^3$ for Airy, $h(\xi) = \frac{1}{2}\xi^2$ for Schrödinger). From Exercise 2.4 or Principle 2.1, the relationship between group velocity and frequency for the Airy equation is $v = -3\xi^2$, as opposed to $v = \xi$ for Schrödinger. Thus high frequencies move even faster in Airy than in Schrödinger (leading to more smoothing), but low frequencies move more slowly[12] (leading to less decay).

Now we turn to the wave equation with $c = 1$, for which the situation is more complicated. First of all, as the equation is second-order in time, there are two fundamental solutions of importance, namely

$$K_t^0(x) := \cos(t\sqrt{-\Delta})\delta(x) = \frac{1}{(2\pi)^d}\int_{\mathbf{R}^d} \cos(2\pi t|\xi|)e^{ix\cdot\xi}\ d\xi$$

and

$$K_t^1(x) := \frac{\sin(t\sqrt{-\Delta})}{\sqrt{-\Delta}}\delta(x) = \frac{1}{(2\pi)^d}\int_{\mathbf{R}^d} \frac{\sin(2\pi t|\xi|)}{|\xi|}e^{ix\cdot\xi}\ d\xi$$

(see Exercise 2.18, 2.22). In principle, this is not a difficulty, as one can easily verify (e.g. using the Fourier transform) that the two solutions are related by the formula $K_t^0 = \partial_t K_t^1$. However, in this case the fundamental solutions are only distributions rather than smooth functions, because one now only has finite speed of propagation (see Section 2.5) and hence no fixed-time smoothing effects exist. Nevertheless, the above formulae do give some insight as to the character of these

[12]The reader may wonder at this point whether the vague heuristics from Principle 2.1 can be placed on a more rigorous footing. This is possible, but requires the tools of microlocal analysis, and in particular the principle of stationary phase and the wave packet decomposition. We will not attempt a detailed description of these tools here, but see for instance [**Stei2**].

distributions. Indeed one easily sees that K_t^0 and K_t^1 are radially symmetric, real, and verify the scaling rules

$$K_t^0(x) = \frac{1}{|t|^d} K_1^0(\frac{x}{t}); \quad K_t^1(x) = \frac{\mathrm{sgn}(t)}{|t|^{n-1}} K_1^1(\frac{x}{t}).$$

In the case when d is odd, one can describe K_t^0 and K_t^1 explicitly (see Exercise 2.31), and then one can then use the method of descent (as in Exercise 2.33) to obtain a formula for even dimensions d. It will however be more important for us not to have explicit formulae for these fundamental solutions, but instead to obtain good estimates on the solutions and on various smoothed out versions of these solutions. A typical estimate is as follows. Let $\phi \in \mathcal{S}_x(\mathbf{R}^d)$ be a Schwartz function, and for any $\lambda > 0$ let $\phi_\lambda := \lambda^n \phi(\lambda x)$; thus for large λ this resembles an approximation to the identity. Then we have the pointwise estimates

$$|K_t^0 * \phi_\lambda(x)| \lesssim_{\phi,d} \lambda^d \langle \lambda t \rangle^{-(d-1)/2} \tag{2.19}$$

and

$$|K_t^1 * \phi_\lambda(x)| \lesssim_{\phi,d} \lambda^{d-1} \langle \lambda t \rangle^{-(d-1)/2} \tag{2.20}$$

for all $t \in \mathbf{R}$ and $x \in \mathbf{R}^d$. These estimates can be proven via the Fourier transform and the principle of stationary phase; we shall indicate an alternate approach using commuting vector fields in Exercise 2.65. Roughly speaking, these estimates assert that K_t^0 and K_t^1 decay like $t^{-(d-1)/2}$, but only after integrating K_t^0 and K_t^1 $\frac{d+1}{2}$ times and $\frac{d-1}{2}$ times respectively.

The situation for the Klein-Gordon equation (2.10) is even more complicated; formulae for the explicit solution are in principle obtainable from the method of descent (see Exercise 2.33) but they are not particularly simple to work with. It is convenient to split the frequency domain into the *nonrelativistic region*, when $\hbar|\xi| \ll mc$, and the *relativistic region*, when $\hbar|\xi| \gg mc$. The basic rule of thumb here is that the Klein-Gordon equation behaves like the Schrödinger equation in the non-relativistic region and like the wave equation in the relativistic region. For some more precise formulations of this heuristic, see [**MN**] and the references therein.

One can also construct fundamental solutions for these equations on the torus $\mathbf{T}^d = \mathbf{R}^d/2\pi\mathbf{Z}^d$. In the case of the wave and Klein-Gordon equations, which have finite speed of propagation, there is little difference between the two domains (for short times at least). However, the fundamental solution for dispersive equations such as the Schrödinger equation become significantly more complicated to control on torii, with some very delicate number theoretic issues arising; as such, the concept of a fundamental solution has only limited importance in these settings.

Exercise 2.26 (Gaussian integrals). Use contour integration to establish the identity

$$\int_{-\infty}^{\infty} e^{-\alpha x^2} e^{\beta x}\, dx = \sqrt{\frac{\pi}{\alpha}} e^{\beta^2/4\alpha}$$

whenever α, β are complex numbers with $\mathrm{Re}(\alpha) > 0$, where one takes the standard branch of the square root. Use this and (2.16) to establish (2.17). (You may wish to first reduce to the case when $d = 1$ and $\frac{\hbar}{2m} = 1$. With a rescaling trick one can also assume $t = 1$.)

Exercise 2.27 (Schrödinger fundamental solution). Show that up to multiplication by a constant, the fundamental solution (2.17) for the Schrödinger equation is the only tempered distribution which is invariant under spatial rotations, Galilean transforms, time reversal symmetry, and the scaling $u(t,x) \mapsto \lambda^{-n} u(\frac{t}{\lambda^2}, \frac{x}{\lambda})$. This can be used to give an alternate derivation of the fundamental solution (except for the constant).

EXERCISE 2.28 (Pseudoconformal transformation). Let $u \in C^1_{t,\mathrm{loc}}\mathcal{S}_x(\mathbf{R} \times \mathbf{R}^d)$ be a complex scalar field, and define the field $v : \mathbf{R} \times \mathbf{R}^d \to \mathbf{C}$ by

$$v(t,x) := \frac{1}{(it)^{d/2}} \overline{u(\frac{1}{t}, \frac{x}{t})} e^{i|x|^2/2t},$$

with the convention

$$v(0,x) = \frac{1}{(2\pi)^{d/2}} \hat{u}_0(x).$$

Establish the identity

$$(i\partial_t v + \Delta v)(t,x) = \frac{1}{t^2} \frac{1}{(it)^{d/2}} \overline{(i\partial_t u + \Delta u)(\frac{1}{t}, \frac{x}{t})} e^{i|x|^2/2t}$$

for $t \neq 0$. In particular, show that if u solves the Schrödinger equation (2.3) with $\hbar = m = 1$, then v also solves the Schrödinger equation (even at time $t = 0$). Thus the pseudoconformal transformation manages the remarkable feat of swapping the initial datum $u_0(x)$ with its Fourier transform $\hat{u}_0(x)$ (up to constants). Also verify that the psedoconformal transformation $u \mapsto v$ is its own inverse. For an additional challenge, see if you can link this transformation to the conformal transformation in Exercise 2.14 using Exercise 2.11.

EXERCISE 2.29 (Van der Corput lemma). Let $I \subset \mathbf{R}$ be a compact interval. If $\phi \in C^2_x(I \to \mathbf{R})$ is either convex or concave, and $|\partial_x \phi(x)| \geq \lambda$ for all $x \in I$ and some $\lambda > 0$, establish the estimate $|\int_I e^{i\phi(x)}\ dx| \leq \frac{2}{\lambda}$. (Hint: write $e^{i\phi(x)} = \frac{1}{i\phi'(x)} \partial_x e^{i\phi(x)}$ and integrate by parts.) From this and induction, conclude that if $k \geq 2$ and $\phi \in C^k(I \to \mathbf{R})$ is such that $|\partial_x^k \phi(x)| \geq \lambda$ for all $x \in I$ and some $\lambda > 0$, then $|\int_I e^{i\phi(x)}\ dx| \lesssim_k \lambda^{-1/k}$. Obtain a similar estimate for integrals of the form $\int_{\mathbf{R}} e^{i\phi(x)} \psi(x)\ dx$ when ψ has bounded variation.

EXERCISE 2.30 (Airy fundamental solution). Let $K_t(x)$ be the fundamental solution of the Airy function. Establish the bounds $K_1(x) = O_N(\langle x \rangle^{-N})$ for any $N \geq 0$ and $x > 0$, and $K_1(x) = O(\langle x \rangle^{-1/4})$ for any $x \leq 0$. (Hint: when $x \geq 1$, use repeated integration by parts. When x is bounded, use van der Corput's lemma. When $x \leq -1$, split the integral up into pieces and apply van der Corput's lemma or integration by parts to each.) Explain why this disparity in decay is consistent with Principle 2.1.

EXERCISE 2.31 (Wave fundamental solution). Let $d \geq 3$ be odd and $c = 1$, and consider the fundamental solutions $K^0_t = \cos(t\sqrt{-\Delta})\delta$ and $K^1_t = \frac{\sin(t\sqrt{-\Delta})}{\sqrt{-\Delta}}\delta$ for the wave equation. Show that these distributions are given by

$$\int_{\mathbf{R}^3} K^0_t(x)\phi(x)\ dx = c_d \partial_t (\frac{1}{t}\partial_t)^{(d-3)/2} (t^{d-2} \int_{S^{d-1}} \phi(t\omega) d\omega);$$

$$\int_{\mathbf{R}^3} K^1_t(x)\phi(x)\ dx = c_d (\frac{1}{t}\partial_t)^{(d-3)/2} (t^{d-2} \int_{S^{d-1}} \phi(t\omega) d\omega)$$

for all test functions ϕ, where $d\omega$ is surface measure on the sphere S^{d-1}, and c_d is a constant depending only on d. (Hint: since K^0_t and K^1_t are radial, it suffices to establish this when ϕ is also radial. Now apply Exercise 2.12 and Exercise 2.13. Alternatively, one can compute the Fourier transforms of the distributions listed above and verify that they agree with those of K_t^0 and K_x^1. Yet another approach is to show that the expressions given above solve the wave equation with the specified initial data.) What happens in the case $d = 1$? (Use Exercise 2.13.)

EXERCISE 2.32 (Sharp Huygens' principle). Let $d \geq 3$ be odd, and let $u \in C^2_{t,\mathrm{loc}}\mathcal{S}_x(\mathbf{R} \times \mathbf{R}^d)$ be a classical solution to the wave equation (2.9), such that the initial data u_0, u_1 is supported on a closed set $\Omega \subseteq \mathbf{R}^d$. Show that for any time t, $u(t)$ is supported on the set $\Omega_t := \{x \in \mathbf{R}^d : |x - y| = ct \text{ for some } y \in \Omega\}$.

EXERCISE 2.33 (Klein-Gordon fundamental solution). Let $d \geq 1$ and $c = \hbar = 1$, let K_t^0 and K_t^1 be the fundamental solutions for the Klein-Gordon equation in d dimensions, and let $\tilde{K}_t^0$ and $\tilde{K}_t^1$ be the fundamental solutions for the wave equation in $d+1$ dimensions. Establish the distributional identities

$$K_t^j(x_1, \ldots, x_n) = \int_{\mathbf{R}} \tilde{K}_t^j(x_1, \ldots, x_n, x_{d+1}) e^{imx_{d+1}}\ dx_{d+1}$$

for $j = 0, 1$. (Hint: use Exercise 2.10.) In principle, this gives an explicit description of the fundamental solution of the Klein-Gordon equation, although it is somewhat unwieldy to work with in practice.

2.3. Dispersion and Strichartz estimates

Like as the waves make towards the pebbled shore, so do our minutes hasten to their end. (William Shakespeare, Sonnet 60)

In order to be able to perturb linear dispersive equations such as (2.1) to nonlinear dispersive equations, it is crucial that we have some efficient ways to control the size of solutions to the linear problem in terms of the size of the initial datum (or of the forcing term, if one exists). Of course, to do this, one has to quantify the notion of "size" by specifying a suitable function space norm. It turns out for semilinear equations such as NLS and NLW, the mixed Lebesgue norms $L_t^q L_x^r(I \times \mathbf{R}^d)$, and more generally the mixed Sobolev norms $L_t^q W_x^{s,r}(I \times \mathbf{R}^d)$, are particularly useful[13].

To make the discussion more concrete, let us consider the Schrödinger equation (2.3) in $\mathbf{R}^d$ with $\hbar = m = 1$, so the propagator operator is simply[14] $e^{it\Delta/2}$. We first ask what *fixed-time* estimates are available: if we fix a time $t \neq 0$, and we know the initial datum u_0 lies in some Sobolev space, what do we learn about the size of the solution $u(t) = e^{it\Delta/2}u_0$? To avoid technicalities let us take the solution to be Schwartz; the general case can usually be handled by limiting arguments.

Since $e^{it\Delta/2}$ is unitary, we obtain the L_x^2 conservation law

$$\|e^{it\Delta/2}u_0\|_{L_x^2(\mathbf{R}^d)} = \|u_0\|_{L_x^2(\mathbf{R}^d)} \tag{2.21}$$

and then (since $e^{it\Delta/2}$ commutes with other Fourier multipliers)

$$\|e^{it\Delta/2}u_0\|_{H_x^s(\mathbf{R}^d)} = \|u_0\|_{H_x^s(\mathbf{R}^d)};$$

[13]There are also Besov refinements of the Strichartz spaces which convey slightly more precise information on the distribution of the solution among low, medium, and high frequencies, as well as Lorentz refinements that extend a little more control over the distribution of large and small values in physical space, but this is a more technical topic which we will skip lightly over here.

[14]The factor of 1/2 in the exponent is not particularly important, and all the estimates in this section hold if it is omitted; we retain it for compatibility with other sections of the book.

one can also obtain this from Plancherel's identity. From the fundamental solution (2.18) and the triangle inequality we also have the *dispersive inequality*

$$\|e^{it\Delta/2}u_0\|_{L^\infty_x(\mathbf{R}^d)} \lesssim_d t^{-d/2}\|u_0\|_{L^1_x(\mathbf{R}^d)}. \tag{2.22}$$

This shows that if the initial datum u_0 has suitable integrability in space, then the evolution will have a power-type decay in time; the L^2_x mass of the solution is conserved, but is dispersed over an increasingly larger region as time progresses (see Exercise 2.34). We can interpolate this (using the Marcinkeiwicz real interpolation theorem, see e.g. [**Sad**]; one can also modify the arguments from Exercise A.5) with (2.21) to obtain the further estimate

$$\|e^{it\Delta/2}u_0\|_{L^{p'}_x(\mathbf{R}^d)} \lesssim_d t^{-d(\frac{1}{p}-\frac{1}{2})}\|u_0\|_{L^p_x(\mathbf{R}^d)} \tag{2.23}$$

for all $1 \leq p \leq 2$, where p' is the dual exponent of p, defined by the formula $\frac{1}{p}+\frac{1}{p'}=1$. These are the complete range of L^p_x to L^q_x fixed-time estimates available (see Exercise 2.35). In particular, the Schrödinger flow does not preserve any L^p_x norm other than the L^2_x norm. We can insert fractional differentiation operators as before and conclude

$$\|e^{it\Delta/2}u_0\|_{W^{s,p'}_x(\mathbf{R}^d)} \lesssim_d t^{-d(\frac{1}{p}-\frac{1}{2})}\|u_0\|_{W^{s,p}_x(\mathbf{R}^d)}$$

for all $s \in \mathbf{R}$. By using Sobolev embedding, one can trade some of the regularity on the left-hand side for integrability, however one cannot hope for any sort of smoothing effect that achieves more regularity on the left-hand side than on the left (see Exercise 2.36). This is in contrast with the smoothing effects of dissipative propagators such as the heat kernels $e^{t\Delta}$.

These decay estimates are useful in the long-time asymptotic theory of nonlinear Schrödinger equations, especially when the dimension d is large and the initial datum u_0 has good integrability properties. However in many situations, the initial data is only assumed to lie in an L^2_x Sobolev space such as $H^s_x(\mathbf{R}^d)$. Fortunately, by combining the above dispersive estimates with some duality arguments, one can obtain an extremely useful set of estimates, known as *Strichartz estimates*, which can handle this type of data:

THEOREM 2.3 (Strichartz estimates for Schrödinger). [**GV**], [**Yaj**], [**KTao**] *Fix $d \geq 1$ and $\hbar = m = 1$, and call a pair (q,r) of exponents* admissible *if $2 \leq q, r \leq \infty$, $\frac{2}{q}+\frac{d}{r}=\frac{d}{2}$ and $(q,r,d) \neq (2,\infty,2)$. Then for any admissible exponents (q,r) and $(\tilde q,\tilde r)$ we have the homogeneous Strichartz estimate*

$$\|e^{it\Delta/2}u_0\|_{L^q_t L^r_x(\mathbf{R}\times\mathbf{R}^d)} \lesssim_{d,q,r} \|u_0\|_{L^2_x(\mathbf{R}^d)} \tag{2.24}$$

the dual homogeneous Strichartz estimate

$$\|\int_{\mathbf{R}} e^{-is\Delta/2}F(s)\ ds\|_{L^2_x(\mathbf{R}^d)} \lesssim_{d,\tilde q,\tilde r} \|F\|_{L^{\tilde q'}_t L^{\tilde r'}_x(\mathbf{R}\times\mathbf{R}^d)} \tag{2.25}$$

and the inhomogeneous (or retarded) Strichartz estimate

$$\|\int_{t'<t} e^{i(t-t')\Delta/2}F(t')\ ds\|_{L^q_t L^r_x(\mathbf{R}^d)} \lesssim_{d,q,r,\tilde q,\tilde r} \|F\|_{L^{\tilde q'}_t L^{\tilde r'}_x(\mathbf{R}\times\mathbf{R}^d)} \tag{2.26}$$

The non-endpoint version of this theorem (when $q, \tilde q \neq 2$) had been established in [**GV**], [**Yaj**], and of course the original work of Strichartz [**Stri**] (which in turn had precursors in [**Seg3**], [**Tomas**]). The more delicate endpoint cases are treated in [**KTao**]. The estimates are known to fail in a number of ways at the endpoint

$(q, r, d) = (2, \infty, 2)$, see [**Mon**], although the homogeneous estimate can be salvaged if one assumes spherical symmetry [**Stef**], [**Tao2**], [**MNNO**]. The exponents in the homogeneous estimates are best possible (Exercise 2.42), but some additional estimates are available in the inhomogeneous case [**Kat8**], [**Fos**].

Because the Schrödinger evolution commutes with Fourier multipliers such as $|\nabla|^s$ or $\langle\nabla\rangle^s$, it is easy to convert the above statements into ones at regularities $H^s_x(\mathbf{R}^d)$ or $\dot H^s_x(\mathbf{R}^d)$. In particular, if $u : I \times \mathbf{R}^d \to \mathbf{C}$ is the solution to an inhomogeneous Schrödinger equation

$$i\partial_t u + \frac{1}{2}\Delta u = F; \quad u(0) = u_0 \in H^s_x(\mathbf{R}^d),$$

given by Duhamel's formula (2.13) on some time interval I containing 0, then by applying $\langle\nabla\rangle^s$ to both sides and using the above theorem, we obtain the estimates

$$\|u\|_{L^q_t W^{s,r}_x(I\times\mathbf{R}^d)} \lesssim_{d,q,\tilde q,\tilde r,s} \|u_0\|_{H^s_x(\mathbf{R}^d)} + \|F\|_{L^{\tilde q'}_t W^{s,\tilde r'}_x(I\times\mathbf{R}^d)}$$

for any admissible (q, r) and $(\tilde q, \tilde r)$, though one has to take some care with the endpoint $\tilde r = \infty$ because of issues with defining Sobolev spaces there. Similarly if we replace the Sobolev spaces H^s_x, $W^{s,r}_x$, $W^{s,\tilde r}_x$ by their homogeneous counterparts $\dot H^s_x$, $\dot W^{s,r}_x$, $\dot W^{s,\tilde r}_x$. One can also use Sobolev embedding and (if I is bounded) Hölder's inequality in time to increase the family of exponents for which Strichartz estimates are available; see Figure 1 for some examples.

We shall give a proof of this non-endpoint cases of this theorem shortly, but we first give an abstract lemma, the *Christ-Kiselev lemma* [**CKis**] which is very useful in establishing retarded Strichartz estimates. A proof of the lemma as formulated here can be found in [**SSog**] or [**Tao2**].

LEMMA 2.4 (Christ-Kiselev lemma). *Let X, Y be Banach spaces, let I be a time interval, and let $K \in C^0(I \times I \to B(X \to Y))$ be a kernel taking values in the space of bounded operators from X to Y. Suppose that $1 \le p < q \le \infty$ is such that*

$$\|\int_I K(t, s)f(s)\, ds\|_{L^q_t(I\to Y)} \le A\|f\|_{L^p_t(I\to X)}$$

for all $f \in L^p_t(I \to X)$ and some $A > 0$. Then one also has

$$\|\int_{s\in I: s<t} K(t, s)f(s)\, ds\|_{L^q_t(I\to Y)} \lesssim_{p,q} A\|f\|_{L^p_t(I\to X)}.$$

The principle that motivates this lemma is that if an operator is known to be bounded from one space to another, then any reasonable "localisation" of that operator (in this case, to the causal region $s < t$ of time interactions) should also be bounded. The hypothesis that $p < q$ is unfortunately necessary; see Exercise 2.40.

We can now prove the non-endpoint cases of the Strichartz estimate.

PARTIAL PROOF OF THEOREM 2.3. We shall only prove this theorem in the "non-endpoint cases" when $q, q' \ne 2$. In these cases we can argue using the TT^* *method* (as was carried out for the closely related restriction problem in harmonic analysis in [**Tomas**]) as follows. Let (q, r) be admissible. Applying Minkowski's inequality, (2.23) and the Hardy-Littlewood-Sobolev theorem of fractional integration

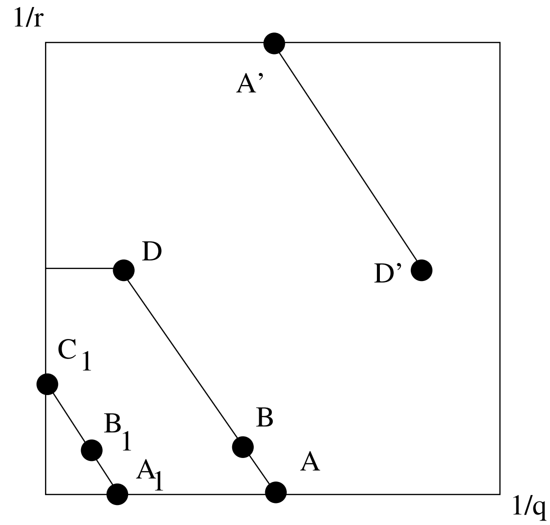

Figure 1. The Strichartz "game board" for Schrödinger equations in $\dot{H}^1_x(\mathbf{R}^3)$. The exponent pairs are $A = L^\infty_t L^2_x$, $B = L^{10}_t L^{30/13}_x$, $D = L^2_t L^6_x$, $A_1 = L^\infty_t L^6_x$, $B_1 = L^{10}_{t,x}$, $C_1 = L^4_t L^\infty_x$, $A' = L^1_t L^2_x$, $D' = L^2_t L^{6/5}_x$. If the initial datum u_0 lies in $\dot{H}^1_x(\mathbf{R}^3)$, and one derivative of the forcing term F lies in a space the closed interval between A' and D', then one derivative of the solution u lies in every space in the closed interval between A and D. Endpoint Sobolev embedding can then "move left", place the solution itself in any space between A_1 and C_1 (though the endpoint C_1 requires a Besov modification to the Strichartz spaces due to failure of endpoint Sobolev embedding; see [**CKSTT11**]). If I is bounded, Hölder in time can also "move up", lowering the r index to gain a power of $|I|$. If one is working with H^1_x instead of $\dot{H}^1_x$ (or is restricting to high frequencies), then non-endpoint Sobolev embedding is also available, allowing one to enter the pentagonal region between AD and A_1C_1. If one restricts to low frequencies, then Bernstein's inequality (A.5) can move further to the left than A_1C_1.

(see (A.10)), we conclude that

$$\begin{aligned}\| \int_{\mathbf{R}} e^{i(t-s)\Delta/2} F(s)\ ds\|_{L^q_t L^r_x(\mathbf{R}\times\mathbf{R}^d)} &\leq \| \int_{\mathbf{R}} \|e^{i(t-s)\Delta/2} F(s)\|_{L^r_x(\mathbf{R}^d)}\ ds\|_{L^q_t(\mathbf{R})} \\ &\lesssim_{d,r} \| \|F\|_{L^{r'}_x(\mathbf{R}^d)} * \frac{1}{|t|^{d(\frac{1}{2}-\frac{1}{r})}} \|_{L^q_t(\mathbf{R})} \\ &\lesssim_{d,q,r} \|F\|_{L^{q'}_t L^{r'}_x(\mathbf{R}\times\mathbf{R}^d)}\end{aligned}$$

whenever $2 < r \leq \infty$ and $2 < q \leq \infty$ are such that $\frac{2}{q} + \frac{d}{r} = \frac{d}{2}$, and for any Schwartz function F in spacetime. Applying Hölder's inequality, we conclude that

$$|\int_{\mathbf{R}} \int_{\mathbf{R}} \langle e^{i(t-s)\Delta/2} F(s), F(t)\rangle\ dsdt| \lesssim_{d,q,r} \|F\|^2_{L^{q'}_t L^{r'}_x(\mathbf{R}\times\mathbf{R}^d)},$$

where $\langle F, G\rangle = \int_{\mathbf{R}^d} F(x)\overline{G(x)}\, dx$ is the usual inner product. But the left-hand side factorises as $\| \int_{\mathbf{R}} e^{-is\Delta/2} F(s)\, ds\|_{L^2_x(\mathbf{R}^d)}^2$, and thus we obtain the dual homogeneous Strichartz estimate

$$\| \int_{\mathbf{R}} e^{-is\Delta/2} F(s)\, ds\|_{L^2_x(\mathbf{R}^d)} \lesssim_{d,q,r} \|F\|_{L^{q'}_t L^{r'}_x(\mathbf{R}\times\mathbf{R}^d)}$$

which is (2.25). By duality we then obtain (2.24). Composing those two estimates, we conclude

$$\| \int_{\mathbf{R}} e^{-is\Delta/2} F(s)\, ds\|_{L^q_t L^r_x(\mathbf{R}^d)} \lesssim_{d,q,r,\tilde q,\tilde r} \|F\|_{L^{\tilde q'}_t L^{\tilde r'}_x(\mathbf{R}\times\mathbf{R}^d)}$$

and (2.26) then follows from the Christ-Kiselev lemma. $\square$

Strichartz estimates can be viewed in two ways. Locally in time, they describe a type of smoothing effect, but reflected in a gain of integrability rather than regularity (if the datum is in L^2_x, the solution $u(t)$ is in L^r_x with $r > 2$ for most of the time), and only if one averages in time. (For fixed time, no gain in integrability is possible; see Exercise 2.35.) Looking globally in time, they describe a decay effect, that the L^r_x norm of a solution $u(t)$ must decay to zero as $t \to \infty$, at least in some L^q_t-averaged sense. Both effects of the Strichartz estimate reflect the dispersive nature of the Schrödinger equation (i.e. that different frequencies propagate in different directions); it is easy to verify that no such estimates are available for the dispersionless transport equation (2.2), except with the trivial pair of exponents $(q, r) = (\infty, 2)$.

Remark 2.5. Readers who are familiar with the uncertainty principle (Principle A.1) can interpret the Strichartz estimates for the homogeneous Schrödinger equation as follows. Consider a solution u to the homogeneous Schrödinger equation with L^2_x norm $O(1)$, and with frequency $\sim N$ (i.e. the Fourier transform is supported in the region $|\xi| \sim N$), for each time t. The uncertainty principle shows that at any given time t, the most that the solution $u(t)$ can concentrate in physical space is in a ball of radius $\sim 1/N$; the L^2_x normalisation then shows the solution can be as large as $N^{d/2}$ on this ball. However, the Strichartz estimates show (roughly speaking) that such a strong concentration effect can only persist for a set of times of measure $\sim 1/N^2$; outside of this set, the solution must disperse in physical space (compare with Proposition A.4). Note that this is also very consistent with Principle 2.1, since concentration in a ball of radius $1/N$ would induce the frequency uncertainty of $\sim N$, hence a velocity uncertainty of $\sim N$, which should disperse the ball after time $\sim 1/N^2$.

Similar Strichartz estimates can be established for any linear evolution equation which enjoys a dispersive estimate, such as the Airy equation. The wave equation also obeys a dispersive inequality, see (2.19), (2.20), but the presence of the regularizing factor ϕ_λ means that one requires some more harmonic analysis (in particular, some Littlewood-Paley theory and the theory of complex analytic interpolation) in order to utilise this estimate properly. Nevertheless, it is still possible to establish suitable Strichartz estimates for the wave equation. Indeed, we have

Theorem 2.6 (Strichartz estimates for wave equation). *Let I be a time interval, and let $u : I \times \mathbf{R}^d \to \mathbf{C}$ be a Schwartz solution to the wave equation $\Box u = F$ with*

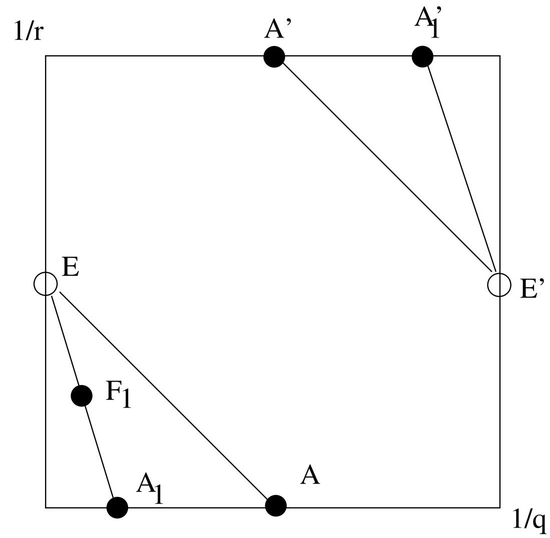

FIGURE 2. The Strichartz "game board" for wave equations in $\dot{H}^1_x(\mathbf{R}^3)$. The exponent pairs are $A = L^\infty_t L^2_x$, $E = L^2_t L^\infty_x$, $A_1 = L^\infty_t L^6_x$, $F_1 = L^4_t L^{12}_x$, $A' = L^1_t L^2_x$, $E' = L^2_t L^1_x$, $A'_1 = L^1_t L^{6/5}_x$. If the initial data (u_0, u_1) lies in $\dot{H}^1_x(\mathbf{R}^3) \times L^2_x(\mathbf{R}^3)$, and a suitable derivative of the forcing term F on a space between A' and E' (excluding E'), then a certain derivative of u lies in every space between A and E (excluding E), and so by Sobolev u itself lies in every space between A_1 and E (excluding E). At the endpoint A', no derivatives on F are required. Also by Sobolev embedding, it would have sufficed to place ∇F in any space between A'_1 and E (excluding E). Other moves, similar to those discussed in Figure 1, are available.

$c = 1$, and with initial data $u(t_0) = u_0$, $\partial_t u(t_0) = u_1$ for some $t_0 \in I$. Then we have the estimates

$$\|u\|_{L^q_t L^r_x(I \times \mathbf{R}^d)} + \|u\|_{C^0_t \dot{H}^s_x(I \times \mathbf{R}^d)} + \|\partial_t u\|_{C^0_t \dot{H}^{s-1}_x(I \times \mathbf{R}^d)}$$
$$\lesssim_{q,r,s,n} \left(\|u_0\|_{\dot{H}^s_x(\mathbf{R}^d)} + \|u_1\|_{\dot{H}^{s-1}_x(\mathbf{R}^d)} + \|F\|_{L^{\tilde{q}'}_t L^{\tilde{r}'}_x(I \times \mathbf{R}^d)}\right)$$

whenever $s \geq 0$, $2 \leq q, \tilde{q} \leq \infty$ and $2 \leq r, \tilde{r} < \infty$ obey the scaling condition

$$\frac{1}{q} + \frac{d}{r} = \frac{d}{2} - s = \frac{1}{\tilde{q}'} + \frac{d}{\tilde{r}'} - 2 \tag{2.27}$$

and the wave admissibility conditions

$$\frac{1}{q} + \frac{d-1}{2r}, \frac{1}{\tilde{q}} + \frac{d-1}{2r'} \leq \frac{d-1}{4}.$$

For a proof of this theorem, see [**Kat8**], [**GV2**], [**Kap**], [**LSog**], [**Sog**], [**SStru3**], [**KTao**]. Again, the exponents are sharp for the homogeneous version of these estimates (i.e. with $F = 0$; see Exercise 2.43) but not the inhomogeneous; see [**Har**], [**Obe**], [**Fos**] for some inhomogeneous estimates not covered by the above theorem.

The endpoint case $r = \infty$ contains a number of subtleties, see [**FW**] for some discussion. As with the Schrödinger equation, the Strichartz estimates for the wave equation encode both local smoothing properties of the equation (the estimates obtained from Strichartz lose fewer derivatives than the fixed-time estimates one would get from Sobolev embedding), as well as global decay estimates, though when the wave equation compared against the Schrödinger equation in the same dimension, the estimates are somewhat inferior; this is due to the weaker dispersion in the wave equation (different frequencies move in different directions, but not in different speeds), the finite speed of propagation, and the concentration of solutions along light cones. This estimate only controls the homogeneous Sobolev norm, but the lower order term in the inhomogeneous Sobolev norm can often be obtained by an integration in time argument.

An important special case of the wave equation Strichartz estimates is the *energy estimate*

$$\begin{aligned}\|\nabla u\|_{C_t^0 H_x^{s-1}(I\times\mathbf{R}^d)} &+ \|\partial_t u\|_{C_t^0 H_x^{s-1}(I\times\mathbf{R}^d)} \\ &\lesssim \|\nabla_x u_0\|_{H_x^{s-1}(\mathbf{R}^d)} + \|u_1\|_{H_x^{s-1}(\mathbf{R}^d)} + \|F\|_{L_t^1 H_x^{s-1}(I\times\mathbf{R}^d)}\end{aligned} \tag{2.28}$$

which can also be proven by other means, for instance via the Fourier transform (Exercise 2.45), or by using the stress energy tensor (see Exercise 2.59). This estimate has the useful feature of gaining a full degree of regularity; the forcing term F is only assumed to have $s-1$ degrees of regularity, but the final solution u has s degrees of regularity. One also has the useful variant

$$\begin{aligned}\|u\|_{C_t^0 H_x^{s}(I\times\mathbf{R}^d)} &+ \|\partial_t u\|_{C_t^0 H_x^{s-1}(I\times\mathbf{R}^d)} \\ &\lesssim_s \langle |I|\rangle (\|u_0\|_{H_x^{s}(\mathbf{R}^d)} + \|u_1\|_{H_x^{s-1}(\mathbf{R}^d)} + \|F\|_{L_t^1 H_x^{s-1}(I\times\mathbf{R}^d)}).\end{aligned} \tag{2.29}$$

The other Strichartz estimates gain fewer than one full degree of regularity, but compensates for this by improving the time and space integrability. One specific Strichartz estimate of interest (and one which was discovered in the original paper [**Stri**]) is the *conformal Strichartz estimate for the wave equation*

$$\begin{aligned}\|u\|_{L_{t,x}^{2(d+1)/(d-1)}(\mathbf{R}\times\mathbf{R}^d)} \lesssim_d &\|u(0)\|_{\dot H_x^{1/2}(\mathbf{R}^d)} + \|\partial_t u(0)\|_{\dot H_x^{-1/2}(\mathbf{R}^d)} \\ &+ \|\Box u\|_{L_{t,x}^{2(d+1)/(d+3)}(\mathbf{R}\times\mathbf{R}^d)},\end{aligned} \tag{2.30}$$

valid for $d \geq 2$.

Strichartz estimates also exist for the Klein-Gordon equation but are more difficult to state. See [**MSW**], [**Nak4**], [**MN**]; for short times one can just rely on the wave equation Strichartz estimates (treating the lower order mass term as a perturbation), while for long times it is easiest to proceed by treating the relativistic and non-relativistic regimes separately. Some Strichartz estimates are also known in periodic settings, but primarily of $L_t^4 L_x^4$ or $L_t^6 L_x^6$ type, and are proven using the spacetime Fourier transform, as one cannot exploit dispersion in these settings (as one can already see by considering plane wave solutions). See Section 2.6. More generally, Strichartz estimates are closely related to *restriction estimates* for the Fourier transform; see [**Tao11**] for a survey of the restriction phenomenon, which has been used to obtain a number of refinements (such as bilinear refinements) to the Strichartz inequalities.

EXERCISE 2.34 (Asymptotic L_x^p behaviour of Schrödinger). Let $u_0 \in \mathcal{S}_x(\mathbf{R}^d)$ be a non-zero Schwartz function whose Fourier transform $\hat u_0(\xi)$ is supported in the ball $|\xi| \leq 1$. Show that we have the estimate $|e^{it\Delta}u_0(x)| \lesssim_{N,u_0} \langle t\rangle^{-N}\langle x\rangle^{-N}$ for all times t, all $N > 0$, and all x such that $|x| \geq 5|t|$. (Hint: use rotational symmetry to make x a multiple of the unit vector e_1, thus $x = (x_1, 0, \ldots, 0)$. Then use the Fourier representation (2.12) of $e^{it\Delta}u_0(x)$, followed by repeated integration by parts in the x_1 direction.) In the region $|x| < 5|t|$, obtain the estimate $|e^{it\Delta}u_0(x)| \lesssim_{u_0} \langle t\rangle^{-d/2}$. Conclude that $\|e^{it\Delta}u_0\|_{L^p_x(\mathbf{R}^d)} \sim_{d,u_0,p} \langle t\rangle^{d(\frac{1}{p}-\frac{1}{2})}$ for all $1 \leq p \leq \infty$. (Hint: first establish this for $p = 2$, obtain the upper bound for all p, and then use Hölder's inequality to obtain the lower bound.)

Exercise 2.35 (Necessary conditions for fixed-time Schrödinger). Suppose $1 \leq p, q \leq \infty$ and $\alpha \in \mathbf{R}$ are such that the fixed-time estimate

$$\|e^{it\Delta/2}u_0\|_{L^q_x(\mathbf{R}^d)} \leq Ct^\alpha \|u_0\|_{L^p_x(\mathbf{R}^d)}$$

for all $u_0 \in \mathcal{S}_x(\mathbf{R}^d)$ and $t \neq 0$, and some C independent of t and u_0. Using scaling arguments, conclude that $\alpha = \frac{d}{2}(\frac{1}{q} - \frac{1}{p})$. Using (2.34) (and time translation invariance), conclude that $q \geq p$ and $q = p'$ (and thus $1 \leq p \leq 2$). In particular the Schrödinger operators are not bounded on any $L^p_x(\mathbf{R}^d)$ space except when $p = 2$. Indeed, a good rule of thumb is that dispersive evolution operators only preserve "L^2_x-based" spaces and no others (with the notable exception of the one-dimensional wave evolution operators, which are not truly dispersive).

Exercise 2.36 (Schrödinger Strichartz cannot gain regularity). Using Galilean invariance (Exercise 2.5), show that no estimate of the form

$$\|e^{i\Delta/2}u_0\|_{W^{s_2,q}_x(\mathbf{R}^d)} \leq C\|u_0\|_{W^{s_1,p}_x(\mathbf{R}^d)}$$

or

$$\|e^{it\Delta/2}u_0\|_{L^q_t W^{s_2,r}_x([0,1]\times\mathbf{R}^d)} \leq C\|u_0\|_{W^{s_1,p}_x(\mathbf{R}^d)}$$

can hold with C independent of u_0, unless $s_2 \leq s_1$.

Exercise 2.37 (Decay of finite energy Schrödinger solutions). Show that the admissible space $L^\infty_t L^2_x$ which appears in Theorem 2.3 can be replaced by the slightly smaller space $C^0_t L^2_x$. Similarly, if $u_0 \in H^1_x(\mathbf{R}^3)$ and $u : \mathbf{R} \times \mathbf{R}^3 \to \mathbf{C}$ is the solution to the Schrödinger equation, show that $\lim_{t\to\pm\infty} \|u(t)\|_{L^p_x(\mathbf{R}^3)}$ for $2 < p \leq 6$ and that

$$\lim_{t\to\pm\infty} \|\langle x\rangle^{-\varepsilon}u(t)\|_{L^2_x} + \|\langle x\rangle^{-\varepsilon}\nabla u(t)\|_{L^2_x} = 0$$

for any $\varepsilon > 0$ (see [**Tao8**] for an application of these estimates to nonlinear Schrödinger equations).

Exercise 2.38 (Pseudoconformal invariance of Strichartz). (Pieter Blue, private communication) Show that if q, r are Schrödinger-admissible exponents, then the space $L^q_t L^r_x(\mathbf{R} \times \mathbf{R}^d)$ is preserved under the pseudoconformal transformation (Exercise 2.28). Conclude that Theorem 2.3 is invariant under the pseudoconformal transformation. (It is also invariant under space and time translations, homogeneity, scaling, phase rotation, time reflection, spatial rotation and Galilean transformations.)

Exercise 2.39 (Conformal invariance of Strichartz). Show that the conformal Strichartz estimate (2.30) is invariant under space and time translations, homogeneity, scaling, phase rotation, time reflection, spatial rotation, Lorentz transformation (Exercise 2.6), and conformal inversion (Exercise 2.14).

EXERCISE 2.40. Show that Lemma 2.4 fails at the endpoint $p = q$, even when X and Y are scalar. (Hint: take $p = q = 2$ and consider truncated versions of the Hilbert transform $Hf(t) = p.v. \int \frac{f(s)}{t-s}\ ds$.)

EXERCISE 2.41. Using Exercise 2.30, establish some Strichartz estimates for the Airy equation. (As it turns out, when it comes to controlling equations of Korteweg-de Vries type, these estimates are not as strong as some other estimates for the Airy equation such as local smoothing and $X^{s,b}$ estimates, which we shall discuss later in this chapter.)

EXERCISE 2.42. Using the scale invariance from Exercise 2.9, show that the condition $\frac{2}{q} + \frac{d}{r} = \frac{d}{2}$ is necessary in order for (2.24). Next, by superimposing multiple time-translated copies of a single solution together (thus replacing $u(t,x)$ by $\sum_{j=1}^N u(t - t_j, x)$ for some widely separated times $t_1, \ldots, t_N$) show that the condition $q \geq 2$ is also necessary. (The double endpoint $(q, r, d) = (2, \infty, 2)$ is harder to rule out, and requires a counterexample constructed using Brownian motion; see [**Mon**].) Note that similar arguments will establish the necessity of (2.27), as well as the requirements $q, \tilde{q} \geq 2$ in Theorem 2.6.

EXERCISE 2.43 (Knapp example). Consider the solution $u : \mathbf{R}^{1+d} \to \mathbf{C}$ to the wave equation with $c = 1$ given by

$$u(t,x) := \int_{1 \leq \xi_1 \leq 2; |\xi_2|, \ldots, |\xi_n| \leq \varepsilon} e^{ix\cdot\xi} e^{it|\xi|}\ d\xi$$

where $0 < \varepsilon < 1$ is a small number. Show that $u(t,x)$ is comparable in magnitude to ε^{d-1} whenever $|t| \ll 1/\varepsilon$, $|x_1 - t| \ll \varepsilon$ and $|x_2|, \ldots, |x_n| \ll 1$. Use this to conclude that the condition $\frac{1}{q} + \frac{d-1}{2r} \leq \frac{d-1}{4}$ in Theorem 2.6. One can also obtain essentially the same example by starting with a bump function initial datum, and applying a Lorentz transform (see Exercise 2.6) with velocity $v := (1 - \varepsilon^2)e_1$; the strange rectangular region of spacetime obtained above is then explainable using the familiar phenomena of time dilation and length contraction in special relativity.

EXERCISE 2.44 (Stein example). Let $1/2 < \alpha \leq 1$, and let $g \in L_x^2(\mathbf{R}^3)$ be the function

$$g(x) := \frac{1_{B(2e_3,2)\backslash B(e_3,1)}(x)}{|x|^2 \langle \log |x| \rangle^\alpha},$$

where e_3 is the third unit vector of $\mathbf{R}^3$ and $B(x,r)$ denotes the ball of radius r centred at x. Show that $\|g\|_{L_x^2(\mathbf{R}^3)} = O_\alpha(1)$, but that the solution $u(t) := \frac{\sin(t\sqrt{-\Delta}}{\sqrt{-\Delta}} g$ to the homogeneous wave equation with initial position 0 and initial velocity g obeys $u(te_3, t) = \infty$ for all $1 \leq t \leq 2$. (Hint: use Exercise 2.31.) In particular, Theorem 2.6 fails at the endpoint $(q,r) = (2,\infty)$, a fact first observed (with a more indirect proof) in [**KM**].

EXERCISE 2.45. Prove the energy estimate (2.28) using (2.14), the Fourier transform, and Minkowski's inequality. Note that the implied constant in the $\lesssim$ notation is absolute (not depending on s or d). Then deduce (2.29).

EXERCISE 2.46 (Besov version of Strichartz estimates). For each dyadic number N, let P_N be the Littlewood-Paley multiplier at frequency N, as defined in Appendix A. By means of complex interpolation, establish the inequalities

$$\|u\|_{L_t^q L_x^r(I \times \mathbf{R}^d)} \lesssim_{q,r} \big(\sum_N \|P_N u\|_{L_t^q L_x^r(I \times \mathbf{R}^d)}^2\big)^{1/2}$$

whenever $2 \leq q, r \leq \infty$ (so in particular whenever q, r are admissible exponents for Strichartz estimates), as well as the "dual" estimate

$$(\sum_N \|P_N F\|^2_{L_t^{q'} L_x^{r'}(I \times \mathbf{R}^d)})^{1/2} \lesssim_{q,r} \|F\|_{L_t^{q'} L_x^{r'}(I \times \mathbf{R}^d)}$$

for the same range of q, r. (Note that to apply interpolation properly for the first inequality, you will need to write u as an appropriate linear operator applied to the functions $u_N = P_N u$.) By exploiting the fact that P_N commutes with the Schrödinger operator $i\partial_t + \Delta$, establish the estimate

$$(\sum_N \|P_N e^{it\Delta/2} u_0\|^2_{L_t^q L_x^r(\mathbf{R} \times \mathbf{R}^d)})^{1/2} \lesssim_{d,q,r} \|u_0\|_{L_x^2(\mathbf{R}^d)}$$

for all Schrödinger-admissible q, r. Similarly establish analogues of (2.25) and (2.26).

2.4. Conservation laws for the Schrödinger equation

> *Knowledge about life is one thing; effective occupation of a place in life, with its dynamic currents passing through your being, is another.* (William James, "The Varieties of Religious Experience")

In Hamiltonian ODE, one often encounters conserved quantities $E(t)$ of the flow such as energy, momentum, mass, and so forth. One has similar conserved quantities in Hamiltonian PDE, but they come with additional structure. Namely, the typical conserved quantity $E(t)$ that one encounters is not just a scalar quantity, but is also an integral of some density $e_0(t, x)$, thus for instance $E(t) = \int_{\mathbf{R}^d} e_0(t, x)\, dx$. The conservation of $E(t)$ can then be manifested in a more local form by the *pointwise conservation law*[15]

$$\partial_t e_0(t, x) + \partial_{x_j} e_j(t, x) = 0 \tag{2.31}$$

for some vector-valued[16] quantity e_j, which is thus the *current* associated to the density e_0. The conservation of E then follows (formally, at least) by integrating the continuity equation in space and using the divergence theorem (assuming some suitable spatial decay on the current, of course). Thus in PDE, conservation laws come in both integral and differential forms. The differential form of the conservation law is significantly more flexible and powerful than the integral formulation, as it can be *localised* to any given region of spacetime by integrating against a suitable cutoff function (or contracting against a suitable vector field). Of course, when one does so, one may no longer get a perfectly conserved quantity, but one often obtains a quantity which is almost conserved (the derivative in time is small or somehow "lower order") or perhaps monotone. Thus from a single conservation law one can generate a variety of useful estimates, which can serve to constrain the direction of propagation of a solution, or at least of various components of that solution (e.g. the high frequency components).

These conservation laws are particularly useful for controlling the long-time dispersive behaviour, or short-time smoothing behaviour, of *nonlinear* PDE, but to illustrate the concepts let us just work first with the linear analogues of these

[15] Roman indices such as j and k will be summed over the spatial indices $1, \ldots, d$ in the usual manner; Greek indices such as α and β will be summed over $0, \ldots, d$.

[16] In some cases E and e_0 are themselves vector-valued instead of scalar, in which case e_j will be tensor-valued.

PDE. To make the discussion concrete, we focus on the Schrödinger equation (2.3) with $\hbar = m = 1$, thus

$$i\partial_t u + \frac{1}{2}\Delta u = 0; \tag{2.32}$$

to avoid technicalities we consider only smooth solutions $u \in C^\infty_{t,\mathrm{loc}}\mathcal{S}_x(\mathbf{R} \times \mathbf{R}^d)$. In practice one can usually extend the estimates here to rougher solutions by limiting arguments or duality.

Before we discuss the differential form of the conservation laws, let us first recall the integral form of the conservation laws and their connection (via Noether's theorem) to various symmetries (cf. Table 1 from Chapter 1); we will justify their conservation later when considering the differential form of these laws, and also more informally in Exercise 2.47 below. The symmetry of phase rotation, $u(t,x) \mapsto e^{i\theta}u(t,x)$, leads to the scalar conserved quantity

$$M(t) := \int_{\mathbf{R}^d} |u(t,x)|^2 \; dx,$$

which is variously referred to as the *total probability*, *charge*, or *mass* in the literature. The symmetry of space translation, $u(t,x) \mapsto u(t, x - x_0)$, leads to the vector-valued conserved quantity $\vec{p}(t)$, defined in coordinates as

$$p_j(t) := \int_{\mathbf{R}^d} \mathrm{Im}(\overline{u(t,x)}\partial_{x_j} u(t,x)) \; dx,$$

which is the *total momentum*. The symmetry of time translation, $u(t,x) \mapsto u(t - t_0, x)$, leads to the conserved quantity

$$E(t) := \frac{1}{2}\int_{\mathbf{R}^d} |\nabla u(t,x)|^2 \; dx,$$

known as the *total energy* or *Hamiltonian*. The symmetry of Galilean invariance (Exercise 2.5) leads to the conserved quantity

$$\int_{\mathbf{R}^d} x|u(t,x)|^2 \; dx - t\vec{p}(t),$$

which is the normalised centre-of-mass. The pseudo-conformal symmetry (Exercise 2.28) does not directly lead to a conserved quantity, as it is a discrete symmetry rather than a continuous one. However, it can be used to conjugate some of the preceding symmetries and generate new ones. The pseudo-conformal symmetry leaves the mass invariant, as that symmetry commutes with phase rotation; it also swaps momentum conservation with the conservation of normalised centre-of-mass, as the pseudo-conformal symmetry intertwines spatial transation and Galilean invariance. The energy conjugates to the *pseudo-conformal energy*

$$\frac{1}{2}\|(x + it\nabla)u(t)\|^2_{L^2_x(\mathbf{R}^d)} \tag{2.33}$$

which is also conserved. In particular, we see from the triangle inequality and conservation of energy that

$$\|xu(t)\|_{L^2_x(\mathbf{R}^d)} \leq \|xu(0)\|_{L^2_x(\mathbf{R}^d)} + t\|\nabla u(0)\|_{L^2_x(\mathbf{R}^d)}, \tag{2.34}$$

which is an estimate establishing a kind of approximate finite speed of propagation result for finite energy solutions. See also Exercise 2.49.

Now we analyze the local versions of these conserved quantities. It is convenient to introduce the *pseudo-stress-energy tensor* $\mathrm{T}_{\alpha\beta}$, defined for $\alpha, \beta = 0, 1, \ldots, n$ and on the spacetime $\mathbf{R} \times \mathbf{R}^d$ by

$$\begin{aligned} \mathrm{T}_{00} &= |u|^2 \\ \mathrm{T}_{0j} = \mathrm{T}_{j0} &= \mathrm{Im}(\overline{u}\partial_{x_j} u) \\ \mathrm{T}_{jk} &= \mathrm{Re}(\partial_{x_j} u \overline{\partial_{x_k} u}) - \frac{1}{4}\delta_{jk}\Delta(|u|^2) \end{aligned}$$

for all $j, k = 1, \ldots, d$, where δ_{jk} is the Kronecker delta (thus it equals 1 when $j = k$ and zero otherwise). A direct computation using (2.32) then verifies (for smooth fields u, at least) the local conservation laws

$$\partial_t \mathrm{T}_{00} + \partial_{x_j} \mathrm{T}_{0j} = 0; \quad \partial_t \mathrm{T}_{j0} + \partial_{x_k} \mathrm{T}_{jk} = 0. \tag{2.35}$$

The quantity T_{00} is known as the *mass density*, the quantity $\mathrm{T}_{0j} = \mathrm{T}_{j0}$ is known as both the *mass current* and the *momentum density*, and the quantity T_{jk} is the *momentum current* or *stress tensor*. For smooth, rapidly decreasing solutions u, this verifies the conservation of mass and momentum asserted earlier, since

$$M(t) = \int_{\mathbf{R}^d} \mathrm{T}_{00}(t,x)\ dx; \quad p_j(t) = -\int_{\mathbf{R}^d} \mathrm{T}_{0j}(t,x)\ dx.$$

Conservation of energy also has a local formulation, though not one related to the pseudo-stress-energy tensor[17] $\mathrm{T}_{\alpha\beta}$ (Exercise 2.48).

By multiplying first equation in (2.35) by a smooth function $a(x)$ of at most polynomial growth, integrating in space, and then integrating by parts, we obtain the identity

$$\partial_t \int_{\mathbf{R}^d} a(x)|u(t,x)|^2\ dx = \int_{\mathbf{R}^d} \partial_{x_j} a(x) \mathrm{Im}(\overline{u}(t,x)\partial_{x_j} u(t,x))\ dx. \tag{2.36}$$

This identity can be used to control the local rate of change of mass by quantities such as the energy (see Exercises 2.50, 2.51).

One can also obtain a useful identity by differentiating (2.36) in time again and using (2.35) to obtain the useful identity

$$\begin{aligned} \partial_t^2 \int_{\mathbf{R}^d} a(x)|u(t,x)|^2\ dx &= \partial_t \int_{\mathbf{R}^d} \partial_{x_j} a(x) \mathrm{T}_{0j}(t,x)\ dx \\ &= -\int_{\mathbf{R}^d} \partial_{x_j} a(x) \partial_{x_k} \mathrm{T}_{jk}(t,x)\ dx \\ &= \int_{\mathbf{R}^d} (\partial_{x_j}\partial_{x_k} a(x)) \mathrm{T}_{jk}(t,x)\ dx \\ &= \int_{\mathbf{R}^d} (\partial_{x_j}\partial_{x_k} a(x)) \mathrm{Re}(\partial_{x_j} u \overline{\partial_{x_k} u})\ dx \\ &\quad - \frac{1}{4}\int_{\mathbf{R}^d} |u(t,x)|^2 \Delta^2 a(x)\ dx. \end{aligned} \tag{2.37}$$

This identity has some useful consequences for various values of a. For instance, letting $a(x) = 1$ we recover mass conservation, and letting $a(x) = x_j$ we recover

[17] See however Exercise 3.30.

the conservation of normalised centre of mass. Setting $a(x) = |x|^2$, we obtain the *virial identity*

$$\partial_t^2 \int_{\mathbf{R}^d} |x|^2 |u(t,x)|^2 \, dx = 2 \int_{\mathbf{R}^d} |\nabla u(t,x)|^2 \, dx = 4E(t); \tag{2.38}$$

thus the energy controls the convexity of the quantity $\int_{\mathbf{R}^d} |x|^2|u(t,x)|^2 \, dx$; compare with the classical counterpart in Example 1.31. (Actually, this identity can easily be seen to be equivalent to the pseudo-conformal conservation law after using mass and momentum conservation.) Setting[18] $a(x) = |x|$, we obtain

$$\begin{aligned} \partial_t \int_{\mathbf{R}^d} \mathrm{Im}(\overline{u}(t,x) \frac{x}{|x|} \cdot \nabla u(t,x)) \, dx = & \int_{\mathbf{R}^d} (\partial_{x_j}\partial_{x_k} a(x)) \mathrm{Re}(\partial_{x_j} u \overline{\partial_{x_k} u}) \, dx \\ & - \frac{1}{4} \int_{\mathbf{R}^d} |u(t,x)|^2 \Delta^2 a(x) \, dx. \end{aligned} \tag{2.39}$$

Now observe that

$$(\partial_{x_j}\partial_{x_k} a(x)) \mathrm{Re}(\partial_{x_j} u \overline{\partial_{x_k} u}) = |\nabla\!\!\!/ u|^2/|x|,$$

where $|\nabla\!\!\!/ u|^2 := |\nabla u|^2 - |\frac{x}{|x|} \cdot \nabla u|^2$ is the angular component of the gradient. If we specialise to three dimensions, we also have $\Delta^2 a = -4\pi\delta$. If we integrate in time, we thus obtain the *Morawetz identity*

$$\begin{aligned} &\int_{-T}^{T} \int_{\mathbf{R}^3} \frac{|\nabla\!\!\!/ u(t,x)|^2}{|x|} \, dxdt + 4\pi \int_{-T}^{T} |u(t,0)|^2 \, dt \\ &\quad = \int_{\mathbf{R}^3} \mathrm{Im}(\overline{u}(T,x) \frac{x}{|x|} \cdot \nabla u(T,x)) \, dx - \int_{\mathbf{R}^3} \mathrm{Im}(\overline{u}(-T,x) \frac{x}{|x|} \cdot \nabla u(-T,x)) \, dx \end{aligned} \tag{2.40}$$

for any time $T > 0$. Using Lemma A.10, and observing (using the Fourier transform) that $\|u(\pm T)\|_{\dot H^{1/2}_x(\mathbf{R}^3)} = \|u(0)\|_{\dot H^{1/2}_x(\mathbf{R}^3)}$, we conclude the *Morawetz estimate*

$$\int_{\mathbf{R}} \int_{\mathbf{R}^3} \frac{|\nabla\!\!\!/ u(t,x)|^2}{|x|} \, dxdt + \int_{\mathbf{R}} |u(t,0)|^2 \, dt \lesssim \|u(0)\|^2_{\dot H^{1/2}_x(\mathbf{R}^3)}. \tag{2.41}$$

This estimate can be thought of as a variant of a Strichartz estimate, obtaining some smoothing and decay near the spatial origin $x = 0$; it should be compared with Example 1.32. It has a significant advantage over the Strichartz estimate, in that it extends with only minor modification to the case of defocusing nonlinear Schrödinger equations; see Section 3.5, in which we also discuss an interaction version of the above estimate. One can also extend this three-dimensional estimate to higher dimensions. For lower dimensions there are some additional subtleties; see [**Nak2**].

Exercise 2.47. Let us formally consider $L^2_x(\mathbf{R}^d \to \mathbf{C})$ as a symplectic phase space with symplectic form $\omega(u,v) = -2\int_{\mathbf{R}^d} \mathrm{Im}(u(x)\overline{v(x)}) \, dx$. Show that the Schrödinger equation (2.3) with $\hbar = m = 1$ is then the formal Hamiltonian flow associated to the (densely defined) Hamiltonian $H(u) := \frac{1}{2}\int_{\mathbf{R}^d} |\nabla u|^2 \, dx$. Also use this flow to formally connect the symmetries and conserved quantities mentioned in this section via Noether's theorem (ignoring irrelevant constants such as factors

[18] Strictly speaking, (2.37) does not directly apply here because a is not smooth, but this can be fixed by a regularisation argument; see Exercise 2.57.

of 2). See [**Kuk3**] for a more rigorous treatment of this infinite-dimensional Hamiltonian perspective, and [**SSul**] for a Lagrangian perspective of the material in this section.

EXERCISE 2.48. Obtain a local conservation law (2.31) for the energy density $e_0 = \frac{1}{2}|\nabla u|^2$ for the Schrödinger equation.

EXERCISE 2.49 (Local smoothing from localised data). Let $u \in C^\infty_{t,\mathrm{loc}}\mathcal{S}_x(\mathbf{R} \times \mathbf{R}^d)$ be a smooth solution to the Schrödinger equation (2.32) with $\hbar = m = 1$. By using mass conservation and the pseudo-conformal conservation law, establish the bound

$$\|\nabla u(t)\|_{L^2_x(B_R)} \lesssim_d \frac{\langle R\rangle}{|t|}\|\langle x\rangle u(0)\|_{L^2_x(\mathbf{R}^d)}$$

for all $t \neq 0$ and $R > 0$, where $B_R := \{x \in \mathbf{R}^d : |x| \leq R\}$ is the spatial ball of radius R. This shows that localisation of initial data leads to a local gain of regularity (by a full derivative, in this case) at later times, together with some local decay in time.

EXERCISE 2.50 (Local near-conservation of mass). Let $u \in C^\infty_{t,\mathrm{loc}}\mathcal{S}_x(\mathbf{R} \times \mathbf{R}^d)$ be a smooth solution to the Schrödinger equation (2.32) with $\hbar = m = 1$, and with energy $E = \frac{1}{2}\|\nabla u(0)\|^2_{L^2_x(\mathbf{R}^d)}$. Show that for any $R > 0$ and $t \neq 0$ we have

$$(\int_{|x|\leq R} |u(t,x)|^2\ dx)^{1/2} \leq (\int_{|x|\leq 2R} |u(0,x)|^2\ dx)^{1/2} + O_d(\frac{E^{1/2}|t|}{R}).$$

(Hint: apply (2.36) with $a(x) = \phi^2(x/R)$, where ϕ is a bump function supported on the ball of radius 2 which equals 1 on the ball of radius 1, and obtain a differential inequality for the quantity $M(t) := (\int_{\mathbf{R}^d} \phi^2(x/R)|u(t,x)|^2\ dx)^{1/2}$.) This estimate (and the generalisation in Exercise 2.51 below) is especially useful for the energy-critical nonlinear Schrödinger equation, as the error term of $\frac{CE^{1/2}|t|}{R}$ depends only on the energy of u and not on other quantities such as the mass; see [**Gri5**], [**Bou7**], [**Bou9**], [**Tao9**] for some applications of this estimate.

EXERCISE 2.51 (Local near-conservation of mass, II). [**Bou7**], [**Bou9**] Let the notation and assumptions be as in Exercise 2.50. Prove the more general inequality

$$(\int_{|x|\leq R} |u(t,x)|^2\ dx)^{1/2} \leq (\int_{|x|\leq 2^n R} |u(0,x)|^2\ dx)^{1/2} + O_d(\frac{E^{1/2}|t|}{Rn^{1/2}})$$

for any integer $n \geq 1$; this improves the error term at the cost of enlarging the ball in the main term. (Hint: We may use time reversal symmetry to take $t > 0$. Use conservation of energy and the pigeonhole principle to locate an integer $1 \leq j \leq n$ such that $\int_0^t \int_{2^j R\leq|x|\leq 2^{j+1}R} |\nabla u(t,x)|^2\ dxdt \lesssim Et/n$. Then apply the argument used to establish Exercise 2.50, but with R replaced by $2^j R$.) This example shows that one can sometimes exploit the pigeonhole principle to ameliorate "boundary effects" caused by cutoff functions; for another instance of this idea, see [**CKSTT11**]. In this particular case it turns out that one can use Exercise 2.50 (with R replaced by $2^{n-1}R$) to replace the $n^{1/2}$ denominator by the much stronger 2^n term, however there are other situations in which the pigeonhole principle is the only way to obtain a satisfactory estimate.

EXERCISE 2.52 (Weighted Sobolev spaces). For any integer $k \geq 0$, define the *weighted Sobolev space* $H^{k,k}_x(\mathbf{R}^d)$ be the closure of the Schwartz functions under the norm

$$\|u\|_{H^{k,k}_x(\mathbf{R}^d)} := \sum_{j=0}^k \|\langle x\rangle^j u\|_{H^{k-j}_x(\mathbf{R}^d)}.$$

Establish the estimate

$$\|e^{it\Delta/2} f\|_{H^{k,k}_x(\mathbf{R}^d)} \lesssim_{k,d} \langle t\rangle^k \|f\|_{H^{k,k}_x(\mathbf{R}^d)}$$

for all Schwartz f, either using the Fourier transform or by using mass conservation laws (and higher order variants of these laws, exploiting the fact that the Schrödinger equation commutes with derivatives). (Hint: it may help to first work out the $k = 1$ case.)

EXERCISE 2.53 (Local smoothing for Schrödinger). [**Sjo**],[**Veg**],[**CS**] Let $u \in C^\infty_{t,\mathrm{loc}}\mathcal{S}_x(\mathbf{R} \times \mathbf{R}^3)$ be a smooth solution to the Schrödinger equation (2.32) with $\hbar = m = 1$. Establish the *homogeneous local smoothing estimate*

$$\int_{\mathbf{R}}\int_{\mathbf{R}^3} \langle x\rangle^{-1-\varepsilon}|\nabla u(t,x)|^2 + \langle x\rangle^{-3-\varepsilon}|u(t,x)|^2\ dxdt \lesssim_\varepsilon \|u(0,x)\|^2_{\dot H^{1/2}_x(\mathbf{R}^3)}$$

for all $\varepsilon > 0$. (Hint: One can take ε to be small. Then adapt the Morawetz argument in the text with $a(x) := \langle x\rangle - \varepsilon\langle x\rangle^{1-\varepsilon}$.) This shows a local gain of half a derivative for the homogeneous Schrödinger equation. Give an informal explanation as to why this estimate is consistent with the Heisenberg law $v = \xi$ that arises from Principle 2.1; compare it also with Example 1.33.

EXERCISE 2.54 (Local smoothing for Schrödinger, II). Let $u \in \mathbf{C}^\infty_{t,\mathrm{loc}}\mathcal{S}_x(\mathbf{R}\times\mathbf{R}^3)$ be a smooth solution to the inhomogeneous Schrödinger equation

$$i\partial_t u + \Delta u = F.$$

Establish the *dual local smoothing estimate*

$$\sup_{t\in\mathbf{R}} \|u(t)\|_{\dot H^{1/2}_x(\mathbf{R}^3)} \lesssim_\varepsilon \|u(0)\|_{\dot H^{1/2}_x(\mathbf{R}^3)} + \int_{\mathbf{R}}\int_{\mathbf{R}^3} \langle x\rangle^{1+\varepsilon}|F(t,x)|^2\ dxdt$$

for any $\varepsilon > 0$. (Hint: use Exercise 2.53 and Duhamel's formula.) Then establish the *retarded local smoothing estimate*

$$\int_{\mathbf{R}}\int_{\mathbf{R}^3} \langle x\rangle^{-1-\varepsilon}|\nabla u(t,x)|^2 + \langle x\rangle^{-3-\varepsilon}|u(t,x)|^2\ dxdt$$
$$\lesssim_\varepsilon \|u(0,x)\|^2_{\dot H^{1/2}_x(\mathbf{R}^3)} + \int_{\mathbf{R}}\int_{\mathbf{R}^3} \langle x\rangle^{1+\varepsilon}|F(t,x)|^2\ dxdt.$$

(Hint: use the same argument[19] as in Exercise 2.53. An additional interaction term between u and F will appear, and can be controlled using Cauchy-Schwarz.) This shows a local gain of a full derivative for the in homogeneous Schrödinger equation.

EXERCISE 2.55 (Local smoothing for Airy equation). [**Kat2**], [**KF**] Show that smooth solutions $u \in C^\infty_{t,\mathrm{loc}}\mathcal{S}_x(\mathbf{R} \times \mathbf{R} \to \mathbf{R})$ to the Airy equation $\partial_t u + \partial_x^3 u = 0$ obey the conservation law

$$\partial_t(u^2) = -\partial_x^3(u^2) + 3\partial_x(u_x^2) \tag{2.42}$$

[19] We thank Jared Wunsch for pointing out this simple argument.

where $u_x := \partial_x u$, and use this to conclude the *local smoothing estimate*

$$\int_0^T \int_{|x|\leq R} u_x^2 \, dxdt \lesssim \frac{T+R^3}{R^2} \int_{\mathbf{R}} u(0,x)^2 \, dx \tag{2.43}$$

for all $T > 0$ and $R > 0$. (Hint: first integrate (2.42) against the constant function 1 to establish L_x^2 conservation, and then integrate instead against a suitable cutoff function which equals 1 for $x > 2R$, zero for $x < -2R$, and increases steadily for $-R < x < R$.)

EXERCISE 2.56 (Sharp local smoothing for Airy equation). [**KPV2**] With the notation as in the preceding exercise, prove the sharper estimate

$$\int_{\mathbf{R}} u_x(t,x_0)^2 \, dt \lesssim \int_{\mathbf{R}} u(0,x)^2 \, dx$$

for any $x_0 \in \mathbf{R}$, so that the factor $\frac{T+R^3}{R^2}$ in (2.43) can be replaced with R. (Hint: use translation invariance to set $x_0 = 0$, and use the Fourier representation of $u_x(t,0)$, followed by Plancherel's theorem. Unlike (2.43), it seems difficult to establish this estimate purely using conservation law techniques, thus suggesting some limitations to that method.) Give an informal explanation as to why (2.43) is consistent with the dispersion relation $v = -3\xi^2$ that arises from Principle 2.1. What is the analogous estimate for the one-dimensional Schrödinger equation?

EXERCISE 2.57. Justify the derivation of (2.39) from (2.37) (for $C_{t,\mathrm{loc}}^\infty \mathcal{S}_x$ solutions u to the Schrödinger equation) by applying (2.37) with $a(x) := \sqrt{\varepsilon^2 + |x|^2}$ and then taking limits as $\varepsilon \to 0$. These types of regularisation arguments are quite common in the theory of linear and nonlinear PDE, and allow one to extend the validity of many formal computations well beyond the regularities that would be needed to justify them classically.

2.5. The wave equation stress-energy tensor

A man must drive his energy, not be driven by it. (William Frederick Book)

Having considered conservation laws for the Schrödinger equation, we turn to the wave equation

$$\partial^\alpha \partial_\alpha u = F, \tag{2.44}$$

where we take smooth scalar fields $u \in C_{t,\mathrm{loc}}^\infty \mathcal{S}_x(\mathbf{R}^{1+d} \to \mathbf{C})$ for simplicity. In particular u has sufficient regularity and decay at infinity to justify all integration by parts computations.

While it is possible to view the wave equation as a Hamiltonian system (see Exercise 2.58), the geometry of Minkowski space suggests that one can also alternately view the wave equation (and the associated conservation laws) in a more Lorentz-invariant way; thus our perspective here will be slightly different from that in the preceding section. Indeed, one can view the wave equation in a Lagrangian manner instead, viewing u as a formal critical point of the Lagrangian functional $\frac{1}{2}\int_{\mathbf{R}^{1+d}} \partial^\alpha u \partial_\alpha u \, dg$. This functional is formally invariant under diffeomorphic changes of variable; by considering the variation of the functional along an infinitesimal such change of variable (see Exercise 2.60) one can then construct a stress-energy tensor $\mathrm{T}^{\alpha\beta}$ for this equation, which in this case is

$$\mathrm{T}^{\alpha\beta} := \mathrm{Re}(\partial^\alpha u \overline{\partial^\beta u}) - \frac{1}{2} g^{\alpha\beta} \mathrm{Re}(\partial^\gamma u \overline{\partial_\gamma u}). \tag{2.45}$$

In coordinates with the normalisation $c = 1$, we have

$$\begin{aligned} \mathrm{T}^{00} = \mathrm{T}_{00} &= \frac{1}{2}|\partial_t u|^2 + \frac{1}{2}|\nabla u|^2 \\ \mathrm{T}^{0j} = -\mathrm{T}_{0j} &= -\mathrm{Re}(\partial_t u \overline{\partial_{x_j} u}) \\ \mathrm{T}^{jk} = \mathrm{T}_{jk} &= \mathrm{Re}(\partial_{x_j} u \overline{\partial_{x_k} u}) - \frac{\delta_{jk}}{2}(|\nabla u|^2 - |\partial_t u|^2). \end{aligned}$$

This tensor is real and symmetric. The quantity T^{00} is known as the *energy density*, the quantity T^{0j} is the *energy current* or *momentum density*, and T^{jk} is the *momentum current* or *stress tensor*. The similarity with the Schrödinger pseudo-stress-energy tensor is not accidental; see Exercise 3.30.

The tensor $\mathrm{T}^{\alpha\beta}$ is clearly symmetric, and a quick computation using (2.44) yields the divergence equation

$$\partial_\alpha \mathrm{T}^{\alpha\beta} = \mathrm{Re}((\partial_\beta u)\overline{F}). \tag{2.46}$$

Henceforth we consider the homogeneous equation $F \equiv 0$, so that T is divergence free. In coordinates, we thus have

$$\partial_t \mathrm{T}^{00} + \partial_{x_j} \mathrm{T}^{0j} = 0; \quad \partial_t \mathrm{T}^{0j} + \partial_{x_k} \mathrm{T}^{jk} = 0$$

(compare with (2.35)). This already yields conservation of the total energy

$$E[u(t)] = E(t) := \int_{\mathbf{R}^d} \mathrm{T}^{00}(t,x)\, dx$$

and the total momentum

$$p^j(t) := \int_{\mathbf{R}^d} \mathrm{T}^{0j}(t,x)\, dx,$$

assuming sufficient spatial decay of course on the solution u and its derivatives.

It turns out that these conservation laws can be nicely localised in spacetime by exploiting the *positivity property*

$$\mathrm{T}^{\alpha\beta} v_\alpha v_\beta \geq 0 \tag{2.47}$$

whenever v^α is a time-like or light-like vector (so $v^\alpha v_\alpha \leq 0$). Indeed from (2.45) we have

$$\mathrm{T}^{\alpha\beta} v_\alpha v_\beta = (v^\alpha \partial_\alpha u)^2 - \frac{1}{2}(v^\alpha v_\alpha)(\partial^\gamma u)(\partial_\gamma u)$$

which is clearly non-negative in the light-like case $v^\alpha v_\alpha = 0$, and in the timelike case one can check positivity by dividing the gradient $\partial^\alpha u$ into the component parallel to v and the component Minkowski-orthogonal to v (on which the metric g is spacelike, hence non-negative); alternatively one can use Lorentz invariance to reduce to the case[20] $v = \partial_t$. More generally we have $\mathrm{T}^{\alpha\beta} v_\alpha w_\beta \geq 0$ whenever v, w

[20] We of course equate a vector field X^α with the associated first-order differential operator $X^\alpha \partial_\alpha$ in the usual manner.

are time-like or light-like, and both future-oriented (see Exercise 3.41). To exploit this positivity, we can use Stokes' theorem to obtain the identity

$$\int_{\Sigma_1} \mathrm{T}^{\alpha\beta} X_\alpha n_\beta dS = \int_{\Sigma_0} \mathrm{T}^{\alpha\beta} X_\alpha n_\beta dS + \int_{\Sigma} \partial_\beta(\mathrm{T}^{\alpha\beta} X_\alpha) dg \tag{2.48}$$

for an arbitrary smooth vector field X_α, where Σ is an open region in spacetime bounded below by a spacelike hypersurface Σ_0 and above by a spacelike hypersurface Σ_1, n_β is the positive timelike unit normal and dS is the induced measure from the metric g (which is positive on the spacelike surfaces Σ_0, Σ_1); if Σ is unbounded then of course we need to assume suitable decay hypotheses on u or X. For instance, if $\Sigma_0 = \{(t,x) : t = 0\}$, $\Sigma_1 = \{(t,x) : t = t_1\}$ and $X = \partial_t$ for some arbitrary time $t_1 \in \mathbf{R}$ we obtain the conservation of energy $E(t_1) = E(0)$, while if we instead take $X = \partial_{x_j}$ we obtain conservation of momentum $p^j(t_1) = p^j(0)$. Now suppose that $T_* > t_1 > 0$, and we take Σ_0 to be the disk $\{(t,x) : t = 0, |x| \leq T_*\}$ and Σ_1 to be the truncated cone[21]

$$\Sigma_1 := \{(t,x) : 0 < t < t_1, |x| = T_* - t\} \cup \{(t,x) : t = t_1, |x| \leq T_* - t_1\}.$$

Setting $X = \partial_t$, we conclude the *energy flux identity*

$$\int_{|x| \leq T_* - t_1} \mathrm{T}_{00}(t_1, x)\ dx + \mathrm{Flux}_{T_*}[0, t_1] = \int_{|x| \leq T_*} \mathrm{T}_{00}(0, x)\ dx \tag{2.49}$$

where $\mathrm{Flux}_{T_*}[t_0, t_1]$ is defined for $0 \leq t_0 < t_1 < T_*$ by

$$\mathrm{Flux}_{T_*}[t_0, t_1] := \int_{t_0 < t < t_1, |x| = T_* - t} \mathrm{T}^{\alpha\beta} X_\alpha n_\beta dS.$$

Intuitively, this identity asserts that the energy at the top of the truncated cone, plus the energy flux escaping the sides of the cone, is equal to the original energy at the base of the cone. From (2.47) we see that $\mathrm{Flux}_{T_*}[0, t_1]$ is non-negative, and so we have the localised energy monotonicity formula

$$\int_{|x| \leq T_* - t_1} \mathrm{T}_{00}(t_1, x)\ dx \leq \int_{|x| \leq T_*} \mathrm{T}_{00}(0, x)\ dx$$

In particular, we have obtained the *finite speed of propagation property*: if u solves the wave equation, and vanishes on the ball $|x| \leq T_*$ at time $t = 0$, then it vanishes on the smaller ball $|x| \leq T_* - t_1$ for all times $0 < t_1 < T_*$; this reflects the well-known fact that solutions to wave equations cannot propagate faster than the speed of light (which has been normalised to $c = 1$ here). Also, from the energy flux identity and energy conservation we have

$$\mathrm{Flux}_{T_*}[0, t_1] \leq E(0); \tag{2.50}$$

thus if the solution has finite energy, then $\mathrm{Flux}_{T_*}[0, t_1]$ is monotone increasing in T and is also bounded. It therefore converges[22] to some limit as $t_1 \to T_*$; since

[21]Strictly speaking, Σ_1 is not quite spacelike, which causes dS to degenerate to zero and n_β to elongate to infinity. But the area form $n_\beta dS$ remains well defined in the limit; we omit the standard details.

[22]This innocuous statement from basic real analysis - that every monotone bounded sequence converges - is surprisingly useful in the analysis of PDE, as it allows one to convert a monotonicity formula (say for a sequence a_n) into a decay estimate (for the Cauchy differences $a_n - a_m$). The drawback is that the decay is qualitative only; there is no uniform control on the decay of the $a_n - a_m$, although one can see that these differences cannot be too large for too many disjoint intervals $[n, m]$.

$\mathrm{Flux}_{T_*}[t_0, t_1] = \mathrm{Flux}_{T_*}[0, t_1] - \mathrm{Flux}_{T_*}[0, t_0]$, we conclude in particular the *flux decay* property

$$\lim_{t_0, t_1 \to T_*} \mathrm{Flux}_{T_*}[t_0, t_1] = 0.$$

This shows that near the tip $(T_*, 0)$ of the cone, an asymptotically vanishing amount of energy will escape the sides of the cone. There is however still the possibility of *energy concentration*, in which the energy stays inside the cone and concentrates to the tip as $t \to T_*$; we shall discuss this possibility further in Section 5.1.

To exploit (2.48) and (2.47) further, it is of interest to understand the divergence[23] of vector fields of the form $\mathrm{T}^{\alpha\beta} X_\beta$. Indeed we see from (2.46) and the symmetry of T that the vector field $\mathrm{T}^{\alpha\beta} X_\beta$ has divergence

$$\partial_\alpha (\mathrm{T}^{\alpha\beta} X_\beta) = \frac{1}{2} \mathrm{T}^{\alpha\beta} \pi_{\alpha\beta} \tag{2.51}$$

where $\pi_{\alpha\beta}$ is the *deformation tensor*

$$\pi_{\alpha\beta} := \partial_\alpha X_\beta + \partial_\beta X_\alpha = \mathcal{L}_X g_{\alpha\beta}, \tag{2.52}$$

where $\mathcal{L}_X$ denotes the Lie derivative. In particular, if the vector field X is a *Killing vector field* (i.e. the diffeomorphism induced by X preserves the metric), then $\mathcal{L}_X g_{\alpha\beta} = 0$ and hence $\mathrm{T}^{\alpha\beta} X_\beta$ is divergence-free. In particular, we obtain conservation of the quantity

$$\int_{\mathbf{R}^d} \mathrm{T}^{0\beta} X_\beta(t, x)\ dx$$

For instance, the vector fields ∂_t and ∂_{x_j}, which correspond to time translation and spatial translation respectively, yield the conservation of energy and momentum respectively. If instead we take the rotation vector field $x_j \partial_{x_k} - x_k \partial_{x_j}$ for some fixed $1 \leq j < k \leq n$, which is clearly Killing, we obtain the conservation of angular momentum

$$\int_{\mathbf{R}^d} x_j \mathrm{T}^{0k}(t, x) - x_k \mathrm{T}^{0j}(t, x)\ dx = \mathrm{Re} \int_{\mathbf{R}^d} \partial_t \overline{u(t, x)} (x_k \partial_{x_j} - x_j \partial_{x_k}) u(t, x)\ dx.$$

Taking instead the Lorentz vector field $x_j \partial_t + t \partial_{x_j}$, which is also Killing, we obtain conservation of normalised centre-of-mass

$$\int_{\mathbf{R}^d} x_j \mathrm{T}^{00}(t, x) - t \mathrm{T}^{0j}(t, x)\ dx = \int_{\mathbf{R}^d} x_j \mathrm{T}^{00}(t, x)\ dx - t p_j.$$

Thus Lorentz invariance plays the role in the wave equation that Galilean invariance does for the Schrödinger equation.

Unfortunately, there are no further Killing vector fields for Minkowski space (other than taking linear combinations of the ones listed above); this is the hyperbolic analogue of the well-known fact that the only orientation-preserving isometries of Euclidean space (i.e. the rigid motions) are the translations and rotations, and combinations thereof; see Exercise 2.62. However, there is a slightly larger class of *conformal Killing vector fields*, where the deformation tensor $\pi_{\alpha\beta}$ does not vanish, but is instead a scalar multiple of $g_{\alpha\beta}$, thus $\pi_{\alpha\beta} = \Omega g_{\alpha\beta}$ for some scalar function

[23]We thank Markus Keel and Sergiu Klainerman for sharing some unpublished notes on this topic, which were very helpful in preparing the material in this section.

Ω. Inserting this and (2.45) into (2.51), and observing that $g_{\alpha\beta}g^{\alpha\beta} = (d+1)$, we conclude that for a conformal Killing field we have

$$\partial_\alpha(\mathrm{T}^{\alpha\beta}X_\beta) = -\frac{d-1}{4}\Omega\mathrm{Re}(\overline{\partial^\gamma u}\partial_\gamma u).$$

Using the equation (2.44), we can rewrite this further as

$$\partial_\alpha(\mathrm{T}^{\alpha\beta}X_\beta) = -\frac{d-1}{8}\Omega\partial^\gamma\partial_\gamma(|u|^2),$$

which rearranges as

$$\partial_\alpha P^\alpha = -\frac{d-1}{8}|u|^2\Box\Omega \tag{2.53}$$

where P^α is the vector field

$$P^\alpha := \mathrm{T}^{\alpha\beta}X_\beta + \frac{d-1}{8}(\Omega\partial^\alpha(|u|^2) - (\partial^\alpha\Omega)|u|^2).$$

This is quite close to being a conservation law; note that P^α contains terms which are quadratic in the first derivatives of u, but the divergence of P^α only contains terms of zeroth order in u.

To give some examples of (2.53), let us first consider the *Morawetz vector field* X associated to $(t^2+|x|^2)\partial_t+2tx_j\partial_{x_j}$ (i.e. $X_0 = -(t^2+|x|^2)$ and $X_j = 2tx_j$); this is the pullback of the time translation vector field $-\partial_t$ under the conformal inversion in Exercise 2.14. This vector field is easily verified to be conformal Killing with $\Omega = 4t$. Since $\Box(4t) = 0$, we thus see that the vector field

$$P^\alpha := -(t^2+|x|^2)\mathrm{T}^{\alpha 0} + 2tx_j\mathrm{T}^{\alpha j} + (d-1)t\mathrm{Re}\overline{u}\partial^\alpha u - \frac{d-1}{2}g^{\alpha 0}|u|^2$$

is divergence free. In particular we see that the *conformal energy*

$$\begin{aligned} Q[u[t],t] := \int_{\mathbf{R}^d} &(t^2+|x|^2)\mathrm{T}^{00}(t,x) - 2tx_j\mathrm{T}^{0j}(t,x) + (d-1)t\mathrm{Re}\overline{u}\partial_t u(t,x) \\ &- \frac{d-1}{2}|u(t.x)|^2\ dx \end{aligned} \tag{2.54}$$

is preserved in time, and in particular is equal to

$$Q[u[0],0] = \int_{\mathbf{R}^d} |x|^2\mathrm{T}^{00}(t,x) - \frac{d-1}{2}|u|^2(0,x)\ dx.$$

This conservation law is the analogue of the conservation of pseudo-conformal energy (2.33) for the Schrödinger equation (cf. the last part of Exercise 2.28). The quantity $Q[u[t],t]$ is not obviously non-negative, but it can eventually be rearranged using the null frame vector fields $L := \partial_t + \frac{x}{|x|}\cdot\nabla_x$, $\underline{L} := \partial_t - \frac{x}{|x|}\cdot\nabla_x$ as

$$\frac{1}{4}\int_{\mathbf{R}^d} |(t+|x|)Lu + (d-1)u|^2 + |(t-|x|)\underline{L}u + (d-1)u|^2 + 2(t^2+|x|^2)|\nabla\!\!\!/ u|^2\ dx;$$

see for instance [**Kla3**] (for a related computation, see Exercise A.16). One thus obtains the decay laws

$$\begin{aligned} &\|(t+|x|)Lu(t) + (d-1)u(t)\|_{L^2_x(\mathbf{R}^d)}, \|(t-|x|)\underline{L}u(t) + (d-1)u(t)\|_{L^2_x(\mathbf{R}^d)}, \\ &\|(t+|x|)|\nabla\!\!\!/ u|\|_{L^2_x(\mathbf{R}^d)} \lesssim \||x|\nabla_{x,t}u(0)\|_{L^2_x(\mathbf{R}^d)} \end{aligned}$$

(compare with (2.34)). On one hand, these estimates are somewhat weak compared with the decay of $t^{-(d-1)/2}$ ordinarily expected for the wave equation; the decay obtained here is more of the order of $1/t$ (although near the light cone one does not obtain any decay in the $\underline{L}$ direction). On the other hand, the argument here is extraordinarily robust and in particular extends to very rough manifolds; see [**Kla3**] for some further discussion.

Another application of (2.53) arises by taking the scaling vector field $x^\alpha \partial_\alpha$, which is conformal Killing with $\Omega = 2$. This shows that the vector field

$$\mathrm{T}^{\alpha\beta} x_\beta + \frac{d-1}{2} \mathrm{Re} \overline{u} \partial^\alpha u \tag{2.55}$$

is divergence free, and thus

$$\int_{\mathbf{R}^d} t\mathrm{T}^{00} - x_j \mathrm{T}^{0j} - \frac{d-1}{2} \mathrm{Re} \overline{u} \partial_t u \ dx$$

is a conserved quantity. This particular conservation law is not of much direct use, as the top order terms $t\mathrm{T}^{00} - x_j \mathrm{T}^{0j}$ do not have a definite sign[24]. However, if one localises this law to the cone $\{|x| \leq t\}$, then one recovers some positivity (at least to top order), and one can obtain a useful estimate, especially in the context of establishing non-concentration for the semilinear wave equation. See Section 5.1.

Another use of Killing vector fields X^α (introduced by Klainerman [**Kla**]) lies in the fact that they commute with the d'Lambertian $\Box$, in the sense that

$$\Box(X^\alpha \partial_\alpha u) = X^\alpha \partial_\alpha (\Box u)$$

for all smooth u. This can be seen by direct computation (using Exercise 2.62) or by noting that the d'Lambertian is determined entirely by the metric g, and is hence preserved by the infinitesimal diffeomorphism associated to X^α. We can iterate this and obtain

$$\Box(K_1 \ldots K_k u) = K_1 \ldots K_k \Box u$$

whenever $K_1, \ldots, K_k$ are one of the Killing first order differential operators ∂_t, ∂_{x_j}, $x_j \partial_{x_k} - x_k \partial_{x_j}$, or $t\partial_{x_j} + x_j \partial_t$. In particular, if u is a smooth solution to the wave equation (2.44), then so does $K_1 \ldots K_k u$. In particular we have energy conservation

$$E[K_1 \ldots K_k u](t) = E[K_1 \ldots K_k u](0).$$

given sufficient decay of u and its derivatives, of course.

Let us apply this to a smooth solution with initial data $u(0)$, $\partial_t u(0)$ supported in the ball $\{|x| \leq 1\}$; by finite speed of propagation we then see that for future times $t > 0$, the solution is supported in the cone $\{|x| \leq 1 + t\}$. Then we see that

$$E[K_1 \ldots K_k u](t) \lesssim_{k,u} 1 \tag{2.56}$$

for all $k \geq 0$ and all Killing vector fields $K_1, \ldots, K_k$. This can be used to establish some decay of the solution u. Let us fix a time $t \geq 1$ and look at the region $t/2 < |x| \leq 1 + t$ using polar coordinates $x = r\omega$ for some $t/2 < r \leq 1 + t$ and $\omega \in S^{d-1}$. We can then conclude from (2.56) (using the Killing vector fields ∂_{x_j} and $x_j \partial_{x_j} - x_k \partial_{x_j}$) that

$$\int_{S^{d-1}} \int_{t/2 < r \leq 1+t} |\nabla_\omega^l \partial_r^m \nabla_{x,t} u(t, r\omega)|^2 \ r^{d-1} dr d\omega \lesssim_{l,m,u} 1$$

[24]A general principle is that if the top order terms in an expression are ill-behaved, then no amount of structure arising from the lower-order terms will rescue this. Thus one should always attend to the top order terms first.

for all $l, m \geq 0$. Note that we may replace r^{d-1} by t^{d-1} since r is copmarable to t. If in particular we define the spherical energies

$$f(r) := \sum_{0 \leq l \leq d} \left(\int^{S_{d-1}} |\nabla_\omega^l \nabla_{x,t} u(t, r\omega)|^2 \, d\omega \right)^{1/2}$$

then a simple application of Cauchy-Schwarz and Minkowski's inequality yields

$$\int_{t/2 < r \leq 1+t} |\partial_r^m f(r)|^2 \, dr \lesssim_u t^{-1-d}$$

for $m = 0, 1$. Using the Poincaré inequality

$$|f(x)|^2 \lesssim \int_I |f(y)|^2 \, dy + \int_I |f'(y)|^2 \, dy$$

whenever $x \in I$ and I is an interval of length at least 1, we see that

$$|f(r)|^2 \lesssim_u t^{-1-d}$$

for all $t/2 < r \leq 1+t$. Applying the Sobolev embedding theorem on the sphere (or by using Poincaré type inequalities) we then obtain the pointwise bound

$$|\nabla_{t,x} u(t,x)| \lesssim_u t^{-(d-1)/2}$$

for all $t \geq 1$ and $t/2 < r \leq 1+t$. If we combine this with both the finite speed of propagation and some Lorentz transforms (see Exercise 2.6) to cover the interior region $r \leq t/2$, we conclude the more global estimate

$$\|\nabla_{t,x} u(t)\|_{L_x^\infty(\mathbf{R}^d)} \lesssim_u \langle t \rangle^{-(d-1)/2}. \tag{2.57}$$

This can be used for instance to establish the dispersive bounds in (2.19), (2.20) (Exercise 2.65). In fact, more precise estimates than (2.57) are available, which establish more concentration properties near the light cone; see [**Sog**] for a more detailed treatment of this method (which is especially useful for establishing global solutions of semilinear and quasilinear wave equations from small, rapidly decaying initial data).

Exercise 2.58 (Hamiltonian formulation of wave equation). Let us formally consider $\dot{H}_x^{1/2}(\mathbf{R}^d \to \mathbf{R}) \times \dot{H}_x^{-1/2}(\mathbf{R}^d \to \mathbf{R})$ as a symplectic phase space with symplectic form $\omega((u_0, u_1), (v_0, v_1)) = \int_{\mathbf{R}^d} u_0 v_1 - u_1 v_0$. Show that u is a formal solution to the wave equation (2.32) if and only if the curve $t \mapsto (u(t), \partial_t u(t))$ follows the formal Hamiltonian flow associated to the (densely defined) Hamiltonian

$$H(u_0, u_1) := \frac{1}{2} \int_{\mathbf{R}^d} |\nabla u_0|^2 + |u_1|^2 \, dx.$$

Exercise 2.59 (Energy estimate). Let $u \in C_{t,\mathrm{loc}}^\infty \mathcal{S}_x(\mathbf{R}^{1+d} \to \mathbf{C})$ be a Schwartz solution to the inhomogeneous wave equation $\Box u = F$, and let $\mathrm{T}_{\alpha\beta}$ be as above. By analyzing the energies $E(t) := \int_{\mathbf{R}^d} \mathrm{T}_{00}(t,x) \, dx$, establish the energy identity

$$\partial_t \int_{\mathbf{R}^d} \mathrm{T}_{00}(t,x) \, dx = -\int_{\mathbf{R}^d} \partial_t u(t,x) F(t,x) \, dx \tag{2.58}$$

and conclude the energy estimate (2.28) with $s = 1$. Then use the commutativity of the wave equation with Fourier multipliers to establish this estimate for general values of s. This is an example of the *energy method*; see also the proof of Proposition 3.3 for another instance of this technique.

EXERCISE 2.60 (Lagrangian derivation of stress-energy tensor). In this exercise we shall work formally, ignoring issues of differentiability or integrability, and will assume familiarity with (pseudo-)Riemannian geometry. Given any Lorentzian metric $g_{\alpha\beta}$ on $\mathbf{R}^{1+d}$, and any scalar field $u : \mathbf{R}^{1+d} \to \mathbf{R}$, define the Lagrangian functional

$$S(u,g) := \int_{\mathbf{R}^{1+d}} L(u,g)\ dg$$

where $dg = \sqrt{-\det(g)}dxdt$ is the usual measure induced by g, and $L(u,g)$ is a local quantity which is invariant under diffeomorphic changes of variable. Let X be an arbitrary vector field on $\mathbf{R}^{1+d}$, and let g_s be the deformation of the metric g for $s \in \mathbf{R}$ along the vector field X, thus

$$(g_s)_{\alpha\beta}|_{s=0} = g_{\alpha\beta}; \quad \frac{d}{ds}(g_s)_{\alpha\beta}|_{s=0} = \mathcal{L}_X g_{\alpha\beta} = \pi_{\alpha\beta},$$

where $\mathcal{L}_X$ is the Lie derivative, and $\pi_{\alpha\beta} = \nabla_\alpha X_\beta + \nabla_\beta X_\alpha$ is the deformation tensor (here ∇ denotes the Levi-Civita covariant derivative, see Exercise 6.5). Similarly, let u_s be the deformation of u along X, thus

$$u_s|_{s=0} = u; \quad \frac{d}{ds}u_s|_{s=0} = \mathcal{L}_X u = -X^\alpha \partial_\beta u.$$

As L is invariant under diffeomorphic changes of variable, we have $\frac{d}{ds}S(u_s, g_s) = 0$. Use this fact to conclude that if for fixed g, u is a critical point of the Lagrangian $S(u,g)$, then we have the integral conservation law

$$\int_{\mathbf{R}^{1+d}} \mathrm{T}^{\alpha\beta}\pi_{\alpha\beta}\ dg = 0,$$

where the stress-energy tensor $\mathrm{T}_{\alpha\beta}$ is defined by

$$\mathrm{T}_{\alpha\beta} := \frac{\partial L}{\partial g^{\alpha\beta}} - \frac{1}{2}g_{\alpha\beta}L.$$

Conclude that $\mathrm{T}^{\alpha\beta}$ is divergence-free. In the special case $L(u,g) := g^{\alpha\beta}\partial_\alpha u \partial_\beta u$, with g equal to the Minkowski metric, show that this definition of the stress-energy tensor co-incides with (2.45). See [**SStru3**] for further discussion of the Lagrangian approach to wave equations.

EXERCISE 2.61. Obtain a pointwise angular momentum conservation law for the Schrödinger equation.

EXERCISE 2.62 (Classification of Killing fields). Show that if X^α is a Killing vector field for Minkowski space $\mathbf{R}^{1+d}$, then $\partial_\alpha\partial_\beta X_\gamma = 0$. (Hint: consider various permutations of the identity $\partial_\alpha \pi_{\beta\gamma} = 0$.) Conclude that the only Killing vector fields are the linear combinations of the ones given in the text.

EXERCISE 2.63 (Uniqueness of classical solutions). Show that for any smooth initial data $u_0 \in C^\infty_{x,\mathrm{loc}}(\mathbf{R}^d)$, $u_1 \in C^\infty_{x,\mathrm{loc}}(\mathbf{R}^d)$ there exists a unique smooth solution $u \in C^\infty_{t,x,\mathrm{loc}}(\mathbf{R}^{1+d})$ to the wave equation (2.44) with initial data $u(0,x) = u(x)$, $\partial_t u(0,x) = u_1(x)$. Note that we are not assuming any boundedness or decay at infinity. (Hint: use finite speed of propagation, and the existence and uniqueness theory in the Schwartz class.) This should be contrasted with the breakdown of uniqueness in the smooth category $C^\infty_{x,\mathrm{loc}}$ in the case of infinite speed of propagation, see Exercise 2.24.

EXERCISE 2.64 (Morawetz inequality for wave equation). Let $u \in C_t^\infty \mathcal{S}_x(\mathbf{R}^{1+3})$ be a real-valued smooth solution to the wave equation (2.32) with finite energy $E(t) = E < \infty$. By contracting the stress-energy tensor against the radial vector field X associated $\frac{x}{|x|} \cdot \nabla_x$, obtain the identity

$$\partial_\alpha(\mathrm{T}^{\alpha\beta}X_\beta) = \frac{|\nabla\!\!\!/ u|^2}{|x|} - \frac{1}{2|x|}\Box(|u|^2)$$

and then conclude the *Morawetz inequality*

$$\int_{\mathbf{R}^{1+3}} \frac{|\nabla\!\!\!/ u(t,x)|^2}{|x|}\ dxdt + \int_{\mathbf{R}} |u(t,0)|^2\ dt \lesssim E.$$

(Hint: multiply the previous identity by a smooth cutoff in time to a large interval $[-T, T]$, then integrate by parts, and use the energy bound to control all error terms. You will have to deal with the singularity at $x = 0$ in some fashion (e.g. by approximating $|x|$ by $(\varepsilon^2 + |x|^2)^{1/2}$, or by removing a small neighbourhood of the origin. You may need Hardy's inequality, Lemma A.2.) Compare this with (2.41).

EXERCISE 2.65. Use (2.57) to prove (2.19), (2.20). (Hint: first use (2.57) and a scaling argument to establish (2.19), (2.20) when ϕ is itself the derivative of a compactly supported bump function, then use translation invariance to replace "compactly supported bump function" with "Schwartz function". Finally, use some form of dyadic decomposition (e.g. Littlewood-Paley decomposition) to handle the general case.)

EXERCISE 2.66. Obtain a conserved stress-energy tensor for the Klein-Gordon equation, which collapses to the one given for the wave equation above when $c = 1$ and $m = 0$, and collapses instead to the pseudo-stress-energy tensor given for the Schrödinger equation in the limit $c \to \infty$ and $\hbar = m = 1$, using the connection in Exercise 2.7.

EXERCISE 2.67. Obtain conserved stress-energy tensors for the Maxwell and abelian Yang-Mills equations with $c = 1$, in such a way that the conserved energies are $\frac{1}{2}\int_{\mathbf{R}^3} |E|^2 + |B|^2\ dx$, and $\frac{1}{2}\int_{\mathbf{R}^d} |F_{0i}|^2 + |F_{ij}|^2$ respectively. For the Dirac equation with $m = 0$, show that the rank three stress-energy tensor

$$\mathrm{T}^\lambda_{\alpha\beta} := \{\partial_\alpha u, \gamma^\lambda \partial_\beta u\} - \frac{1}{2} g_{\alpha\beta}\{\partial_\mu u, \gamma^\lambda \partial^\mu u\}$$

is divergence-free in all three variables. Is there an analogue of this tensor in the massive case $m \neq 0$?

EXERCISE 2.68 (Equipartition of energy). Suppose that u is a $C^\infty_{t,\mathrm{loc}}\mathcal{S}_x$ solution to the Klein-Gordon equation $\Box u = m^2 u$, thus the energy $E := \int_{\mathbf{R}^d} \frac{1}{2}|\nabla u|^2 + \frac{1}{2}|\partial_t u|^2 + \frac{m^2}{2}|u|^2\ dx$ is conserved in time. By considering the time derivative of the expression $\int_{\mathbf{R}^d} u(t,x)\partial_t u(t,x)\ dx$, establish the estimate

$$\int_I \int_{\mathbf{R}^d} |\partial_t u|^2 - |\nabla u|^2 - m^2|u|^2\ dxdt = O(E/m)$$

for arbitrary time intervals I. Thus in a time-averaged sense, the energy in E will be equally split between the kinetic component $\frac{1}{2}\int_{\mathbf{R}^d} |\partial_t u|^2\ dx$ and the potential component $\frac{1}{2}\int_{\mathbf{R}^2} |\nabla_x u|^2 + m^2|u|^2$.

EXERCISE 2.69 (Carleman inequality). Let $u \in C^\infty_x(\mathbf{R}^d \to \mathbf{R})$ be a compactly supported solution to the Poisson equation $\Delta u = F$, and define the *stress-energy tensor*

$$\mathrm{T}^{\alpha\beta} := \partial^\alpha u \partial^\beta u - \frac{1}{2} g^{\alpha\beta} \partial^\gamma u \partial_\gamma u = \partial^\alpha u \partial^\beta u - \frac{1}{4} g^{\alpha\beta} \Delta(u^2) + \frac{1}{2} g^{\alpha\beta} u F$$

where $g^{\alpha\beta}$ now denotes the Euclidean metric on $\mathbf{R}^d$ instead of a Minkowski metric. Establish the divergence identity

$$\partial_\alpha \mathrm{T}^{\alpha\beta} = F \partial^\beta u$$

(cf. (2.46)). Now contract this against the vector field $e^{2tx_j}\partial_{x_j}$ for some $j = 1, \ldots, d$ and $t \neq 0$ and integrate by parts to obtain the identity

$$\int_{\mathbf{R}^d} 2t |\partial_{x_j}(e^{tx_j} u)|^2 = -\int_{\mathbf{R}^d} e^{tx_j} F \partial_{x_j}(e^{tx_j} u)$$

and conclude the *Carleman-type inequality*

$$\|\partial_{x_j}(e^{tx_j} u)\|_{L^2_x(\mathbf{R}^d)} \leq \frac{1}{2|t|} \|e^{tx_j} F\|_{L^2_x(\mathbf{R}^d)}.$$

Conclude the following *unique continuation property*: if u is a scalar or vector field on $\mathbf{R}^d$ which is smooth and compactly supported with $\Delta u = O(|u|)$, then u vanishes identically. (Hint: if u is compactly supported, then $\|e^{tx_j} u\|_{L^2_x}$ can be controlled by a bounded multiple of $\|\partial_{x_j}(e^{tx_j} u)\|_{L^2_x}$. Now let $t \to \pm\infty$.) This shows, for instance, that Schrödinger operators $-\Delta + V$ with smooth bounded potentials V cannot have any compactly supported eigenfunctions (bound states); we shall also use it to show a rigidity property of harmonic maps in Exercise 6.40. For a different proof, see Exercise B.6.

2.6. $X^{s,b}$ spaces

I dreamed a thousand new paths... I woke and walked my old one.
(Chinese proverb)

Let us now return to a general scalar constant-coefficient dispersive linear equation

$$\partial_t u = Lu$$

where $L = ih(\nabla/i)$ for some real-valued polynomial h; again, the vector-valued case is more complicated and will not be treated here. As discussed in Section 2.1, the spacetime Fourier transform $\tilde{u}$ of solutions to this equation will be supported on the hypersurface $\{(\tau, \xi) : \tau = h(\xi)\}$. If one then localises the solution in time (for instance by multiplying u by some smooth cutoff function $\eta(t)$), then the uncertainty principle (or the intertwining of multiplication and convolution by the Fourier transform) then suggests that the Fourier transform $\widetilde{\eta u}$ will be concentrated in the region $\{(\tau, \xi) : \tau = h(\xi) + O(1)\}$.

Now consider a nonlinear perturbation of the above equation, such as

$$\partial_t u = Lu + N(u).$$

At first glance one may expect the presence of the nonlinearity to distort the Fourier support of the solution substantially, so that $\tilde{u}$ or $\widetilde{\eta u}$ now has a substantial portion which lies far away from the characteristic hypersurface $\tau = h(\xi)$. Certainly one has a significant distortion if one does not localise in time first (indeed, the nonlinear solution need not even exist globally in time). However, if one applies a suitably

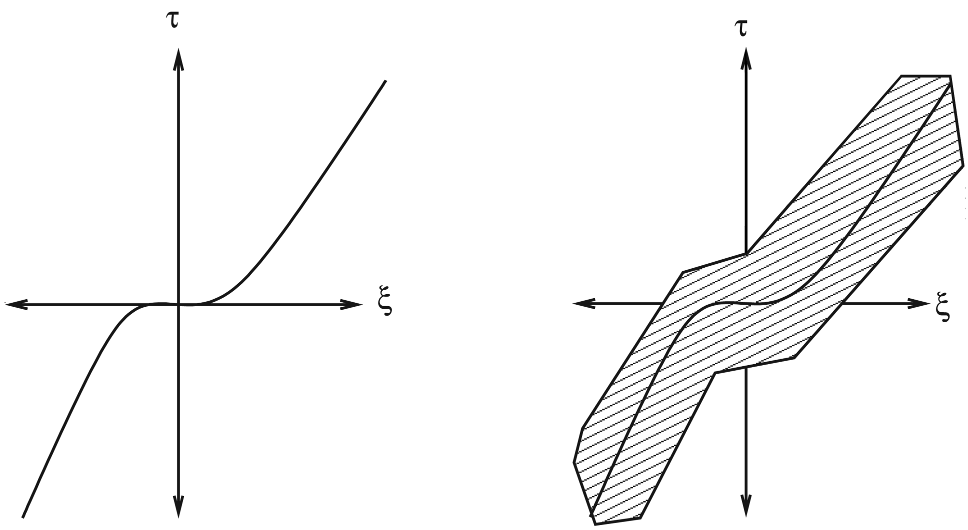

FIGURE 3. A solution to a linear dispersive equation (such as the Airy equation $\partial_t u + \partial_{xxx} u = 0$) will have its spacetime Fourier transform concentrated perfectly on the characteristic surface $\tau = h(\xi)$. Solutions to nonlinear perturbations of that dispersive equation (such as the KdV equation $\partial_t u + \partial_{xxx} u = 6u\partial_x u$) will typically, after localisation in time, have spacetime Fourier transform supported *near* the characteristic surface; thus the nonlinearity does not significantly alter the spacetime Fourier "path" of the solution, at least for short times. The $X^{s,b}$ spaces are an efficient tool to capture this clustering near the characteristic surface.

short time cutoff η, then it turns out that for many types of nonlinearities u, and for surprisingly rough classes $H^s_x(\mathbf{R}^d)$ of initial data, the localised Fourier transform $\widetilde{\eta u}$ still concentrates near the characteristic hypersurface. The reason for this is a "dispersive smoothing effect" for the operator $\partial_t - L$ away from the hypersurface $\tau = h(\xi)$, which can be viewed as the analogue of the more familiar "elliptic regularity" phenomenon for elliptic equations (if $Lu = f$ and L is elliptic, then u is smoother than f).

There are a number of ways to capture this dispersive smoothing effect, but one particularly convenient way is via the $X^{s,b}$-*spaces* (also known as *Fourier restriction spaces*, *Bourgain spaces*, or *dispersive Sobolev spaces*). The full name of these spaces[25] is $X^{s,b}_{\tau=h(\xi)}(\mathbf{R} \times \mathbf{R}^d)$, thus these spaces take $\mathbf{R} \times \mathbf{R}^d$ as their domain and are adapted to a single characteristic hypersurface $\tau = h(\xi)$. Roughly speaking, these spaces are to dispersive equations as Sobolev spaces are to elliptic equations. In a standard Sobolev space $H^s_x(\mathbf{R}^d)$, one can differentiate the function using the elliptic derivative $\langle \nabla \rangle$ s times and still remain square-integrable; for the space $X^{s,b}(\mathbf{R} \times \mathbf{R}^d)$, one can differentiate s times using the elliptic derivative $\langle \nabla \rangle$ and

[25]The terminology $H^{s,\theta} = H^{s,\theta}_{\tau=h(\xi)}(\mathbf{R}^{1+d})$ is also occasionally used in the literature, as these spaces resemble product Sobolev spaces.

b times using the dispersive derivative $\partial_t - L$, and still remain square-integrable. The precise definition is as follows.

DEFINITION 2.7 ($X^{s,b}$ spaces). Let $h : \mathbf{R}^d \to \mathbf{R}$ be a continuous function, and let $s, b \in \mathbf{R}$. The space $X^{s,b}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{R}^d)$, abbreviated $X^{s,b}(\mathbf{R}\times\mathbf{R}^d)$ or simply $X^{s,b}$ is then defined to be the closure of the Schwartz functions $\mathcal{S}_{t,x}(\mathbf{R}\times\mathbf{R}^d)$ under the norm

$$\|u\|_{X^{s,b}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{R}^d)} := \|\langle\xi\rangle^s\langle\tau - h(\xi)\rangle^b\tilde{u}(\tau,\xi)\|_{L^2_\tau L^2_\xi(\mathbf{R}\times\mathbf{R}^d)}.$$

These spaces in their modern form were introduced by Bourgain [**Bou**], although they appear in the context of one-dimensional wave equations in earlier work of Beals [**Bea**] and Rauch-Reed [**RR**], and implicitly in the work of Klainerman and Machedon [**KM**]. A good survey of these spaces and their applications can be found in [**Gin**]. Multilinear estimates for these spaces were systematically studied in [**Tao4**].

In the case $b = 0$, the choice of dispersion relation $\tau = h(\xi)$ is irrelevant, and the $X^{s,b}$ space is simply the space $L^2_t H^s_x$, as can be seen by an application of Plancherel's theorem in time. In the case $h = 0$, the $X^{s,b}$ space becomes the product space $H^b_t H^s_x$, and for general h it is a conjugate of this space (Exercise 2.70). The spatial domain $\mathbf{R}^d$ can be replaced with other abelian groups such as the torus $\mathbf{T}^d$ with minimal modification (just as Sobolev spaces and similarly be defined for the torus), indeed we have

$$\|u\|_{X^{s,b}_{\tau=h(k)}(\mathbf{R}\times\mathbf{T}^d)} := \|\langle k\rangle^s\langle\tau - h(\xi)\rangle^b\tilde{u}(\tau,k)\|_{L^2_\tau l^2_k(\mathbf{R}\times\mathbf{Z}^d)}$$

where $\tilde{u}$ is the spatially periodic, temporally non-periodic Fourier transform

$$\tilde{u}(\tau,k) := \frac{1}{(2\pi)^d}\int_{\mathbf{R}}\int_{\mathbf{T}^d} u(t,x)e^{-i(t\tau+k\cdot\xi)}\ dxdt.$$

Most of the results stated here for the non-periodic setting will extend without any difficulty to the periodic setting; we leave the verification of these details to the reader.

The spaces $X^{s,b}_{\tau=h(\xi)}$ are well adapted to the solutions $u(t) = e^{tL}u(0)$ of the linear dispersive equation $\partial_t u = Lu$, where $L := ih(D) = ih(\nabla/i)$, as the following lemma shows:

LEMMA 2.8 (Free solutions lie in $X^{s,b}$). *Let $f \in H^s_x(\mathbf{R}^d)$ for some $s \in \mathbf{R}$, and let $L = ih(\nabla/i)$ for some polynomial $h : \mathbf{R}^d \to \mathbf{R}$. Then for any Schwartz time cutoff $\eta \in \mathcal{S}_x(\mathbf{R})$, we have*

$$\|\eta(t)e^{tL}f\|_{X^{s,b}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{R}^d)} \lesssim_{\eta,b} \|f\|_{H^s_x(\mathbf{R}^d)}.$$

PROOF. A computation shows that the spacetime Fourier transform of $\eta(t)e^{tL}f$ at (τ,ξ) is simply $\hat{\eta}(\tau - h(\xi))\hat{f}(\xi)$. Since $\hat{\eta}$ is rapidly decreasing, the claim follows. $\square$

We now discuss the basic properties of the $X^{s,b}$ spaces. We first observe the easily verified fact that the $X^{s,b}$ spaces are Banach spaces. We have the trivial nesting

$$X^{s',b'}_{\tau=h(\xi)} \subset X^{s,b}_{\tau=h(\xi)}$$

whenever $s' \leq s$ and $b' \leq b$. From Parseval's identity and Cauchy-Schwarz we have the duality relationship

$$(X^{s,b}_{\tau=h(\xi)})^* = X^{-s,-b}_{\tau=-h(-\xi)}.$$

Also, these spaces interpolate nicely in the s and b indices, as can be seen using the Stein complex interpolation theorem (see e.g. [**Stei2**]). These two facts can save some effort when proving certain estimates regarding the $X^{s,b}$ spaces, particularly the multilinear estimates.

Now we study the invariance and stability properties of these spaces. The $X^{s,b}$ spaces are invariant under translations in space and time, but they are usually not invariant under frequency modulations (e.g. multiplication by a spatial phase $e^{ix\cdot\xi}$ or a temporal phase $e^{it\tau}$). The behaviour under complex conjugation is given by the identity

$$\|\overline{u}\|_{X^{s,b}_{\tau=-h(-\xi)}} = \|u\|_{X^{s,b}_{\tau=h(\xi)}}$$

and thus one has conjugation invariance when h is odd, but not necessarily otherwise.

When $b > 1/2$, one can view the $X^{s,b}$ spaces as being very close to free solutions (i.e. solutions to the equation $\partial_t u = Lu$). This is formalised in the following lemma:

LEMMA 2.9. *Let $L = iP(\nabla/i)$ for some polynomial $P : \mathbf{R}^d \to \mathbf{R}$, let $s \in \mathbf{R}$, and let Y be a Banach space of functions on $\mathbf{R} \times \mathbf{R}^d$ with the property that*

$$\|e^{it\tau_0} e^{tL} f\|_Y \lesssim \|f\|_{H^s_x(\mathbf{R}^d)}$$

for all $f \in H^s_x(\mathbf{R}^d)$ and $\tau_0 \in \mathbf{R}$. Then we have the embedding

$$\|u\|_Y \lesssim_b \|u\|_{X^{s,b}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{R}^d)}.$$

Conversely, free solutions will lie in $X^{s,b}$ once suitably truncated in time; see Lemma 2.11.

PROOF. By Fourier inversion we have

$$u(t,x) = \frac{1}{(2\pi)^{d+1}} \int_{\mathbf{R}} \int_{\mathbf{R}^d} \tilde{u}(\tau,\xi) e^{it\tau} e^{ix\cdot\xi}\ d\xi d\tau.$$

If we write $\tau = h(\xi) + \tau_0$, and set

$$f_{\tau_0}(x) := \frac{1}{(2\pi)^d} \int_{\mathbf{R}^d} \tilde{u}(h(\xi)+\tau_0,\xi) e^{ix\cdot\xi}\ d\xi$$

we have

$$e^{tL} f_{\tau_0}(x) = \frac{1}{(2\pi)^d} \int_{\mathbf{R}^d} \tilde{u}(h(\xi)+\tau_0,\xi) e^{ith(\xi)} e^{ix\cdot\xi}\ d\xi$$

and thus have the representation

$$u(t) = \frac{1}{2\pi} \int_{\mathbf{R}} e^{it\tau_0} e^{tL} f_\tau\ d\tau_0.$$

Taking Y norms and using Minkowski's inequality and the hypothesis on Y, we obtain

$$\|u\|_Y \lesssim \int_{\mathbf{R}} \|f_{\tau_0}\|_{H^s_x(\mathbf{R}^d)}\ d\tau_0,$$

and hence by Cauchy-Schwarz and the hypothesis $b > 1/2$

$$\|u\|_Y \lesssim_b \Big(\int_{\mathbf{R}} \langle\tau_0\rangle^{2b} \|f_{\tau_0}\|^2_{H^s_x(\mathbf{R}^d)}\ d\tau_0\Big)^{1/2}.$$

Using Plancherel's theorem, the right-hand side rearranges to equal $C_b\|u\|_{X^{s,b}}$, and the claim follows. □

Observe that the same argument applies when P is merely a continuous function rather than a polynomial, though in this case L will be a Fourier multiplier rather than a differential operator. Applying this to $Y = C^0_t H^s_x$, we obtain the immediate corollary

Corollary 2.10. *Let $b > 1/2$, $s \in \mathbf{R}$, and $h : \mathbf{R}^d \to \mathbf{R}$ be continuous. Then for any $u \in X^{s,b}_{\tau=h(\xi)}(\mathbf{R} \times \mathbf{R}^d)$ we have*

$$\|u\|_{C^0_t H^s_x(\mathbf{R}\times\mathbf{R}^d)} \lesssim_b \|u\|_{X^{s,b}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{R}^d)}.$$

Furthermore, the $X^{s,b}$ spaces enjoy the same Sobolev embeddings that free solutions to the equation $u_t = Lu$ do. For instance, by combining Lemma 2.9 with (2.24) (and observing that the spaces $L^q_t L^r_x$ are invariant under multiplication by phases such as $e^{it\tau_0}$), one concludes that

$$\|u\|_{L^q_t L^r_x(\mathbf{R}\times\mathbf{R}^d)} \lesssim_{q,r,b} \|u\|_{X^{0,b}_{\tau=|\xi|^2}(\mathbf{R}\times\mathbf{R}^d)}$$

for all Schrödinger-admissible (q, r).

It turns out that the $X^{s,b}$ spaces are only well suited to analyzing nonlinear dispersive equations when one localises in time. Fortunately, these spaces are easy to localise:

Lemma 2.11 ($X^{s,b}$ is stable wrt time localisation). *Let $\eta \in \mathcal{S}_t(\mathbf{R})$ be a Schwartz function in time. Then we have*

$$\|\eta(t)u\|_{X^{s,b}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{R}^d)} \lesssim_{\eta,b} \|u\|_{X^{s,b}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{R}^d)}$$

for any $s, b \in \mathbf{R}$, any $h : \mathbf{R}^d \to \mathbf{R}$, and any field $u \in \mathcal{S}_{t,x}(\mathbf{R} \times \mathbf{R}^d)$. Furthermore, if $-1/2 < b' \leq b < 1/2$, then for any $0 < T < 1$ and $\sigma > 0$ we have

$$\|\eta(t/T)u\|_{X^{s,b'}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{R}^d)} \lesssim_{\eta,b,b'} T^{b-b'}\|u\|_{X^{s,b}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{R}^d)}.$$

The second estimate in this lemma is useful in the large data theory, as it allows one to keep certain $X^{s,b}$ norms of a solution small by localizing to a sufficiently small time interval.

Proof. Let us first understand how the $X^{s,b}$ spaces behave with respect to temporal frequency modulation $u(t, x) \mapsto e^{it\tau_0}u(t, x)$. From the crude estimate

$$\langle\tau - \tau_0 - h(\xi)\rangle^b \lesssim_b \langle\tau_0\rangle^{|b|}\langle\tau - h(\xi)\rangle^b$$

and elementary Fourier analysis, we conclude that

$$\|e^{it\tau_0}u\|_{X^{s,b}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{R}^d)} \lesssim_b \langle\tau_0\rangle^{|b|}\|u\|_{X^{s,b}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{R}^d)}.$$

If we now use Fourier expansion in time to write $\eta(t) = \int_{\mathbf{R}} \hat\eta(\tau_0) e^{it\tau_0}\ d\tau_0$ and use Minkowski's inequality, we conclude

$$\|\eta(t)u\|_{X^{s,b}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{R}^d)} \lesssim_b \left(\int_{\mathbf{R}} |\hat\eta(\tau_0)|\langle\tau_0\rangle^{|b|}\ d\tau_0\right)\|u\|_{X^{s,b}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{R}^d)}.$$

Since η is Schwartz, $\hat\eta$ is rapidly decreasing, and the first claim follows.

Now we prove the second claim. By conjugating by $\langle\nabla\rangle^s$ we may take $s = 0$. By composition it suffices to treat the cases $0 \leq b' \leq b$ or $b' \leq b \leq 0$; by duality we may then take $0 \leq b' \leq b$. By interpolation with the trivial case $b' = b$ we may take $b' = 0$, thus we are now reduced to establishing

$$\|\eta(t/T)u\|_{L^2_t L^2_x(\mathbf{R}\times\mathbf{R}^d)} \lesssim_{\eta,b} T^b\|u\|_{X^{0,b}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{R}^d)}$$

for $0 < b < 1/2$. By partitioning frequency space we can divide into two cases, one where $\tilde u$ is supported on the region $\langle\tau - h(\xi)\rangle \geq 1/T$, and one where $\langle\tau-h(\xi)\rangle \leq 1/T$. In the former case we will have

$$\|u\|_{X^{0,0}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{R}^d)} \leq T^b\|u\|_{X^{0,b}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{R}^d)}$$

and the claim then follows from the boundedness of η. In the latter case, we use a variant of Corollary 2.10, noting for any time t that

$$\begin{aligned}
\|u(t)\|_{L^2_x(\mathbf{R}^d)} &\lesssim \|\widehat{u(t)}(\xi)\|_{L^2_\xi(\mathbf{R}^d)}\\
&\lesssim \left\|\int_{\langle\tau-h(\xi)\rangle\leq 1/T} |\tilde u(\tau,\xi)|\ d\tau\right\|_{L^2_\xi(\mathbf{R}^d)}\\
&\lesssim_b T^{b-1/2}\left\|\left(\int\langle\tau-h(\xi)\rangle^{2b}|\tilde u(\tau,\xi)|^2\ d\tau\right)^{1/2}\right\|_{L^2_\xi(\mathbf{R}^d)}\\
&= T^{b-1/2}\|u\|_{X^{s,b}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{R}^d)}
\end{aligned}$$

thanks to Plancherel, the triangle inequality, Cauchy-Schwarz, and the localisation of $\tilde u$. Integrating this against $\eta(t/T)$, the claim follows. $\square$

The $X^{s,b}$ spaces react well to Fourier multipliers, in much the same way that ordinary Sobolev spaces H^s do. If D^k is a Fourier multiplier of order k, in the sense that

$$\widehat{D^k f}(\xi) = m(\xi)f(\xi)$$

for all Schwartz functions $f \in \mathcal{S}_x(\mathbf{R}^d)$ and some measurable multiplier $m : \mathbf{R}^d \to \mathbf{C}$ obeying the growth condition $|m(\xi)| \lesssim \langle\xi\rangle^k$, then D can be extended to spacetime functions by acting on each time separately, thus $(Du)(t) = D(u(t))$, or in terms of the spacetime Fourier transform

$$\widetilde{D^k u}(\tau,\xi) = m(\xi)\tilde u(\tau,\xi)$$

and then one easily verifies that D maps $X^{s,b}_{\tau=h(\xi)}$ continuously to $X^{s-k,b}_{\tau=h(\xi)}$ for any $s, b \in \mathbf{R}$ and any h:

$$\|D^k u\|_{X^{s-k,b}_{\tau=h(\xi)}} \lesssim \|u\|_{X^{s,b}_{\tau=h(\xi)}}.$$

This is analogous to the well-known estimate $\|D^k u\|_{H^{s-k}} \lesssim \|u\|_{H^s}$ for Sobolev spaces. In the case when k is a non-negative integer, we have the converse

$$\|u\|_{X^{s,b}_{\tau=h(\xi)}} \lesssim \|u\|_{X^{s-k,b}_{\tau=h(\xi)}} + \|\nabla^k_x u\|_{X^{s-k,b}_{\tau=h(\xi)}}, \tag{2.59}$$

which is proven by repeating the above arguments. Similarly, if $h : \mathbf{R}^d \to \mathbf{R}$ is a polynomial and $L := ih(\nabla/i)$, then we have

$$\widetilde{(\partial_t - L)u}(\tau, \xi) = i(\tau - h(\xi))\tilde{u}(\tau, \xi)$$

and hence

$$\|(\partial_t - L)u\|_{X^{s,b-1}_{\tau=h(\xi)}} \lesssim \|u\|_{X^{s,b}_{\tau=h(\xi)}}.$$

It is natural to ask whether there is a converse inequality, in the spirit of (2.59). This is indeed the case:

PROPOSITION 2.12 ($X^{s,b}$ energy estimate). *Let $h : \mathbf{R}^d \to \mathbf{R}$ be a polynomial, let $L := ih(\nabla/i)$, and let $u \in C^\infty_{t,\mathrm{loc}}\mathcal{S}_x(\mathbf{R}\times\mathbf{R}^d)$ be a smooth solution to the equation $u_t = Lu + F$. Then for any $s \in \mathbf{R}$ and $b > 1/2$, and any compactly supported smooth time cutoff $\eta(t)$, we have*

$$\|\eta(t)u\|_{X^{s,b}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{R}^d)} \lesssim_{\eta,b} \|u(0)\|_{H^s_x(\mathbf{R}^d)} + \|F\|_{X^{s,b-1}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{R}^d)}.$$

PROOF. To abbreviate the notation we shall write $X^{s,b}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{R}^d)$ simply as $X^{s,b}$. Let $[-R, R]$ be a time interval containing the support of η. By truncating F smoothly in time, using a compactly supported cutoff that equals 1 on $[-R, R]$ we may assume (using Lemma 2.11) that F is supported on $[-2R, 2R]$ and is Schwartz in spacetime. Also, by applying $\langle\nabla\rangle^s$ to both u and F if necessary, we may take $s = 0$.

Let us first suppose that u vanishes at time $-2R$, thus $u(-2R) = 0$. By Duhamel's formula (2.13) (with $t_0 = -2R$) we thus have

$$\eta(t)u(t) = \eta(t)\int_{-\infty}^t e^{(t-s)L}F(s)\ ds = \eta(t)\int_{\mathbf{R}} \tilde{\eta}(t-s)1_{[0,+\infty)}(t-s)e^{(t-s)L}F(s)\ ds$$

where $\tilde{\eta}$ is a smooth compactly supported function which equals 1 on $[-3R, 3R]$. By Lemma 2.11, it would thus suffice to show that

$$\|\int_{\mathbf{R}} \tilde{\eta}(t-s)1_{[0,+\infty)}(t-s)e^{(t-s)L}F(s)\ ds\|_{X^{0,b}} \lesssim_{\tilde{\eta},b} \|F\|_{X^{0,b-1}}.$$

A routine computation shows that the spacetime Fourier transform of $\int_{\mathbf{R}} \tilde{\eta}(t-s)1_{[0,+\infty)}(t-s)e^{(t-s)L}F(s)\ ds$ at (τ, ξ) is equal to

$$(\int_{\mathbf{R}} \tilde{\eta}(t)1_{[0,+\infty)}(t)e^{-it(\tau-h(\xi))}dt)\tilde{F}(\tau, \xi).$$

The expression inside the parentheses can be shown (via integration by parts) to be at most $O_{\tilde{\eta}}(\langle\tau - \xi\rangle^{-1})$. The claim then follows.

Now we handle the general case. We split $u(t) = (u(t) - e^{(t+2R)L}u(-2R)) + e^{tL}e^{2RL}u(-2R)$. For the first term, the preceding argument applies, and we have

$$\|\eta(t)(u(t) - e^{(t+2R)L}u(-2R))\|_{X^{0,b}(\mathbf{R}\times\mathbf{R}^d)} \lesssim_{\eta,b} \|F\|_{X^{0,b}}.$$

Thus it will suffice to control the remaining term. Applying Lemma 2.8, it thus suffices to show that

$$\|e^{2RL}u(-2R)\|_{L^2_x} \lesssim_{\tilde{\eta},b} \|u(0)\|_{L^2_x} + \|F\|_{X^{0,b-1}}.$$

From Duhamel's formula and the support of F we have

$$e^{2RL}u(-2R) = u(0) + \int_{\mathbf{R}} \tilde{\eta}(s)1_{(-\infty,0]}(s)e^{-sL}F(s)\ ds$$

where $\tilde{\eta}$ is as before. Thus by the triangle inequality it suffices to show that

$$\| \int_{\mathbf{R}} \tilde{\eta}(s) 1_{(-\infty,0]}(s) e^{-sL} F(s)\ ds \|_{L^2_x} \lesssim_{\tilde{\eta},b} \|F\|_{X^{0,b-1}}.$$

Applying Parseval's identity, the left-hand side can be written as

$$\| \int_{\mathbf{R}} (\int_{\mathbf{R}} \tilde{\eta}(s) 1_{(-\infty,0]}(s) e^{is(\tau - h(\xi))}\ ds) \tilde{F}(\tau,\xi)\ d\tau \|_{L^2_\xi}.$$

An integration by parts yields the bound

$$\int_{\mathbf{R}} \tilde{\eta}(s) 1_{(-\infty,0]}(s) e^{is(\tau - h(\xi))}\ ds| \lesssim_\eta \langle \tau - h(\xi) \rangle^{-1}$$

and hence by Cauchy-Schwarz and the hypothesis $b > 1/2$ we have

$$| \int_{\mathbf{R}} (\int_{\mathbf{R}} \tilde{\eta}(s) 1_{(-\infty,0]}(s) e^{is(\tau - h(\xi))}\ ds) \tilde{F}(\tau,\xi)\ d\tau| \lesssim_{\eta,b} | \int_{\mathbf{R}} \langle \tau - h(\xi \rangle^{2(b-1)} |\tilde{F}(\tau,\xi)|^2\ d\tau)^{1/2},$$

and the claim follows. □

As observed earlier, $X^{s,b}$ spaces enjoy all the Strichartz estimates that free solutions do. However, in some cases, particularly in periodic settings, it is not always easy to obtain Strichartz estimates, as dispersive inequalities are typically not available in periodic settings. (When the domain is compact, L^∞_x decay is inconsistent with L^2_x conservation.) However, if one is interested in $L^4_{t,x}$ or $L^6_{t,x}$ type inequalities, one can sometimes establish the Strichartz estimate by a direct Fourier-analytic approach. A typical result, which is of application to the periodic Schrödinger equation[26] is as follows.

Proposition 2.13 (Periodic Schrödinger estimate). [**Bou**] *We have*

$$\|u\|_{L^4_t L^4_x(\mathbf{R}\times\mathbf{T})} \lesssim \|u\|_{X^{0,3/8}_{\tau=k^2}(\mathbf{R}\times\mathbf{T})}.$$

for any $u \in \mathcal{S}_{t,x}(\mathbf{R}\times\mathbf{T})$.

Proof. We use an argument of Nikolay Tzvetkov. Split $u = \sum_M u_M$, where M ranges over integer powers of 2, and u_M is the portion of u localised to the spacetime frequency region $2^M \leq \langle \tau - k^2 \rangle < 2^{M+1}$. From Plancherel's theorem we have

$$\sum_M M^{3/4} \|u_M\|^2_{L^2_t L^2_x(\mathbf{R}\times\mathbf{T})} \lesssim \|u\|^2_{X^{0,3/8}_{\tau=k^2}(\mathbf{R}\times\mathbf{T})}$$

Squaring both sides of the desired inequality and using the triangle inequality, we reduce to proving that

$$\sum \sum_{M \leq M'} \|u_M u_{M'}\|_{L^2_t L^2_x(\mathbf{R}\times\mathbf{T})} \lesssim \sum_M M^{3/4} \|u_M\|^2_{L^2_t L^2_x(\mathbf{R}\times\mathbf{T})}.$$

Setting $M' = 2^m M$, it thus suffices by the triangle inequality to prove that

$$\sum_M \|u_M u_{2^m M}\|_{L^2_t L^2_x(\mathbf{R}\times\mathbf{T})} \lesssim 2^{-\varepsilon m} \sum_M M^{3/4} \|u_M\|^2_{L^2_t L^2_x(\mathbf{R}\times\mathbf{T})}$$

[26]Depending on the choice of normalisation used for the Schrödinger equation, the dispersion relation $\tau = h(k)$ may differ from $\tau = k^2$ by an absolute constant, but this makes no difference to this Strichartz estimate. Note however that for bilinear estimates one needs to distinguish the $X^{s,b}$ space associated to $\tau = k^2$ from the conjugate $X^{s,b}$ space, associated with $\tau = -k^2$.

for all $m \geq 0$ and some absolute constant $\varepsilon > 0$; by Cauchy-Schwarz it thus suffices to establish that

$$\|u_M u_{2^m M}\|_{L^2_t L^2_x(\mathbf{R}\times\mathbf{T})} \lesssim 2^{-\varepsilon m} M^{3/8} \|u_M\|_{L^2_t L^2_x(\mathbf{R}\times\mathbf{T})} (2^m M)^{3/8} \|u_{2^m M}\|_{L^2_t L^2_x(\mathbf{R}\times\mathbf{T})}.$$

Let us now normalise u_M and $u_{2^m M}$ to have $L^2_t L^2_x$ norm one. We use Plancherel and reduce to showing

$$\|\sum_{k_1+k_2=k} \int_{\tau_1+\tau_2=\tau} \tilde u_M(\tau_1,k_1)\tilde u_{2^m M}(\tau_2,k_2)\ d\tau_1\|_{L^2_\tau l^2_k(\mathbf{R}\times\mathbf{Z})} \lesssim 2^{(3/8-\varepsilon)m} M^{3/4}.$$

On the other hand, from the normalisation and Fubini's theorem we have

$$\|(\sum_{k_1+k_2=k} \int_{\tau_1+\tau_2=\tau} |\tilde u_M(\tau_1,k_1)|^2 |\tilde u_{2^m M}(\tau_2,k_2)|^2\ d\tau_1)^{1/2}\|_{L^2_\tau l^2_k(\mathbf{R}\times\mathbf{Z})} = 1$$

so by Cauchy-Schwarz and the support of $\tilde u_M$, $\tilde u_{2^m M}$ it will suffice to show that

$$\sum_{k_1+k_2=k} \int_{\tau_1+\tau_2=\tau;\tau_1=k_1^2+O(M);\tau_2=k_2^2+O(2^m M)} 1\ d\tau_1 \lesssim 2^{(3/4-2\varepsilon m)} M^{3/2}$$

for all k, τ.

Fix k, τ. Observe that for the integral to be non-empty, we must have $\tau = k_1^2 + k_2^2 + O(2^m M)$, in which case the integral is $O(M)$. Thus it suffices to show that

$$\sum_{k_1+k_2=k;\tau=k_1^2+k_2^2+O(2^m M)} 1 \lesssim 2^{(3/4-2\varepsilon)m} M^{1/2}.$$

But if $\tau = k_1^2+k_2^2+O(2^m M)$ and $k_1+k_2 = k$, then $(k_1-k_2)^2 = 2\tau - k + O(2^m M)$, and hence $k_1 - k_2$ is constrained to at most two intervals of length $O(2^{m/2} M^{1/2})$. The claim then follows with $\varepsilon = 1/8$. $\square$

In Section 4.1 we shall encounter some bilinear and trilinear $X^{s,b}$ estimates in a spirit similar to the above (see also the exercises below).

Exercise 2.70 ($X^{s,b}$ vs. product Sobolev spaces). Let $u \in S(\mathbf{R}\times\mathbf{R}^d)$ be a complex field and let $h : \mathbf{R}^d \to \mathbf{R}$ be a polynomial. Let $U(t) := \exp(ith(\nabla/i))$ be the linear propagators for the equation $u_t = Lu$, where $L = ih(\nabla/i)$. Let $v : \mathbf{R}\times\mathbf{R}^d \to \mathbf{C}$ be the function $v(t) := U(-t)u(t)$, thus v is constant in time if and only if u solves the equation $u_t = Lu$. Show that

$$\|v\|_{X^{s,b}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{R}^d)} = \|u\|_{H^b_t H^s_x(\mathbf{R}\times\mathbf{R}^d)}$$

for all $s, b \in \mathbf{R}$.

Exercise 2.71 (Endpoint $X^{s,b}$ spaces). Show that Lemma 2.9, Corollary 2.10, Lemma 2.11, and Proposition 2.12 all break down at the endpoint $b = 1/2$. (But see the next exercise.)

Exercise 2.72 (Endpoint $X^{s,b}$ spaces, II). Let $h : \mathbf{Z}^d \to \mathbf{R}$, and let s, b be real numbers. Define the space $Y^{s,b}_{\tau=h(k)}(\mathbf{R}\times\mathbf{T}^d)$ to be the closure of the Schwartz functions under the norm

$$\|u\|_{Y^{s,b}_{\tau=h(k)}(\mathbf{R}\times\mathbf{T}^d)} := \|\langle k\rangle^s \langle \tau - h(k)\rangle^b \tilde u\|_{l^2_k L^1_\tau(\mathbf{R}\times\mathbf{Z}^d)}.$$

Establish the embeddings

$$\|u\|_{Y^{s,b-1/2-\varepsilon}_{\tau=h(k)}(\mathbf{R}\times\mathbf{T}^d)} \lesssim_\varepsilon \|u\|_{X^{s,b}_{\tau=h(k)}(\mathbf{R}\times\mathbf{T}^d)}$$

and

$$\|u\|_{C_t^0 H_x^s(\mathbf{R}\times\mathbf{T}^d)} \lesssim \|u\|_{Y^{s,0}(\mathbf{R}\times\mathbf{T}^d)}$$

for all Schwartz functions u and all $\varepsilon > 0$; show that the former embedding breaks down at $\varepsilon = 0$. With the notation of Proposition 2.12, establish the energy estimate

$$\begin{aligned}&\|\eta(t)u\|_{Y^{s,0}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{T}^d)} + \|\eta(t)u\|_{X^{s,1/2}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{T}^d)}\\ &\quad\lesssim_{\eta,b} (\|u(0)\|_{H^s_x(\mathbf{T}^d)} + \|F\|_{Y^{s,-1}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{T}^d)} + \|F\|_{X^{s,-1/2}_{\tau=h(\xi)}(\mathbf{R}\times\mathbf{T}^d)}).\end{aligned}$$

In the periodic theory, these estimates allow one to use the endpoint space $X^{s,1/2}$ (which is otherwise very badly behaved, as the preceding exercise showed) by augmenting it with the additional space $Y^{s,0}$.

Exercise 2.73 ($X^{s,b}$ spaces for the wave equation). Let us work in Minkowski space $\mathbf{R}^{1+d}$ with $c = 1$. Define the norm $\|u\|_{s,b} := \|u\|_{X^{s,b}_{|\tau|=|\xi|}(\mathbf{R}^{1+d})}$ by

$$\|u\|_{s,b} := \|\langle\xi\rangle^s \langle|\tau| - |\xi|\rangle^b \tilde{u}\|_{L^2(\mathbf{R}^{1+d})}.$$

Also define the slightly stronger norm $\mathcal{X}^{s,b}$ by

$$\|u\|_{\mathcal{X}^{s,b}} := \|u\|_{s,b} + \|\partial_t u\|_{s-1,b}.$$

Develop analogues of Corollary 2.10 and Lemma 2.11. Establish that

$$\|\Box u\|_{s-1,b-1} \lesssim_{s,b,d} \|u\|_{\mathcal{X}^{s,b}}$$

for all $u \in \mathcal{S}_{t,x}(\mathbf{R}^{1+d})$, and conversely that one has the energy estimate

$$\|\eta(t)u\|_{\mathcal{X}^{s,b}} \lesssim_{s,b,\eta} (\|u(0)\|_{H^s_x(\mathbf{R}^d)} + \|\partial_t u(0)\|_{H^{s-1}_x(\mathbf{R}^d)} + \|\Box u\|_{s-1,b-1})$$

for all compactly supported bump functions η, all $u \in \mathcal{S}_{t,x}(\mathbf{R}^{1+d})$, all $s \in \mathbf{R}$, and all $b > 1/2$. Typically, in applications one would place the solution u in the space $\mathcal{X}^{s,b}$ and the nonlinearity $\Box u$ in $X^{s-1,b-1}_{|\tau|=|\xi|}$. What is the counterpart of Lemma 2.9?

Exercise 2.74. Let $u \in \mathcal{S}_{t,x}(\mathbf{R}\times\mathbf{T}^2)$ solve the inhomogeneous Schrödinger equation $i\partial_t u + \Delta u = F$. Show that

$$\|\eta(t)u\|_{C^0_t L^4_x(\mathbf{R}\times\mathbf{T}^2)} + \|\eta(t)u\|_{L^4_{t,x}(\mathbf{R}\times\mathbf{T}^2)} \lesssim_\eta \|u(0)\|_{L^2_x(\mathbf{T}^2)} + \|F\|_{L^{4/3}_{t,x}(\mathbf{R}\times\mathbf{T}^2)}$$

for all compactly supported cutoff functions η. (Hint: use Proposition 2.13 and the Christ-Kiselev lemma, Lemma 2.4.)

Exercise 2.75 (Bilinear refinement to Strichartz). [**Bou9**], [**CKSTT11**] Let $u, v \in \mathcal{S}_{t,x}(\mathbf{R}\times\mathbf{R}^d)$ be fields whose spacetime Fourier transforms $\tilde{u}, \tilde{v}$ are supported on the sets $|\xi| \leq M$ and $|\xi| \geq N$ respectively for some $N, M \geq 1$. If $d = 1$, let us impose the additional hypothesis $N > 2M$. Show that

$$\|uv\|_{L^2_t L^2_x(\mathbf{R}\times\mathbf{R}^d)} \lesssim_b \frac{M^{(d-1)/2}}{N^{1/2}} \|u\|_{X^{0,b}_{\tau=|\xi|^2}(\mathbf{R}\times\mathbf{R}^d)} \|v\|_{X^{0,b}_{\tau=|\xi|^2}(\mathbf{R}\times\mathbf{R}^d)}.$$

(Hint: use Lemma 2.9 twice to reduce u and v to free solutions of the Schrödinger equation, and compute using Plancherel explicitly. In the case $d = 2$ and $N \leq 2M$, one can use Strichartz estimates.)

EXERCISE 2.76 (Divisor bound). Show that a positive integer d has at most $O_\varepsilon(n^\varepsilon)$ divisors for any $\varepsilon > 0$. (Hint: first show that if d is the power of a prime p, then d has at most $O_\varepsilon(n^\varepsilon)$ divisors, and in fact has at most n^ε divisors if p is sufficiently large depending on ε. For the general case, factorise d into the product of prime powers.)

EXERCISE 2.77 (Periodic Airy $L^6_{t,x}$ estimate). [**Bou**] Using Exercise 2.76 and the identity

$$(k_1+k_2+k_3)^3 - k_1^3 - k_2^3 - k_3^3 = 3(k_1+k_2)(k_2+k_3)(k_3+k_1),$$

show that for any integers k, t that the number of integer solutions to the system $k_1+k_2+k_3 = k$, $k_1^3+k_2^3+k_3^3 = t$ with $k_1, k_2, k_3 = O(N)$ is at most $O_\varepsilon(N^\varepsilon)$ for any $N \geq 1$ and $\varepsilon > 0$. Use this to obtain the estimate

$$\| \sum_{k \in \mathbf{Z}} a_k e^{ikx + ik^3 t} \|_{L^2_t L^2_x(\mathbf{T} \times \mathbf{T})} \lesssim_\varepsilon (\sum_{k \in \mathbf{Z}} \langle k \rangle^\varepsilon a_k^2)^{1/2}$$

for any complex numbers a_k and any $\varepsilon > 0$, and use this in turn to conclude the Strichartz estimate

$$\|\eta(t) u\|_{L^6_t L^6_x(\mathbf{R} \times \mathbf{T})} \lesssim_{\varepsilon, b} \|u\|_{X^{\varepsilon, b}_{\tau = k^3}(\mathbf{R} \times \mathbf{T})}$$

for any $\varepsilon > 0$ and $b > 1/2$ and any field u. It would be of interest to know if this estimate holds for $\varepsilon = 0$, or with the exponent $p = 6$ replaced by a larger exponent such as $p = 8$.

EXERCISE 2.78 (Periodic Airy L^6 estimate, II). [**Bou**] Show that for any integers k, t that the number of integer solutions to the system $k_1 + k_2 - k_3 = k$, $k_1^2 + k_2^2 - k_3^2 = t$ with $k_1, k_2, k_3 = O(N)$ is at most $O_\varepsilon(N^\varepsilon)$ for any $N \geq 1$ and $\varepsilon > 0$. (Hint: use the first equation to eliminate k_3 from the second, and then obtain an identity of the form $(k_1 + a)(k_2 + b) = c$ for some a, b, c given explicitly in terms of k, t.) By arguing as in the preceding exercise, show that

$$\|\eta(t) u\|_{L^6_t L^6_x(\mathbf{R} \times \mathbf{T})} \lesssim_{\varepsilon, b} \|u\|_{X^{\varepsilon, b}_{\tau = k^2}(\mathbf{R} \times \mathbf{T})}$$

for any $\varepsilon > 0$ and $b > 1/2$ and any field u. It is known that the ε cannot be set to zero in this case, though perhaps if the exponent $p = 6$ were lowered slightly then this could be possible. See also [**CKSTT3**], [**CKSTT12**] for some trilinear refinements of this estimate in which the ε loss can be eliminated.

CHAPTER 3

Semilinear dispersive equations

Come what come may,
Time and the hour runs through the roughest day.
(William Shakespeare, "Macbeth")

In this chapter we turn at last to the main subject of this monograph, namely nonlinear dispersive equations. Specifically, we now study the local and global low-regularity wellposedness of the following two Cauchy problems: the *nonlinear Schrödinger equation (NLS)*[1]

$$\begin{aligned} iu_t + \frac{1}{2}\Delta u &= \mu |u|^{p-1} u \\ u(t_0, x) &= u_0(x) \in H^s_x(\mathbf{R}^d) \end{aligned} \tag{3.1}$$

and the *nonlinear wave equation (NLW)*

$$\begin{aligned} \Box u &= \mu |u|^{p-1} u \\ u(t_0, x) &= u_0(x) \in H^s_x(\mathbf{R}^d) \\ \partial_t u(t_0, x) &= u_1(x) \in H^{s-1}(\mathbf{R}^d). \end{aligned} \tag{3.2}$$

In this chapter we have normalised $c = \hbar = m = 1$, so that $\Box = -\partial_t^2 + \Delta$. We will often also take advantage of time translation invariance to normalise $t_0 = 0$. In both cases, the scalar field $u : \mathbf{R}^{1+d} \to \mathbf{C}$ (or $u : I \times \mathbf{R}^d \to \mathbf{C}$, if one only seeks local solutions) is the desired solution, and the initial data u_0 (and u_1, in the case of NLW) is specified and lies in a given Sobolev space $H^s_x(\mathbf{R}^d)$ (or $H^{s-1}_x(\mathbf{R}^d)$). The exponent $1 < p < \infty$ denotes the power of the nonlinearity and is also given; the sign $\mu \in \{-1, 0, +1\}$ denotes whether the nonlinearity is defocusing, absent, or focusing respectively[2]; we will see some reasons for this terminology later in this chapter. The cases when p is an odd integer, and in particular the *cubic* case $p = 3$ and the *quintic* case $p = 5$, are particularly important in mathematical physics, and have the advantage that the nonlinearity $z \mapsto |z|^{p-1} z$ is smooth, indeed it is a polynomial in z and $\overline{z}$. We shall refer to these instances of NLS and NLW as the *algebraic* NLS and NLW respectively. The periodic analogues of these problems, when the domain is the torus $\mathbf{T}^d$ instead of Euclidean space $\mathbf{R}^d$ is also of interest, though our main focus here shall be on the non-periodic case.

[1]The factor of $\frac{1}{2}$ can be easily eliminated by rescaling time by a factor of 2, and so can be safely ignored. We retain it in order to make the dispersion relation (or de Broglie law) between velocity and frequency as simple as possible, namely $v = \xi$.

[2]The defocusing and focusing nonlinearities are sometimes called *repulsive* and *attractive* nonlinearities in the literature.

For the NLW it is convenient to adopt the notation $u[t] := (u(t), \partial_t u(t))$, thus for instance $u[t_0] = (u_0, u_1)$. Thus $u[t]$ describes the total state (both position and velocity) of the solution u at time t.

The power-type nonlinearity function $F(u) := \mu |u|^{p-1} u$ can be replaced by other nonlinearities, and in many cases one obtains results similar to those stated here. But the specific choice of power-type nonlinearity has a number of nice properties that make it well-suited for exposition, in particular enjoying symmetries such as the scaling and phase rotation symmetry $F(zu) = |z|^p F(u)$ for any complex z, which will in turn lead to corresponding symmetries for NLS and NLW. It is also naturally associated to a Hamiltonian potential $V(u) := \frac{1}{p+1}\mu|u|^{p+1}$ via the observation

$$\frac{d}{d\varepsilon} V(u + \varepsilon v)|_{\varepsilon=0} = \mathrm{Re}(F(u)\overline{v})$$

for any $u, v \in \mathbf{C}$; this will lead to a Hamiltonian formulation for NLS and NLW (Exercise 3.1).

We will be particularly interested in the *low regularity problem*: whether one still has existence and uniqueness of solutions even when the initial datum only lies in a very low Sobolev space. There are a number of reasons why one would want to go beyond high-regularity (classical) solutions and consider low-regularity ones[3]. Firstly, a good low-regularity theory gives more control on the nature of singularities of a solution, if they do indeed form; generally speaking, if one has a local wellposedness theory in H^s_x, then that implies that a singularity can only form by making the H^s_x norm go to infinity (or to concentrate at a point, if the norm H^s_x is critical with respect to scaling). Secondly, many of the key structural features of an equation - such as the conserved Hamiltonian, the symplectic form, the scale invariance, or other conserved or monotone quantities - are typically associated to rather low regularities such as L^2_x, $H^{1/2}_x$, or H^1_x, and in order to fully exploit these features it is often important to have a good local theory at those regularities. Thirdly, the technical challenge of working at low regularities (especially near or at the critical regularity) enforces a significant discipline on one's approach to the problem - requiring one to exploit the structural properties of the equation as efficiently and as geometrically as possible - and has in fact led to the development of powerful and robust new techniques and insights, which have provided new applications even for smooth solutions (for instance, in clarifying the dynamics of energy transfer between low and high frequencies). Finally, the task of extending a local existence result to a global existence result can (somewhat paradoxically) be easier if one is working at low regularities than high regularities, particularly if one is working at the scale-invariant regularity, or a regularity associated to a conserved quantity.

The nonlinear Schrödinger and wave models (3.1), (3.2) are among the simplest nonlinear perturbations of the free (linear) Schrödinger and wave equations[4]. Both equations are *semilinear* (the nonlinearity is lower order than the linear terms), and

---

[3]Alternatively, one can continue to work exclusively with classical solutions so that there is no difficulty justifying various formal computations, but demand that all estimates depend only on low-regularity norms of the solution. In practice, the two approaches are essentially equivalent; in most cases one can use limiting arguments to recover the former from the latter.

[4]Indeed, the NLS (together with the KdV equation) frequently arises in physics as the first nonlinear approximation of a dispersive system, by performing a Taylor expansion of the nonlinearity and discarding all but the first term. See for instance [**SSul**].

furthermore the nonlinear term contains no derivatives or non-local terms. Furthermore, these nonlinear equations retain many of the symmetries and structure of their linear counterparts; for instance, the NLS is a Hamiltonian equation with conservation of mass (charge), momentum, energy, enjoys scaling, Galilean, translation, and (partial) pseudoconformal symmetries, and enjoys several monotonicity formulae, including some of virial and Morawetz type. The NLW is also Hamiltonian with a conserved stress-energy tensor, with all its attendant consequences such as Morawetz inequalities and finite speed of propagation, and also enjoys scaling, Lorentz, translation and (partial) conformal symmetries. On the other hand, these equations are not completely integrable (with the notable exception of the one-dimensional cubic ($p = 3$) NLS, as well as a variant of the NLW known as the *sine-Gordon equation*), and so do not admit many explicit solutions (beyond some standard solutions such as the ground state solitons). The large number of parameters present in these equations (the dimension n, the power p, the sign μ, the regularity s, and whether one wishes to consider periodic or non-periodic solutions) means that these equations exhibit a wide range of phenomena and behaviour, and in many ways are quite representative of the much larger class of nonlinear dispersive and wave equations which are studied in the literature. Thus while our understanding of these equations is somewhat better than for most other nonlinear dispersive models (particularly for subcritical and critical regularities, for small data, and for the defocusing regularity), they are still so rich in structure and problems that there is still plenty to be understood.

Broadly speaking, there are two major classes of techniques one can use to analyze these equations. On the one hand, one has *perturbative methods*, which approximate the nonlinear equations (3.1), (3.2) by more tractable and well-understood equations such as[5] the free (and linear) Schrödinger or wave equations. The error between the actual equation and the approximate equation is usually treated by some sort of iteration argument (usually based on Duhamel's formula) or by a Gronwall inequality argument (usually based on energy estimates). Another related example of a perturbative method arises when constructing exact solutions to NLS and NLW by first starting with an approximate solution (that solves the equation up to a small error) and then constructing some sort of iterative scheme or Gronwall inequality argument to convert the approximate solution to an exact one.

As the name implies, perturbative methods only work when the solution is very close to its approximation; typically, this requires the initial datum to be small (or a small perturbation of a special initial datum), or the time interval to be small (or perhaps some spacetime integral of the solution to be well controlled on this time interval). When dealing with large solutions over long times, perturbative techniques no longer work by themselves, and one must combine them with *non-perturbative methods.* Examples of such methods include conservation laws, monotonicity formulae, and algebraic transformations of the equation. Such methods are initially only justified for smooth solutions, but can often be extended to rough solutions by means of the perturbative theory. Thus, global control of a solution is often obtained via a collaboration between the perturbative techniques

[5]In some cases one will use a more complicated equation as the approximating equation. For instance, if one is analyzing the NLS or NLW near a special solution such as a soliton solution, one often uses the linearised equation around that soliton as the approximating equation.

and non-perturbative techniques; typically, the perturbative theory guarantees a well-behaved solution provided that certain integrals of the solution stay bounded, and the non-perturbative theory guarantees control of these integrals provided that the solution remains well-behaved (see e.g. Figure 7.). This basic division of labour already works remarkably well in many situations, although in some recent results (most notably in those employing the *induction on energy* strategy, see Chapter 5) one has had to apply a more advanced scheme.

We conclude this introduction by describing some special (and very explicit) solutions to both (3.1) and (3.2), in order to build some initial intuition about these equations, though we emphasise that for generic initial data we do not expect any similarly explicit formula for the solution. Let us begin by using the classical method of separation of variables, using solutions of ODE to construct special solutions to PDE. For any $\xi \in \mathbf{R}^d$, the plane wave $e^{ix\cdot\xi}$ is an eigenfunction of the Laplacian ξ, and also has magnitude one, which leads one to consider the ansatz

$$u(t,x) = e^{ix\cdot\xi}v(t). \tag{3.3}$$

Simple calculation then shows that in order to solve the NLS (3.1), v must obey the ODE

$$\partial_t v = -i(\frac{|\xi|^2}{2} + \mu|v|^{p-1})v$$

and to solve the NLW (3.2), v must obey the ODE

$$\partial_t^2 v = -(|\xi|^2 + \mu|v|^{p-1})v. \tag{3.4}$$

In the case of NLS, the ODE for v can be explicitly solved, leading to the plane wave solutions

$$u(t,x) := \alpha e^{i\xi\cdot x}e^{i|\xi|^2 t/2}e^{i\mu|\alpha|^{p-1}t} \tag{3.5}$$

for any $\alpha \in \mathbf{C}$ and $\xi \in \mathbf{Z}^d$. Note how the time oscillation of $e^{i|\xi|^2 t/2}$ arising from the linear evolution is augmented by the additional time oscillation $e^{i\mu|\alpha|^{p-1}t}$. In the defocusing case $\mu = +1$, both time oscillations are anti-clockwise, so one can view the defocusing nonlinearity as amplifying the dispersive effect of the linear equation; in the focusing case the focusing nonlinearity is instead trying to cancel the dispersive effect. If the amplitude α is small compared the frequency ξ then the dispersive effect is stronger, but when the amplitude is large then the focusing effect takes over. This already illustrates one useful heuristic: the focusing and defocusing equations behave similarly when the initial data is small or when the frequency is very large[6].

As for the NLW (3.2), one can obtain a similar class of (complex) explicit solutions

$$u(t,x) := \alpha e^{i\xi\cdot x}e^{\pm i(|\xi|^2+\mu|\alpha|^{p-1})^{1/2}t}$$

provided that $|\xi|^2+\mu|\alpha|^{p-1} \geq 0$. This latter condition is automatic in the defocusing case $\mu = +1$ or the linear case $\mu = 0$, but requires either the frequency ξ to be large or the amplitude α to be small in the focusing case $\mu = -1$. This is again consistent with the heuristic mentioned earlier. When the initial data is large and

[6]Actually, this heuristic is only valid in "subcritical" situations, in which the high frequencies scale more favourably than the low frequencies. In critical cases, the high and low frequencies are equally sensitive to the distinction between focusing and defocusing; the supercritical cases are very poorly understood, but it is believed that the high frequency behaviour is radically different in the focusing and defocusing cases.

positive compared to the frequency, then the ODE (3.4) can blow up (this can be seen for instance using Exercise 1.25, in the case when v is real); one explicit family of blowup solutions in the focusing case $\mu = +1$ (with $\xi = 0$) is given by

$$u(t,x) := c_p(t_0 - t)^{-2/(p-1)} \tag{3.6}$$

for $t < t_0$, where $c_p := (\frac{2(p+1)}{(p-1)^2})^{1/(p-1)}$ and $t_0 \in \mathbf{R}$ is an arbitrary parameter. In contrast, in the defocusing or linear cases $\mu = -1, 0$ no blowup solution of the form (3.3) is possible, because (3.4) enjoys a coercive Hamiltonian[7]

$$H(v, \partial_t v) = \frac{1}{2}|\partial_t v|^2 + \frac{1}{2}|\xi|^2|v|^2 + \frac{\mu}{p+1}|v|^{p+1}$$

and thus (by Exercise 1.29) the solutions (3.3) will stay globally bounded for all time. Similarly in the focusing case if the initial data is very small compared to the frequency. Thus we see that the large data focusing behaviour is quite bad when compared to the defocusing or linear cases.

The solutions of the form (3.3) have no decay in space and so will not lie in Sobolev spaces such as $H^s_x(\mathbf{R}^d)$, although if the frequency ξ lies in the integer lattice $\mathbf{Z}^d$ then we can view these solutions as lying in the periodic Sobolev spaces $H^s_x(\mathbf{T}^d)$ for any s. In the (focusing) non-periodic case it is possible to create a different class of solutions by choosing an ansatz which oscillates in time rather than in space:

$$u(t,x) = Q(x)e^{it\tau}, \tag{3.7}$$

where $\omega \in \mathbf{R}$. This leads to the *ground state* equation

$$\Delta Q + \alpha|Q|^p Q = \beta Q \tag{3.8}$$

where $(\alpha, \beta) := (-2\mu, 2\tau)$ for NLS and $(\alpha, \beta) := (-\mu, \tau^2)$ for NLW. In the defocusing case we can take $\alpha, \beta > 0$ (choosing τ to be positive). From Appendix B we then recall that if $1 < p < \infty$ is energy-subcritical in the sense that $\frac{d}{2} - \frac{2}{p-1} < 1$, then there exists a smooth, positive, rapidly decreasing solution $Q \in \mathcal{S}_x(\mathbf{R}^d)$ to the equation (3.8). This leads to the *standard ground state soliton* solution to either NLS or NLW associated to the temporal frequency $\tau > 0$; it lies in every spatial Sobolev space $H^s_x(\mathbf{R}^d)$, and has a very simple behaviour in time. In the next section we will apply the symmetries of NLW and NLS to generate further ground state solitons. These solitons are only available in the focusing case; In Section 3.5 we shall establish *Morawetz inequalities* which show that nothing remotely resembling a soliton can occur in the defocusing equation.

EXERCISE 3.1. Obtain the analogue of Exercise 2.47 for the NLS, and Exercise 2.58 for the NLW, by adding the nonlinear potential energy term $V(u)$ to the Hamiltonian.

EXERCISE 3.2. Let $u \in C^2_{t,x,\mathrm{loc}}(\mathbf{R} \times \mathbf{R}^d \to V)$ be a classical solution to a d-dimensional NLS. Show that the field $v \in C^2_{t,x,\mathrm{loc}}(\mathbf{R} \times \mathbf{R}^{d+1} \to V)$ defined by

$$v(t, x_1, \ldots, x_d, x_{d+1}) := e^{-i(t+x_{d+1})}u(\frac{t - x_{d+1}}{2}, x_1, \ldots, x_d)$$

is a classical solution to the corresponding $d+1$-dimensional NLW (cf. Exercise 2.11). This correspondence may help explain why many of the algebraic expressions

[7]The case $\mu = \xi = 0$ is degenerate coercive, but this case can be treated by hand, leading to solutions of linear growth in time.

defined below for the NLW have a counterpart for NLS, but with d replaced by $d+1$. This correspondence is less useful for the H^s_x wellposedness theory, because the functions v constructed above will not have finite H^s_x norm. One should also caution that this correspondence does *not* link periodic NLS solutions with periodic NLW solutions.

EXERCISE 3.3. By taking formal limits of the Lax pair formulation of the periodic Ablowitz-Ladik system as discussed in Section 1.7, discover a Lax pair formalism for the one-dimensional cubic defocusing Schrödinger equation (in either the periodic or nonperiodic settings).

3.1. On scaling and other symmetries

It has long been an axiom of mine that the little things are infinitely the most important. (Sir Arthur Conan Doyle, "A Case of Identity")

We now describe the concrete symmetries of NLS and NLW; to avoid technicalities let us just work with classical solutions for now (we will discuss more general notions of solution in the next section). The NLS (3.1) enjoys the scaling symmetry

$$u(t,x) \mapsto \lambda^{-2/(p-1)} u(\frac{t}{\lambda^2}, \frac{x}{\lambda}); \quad u_0(x) \mapsto \lambda^{-2/(p-1)} u_0(\frac{x}{\lambda}) \tag{3.9}$$

for any dilation factor $\lambda > 0$ (thus time has twice the dimensionality of space), and the Galilean invariance

$$u(t,x) \mapsto e^{ix\cdot v} e^{it|v|^2/2} u(t, x - vt); \quad u_0(x) \mapsto e^{ix\cdot v} u_0(x) \tag{3.10}$$

for any velocity $v \in \mathbf{R}^d$ (cf. Exercise 2.5). It also enjoys the more mundane symmetries of time translation invariance, space translation invariance, spatial rotation symmetry, phase rotation symmetry $u \mapsto e^{i\theta} u$, as well as time reversal symmetry

$$u(t,x) \mapsto \overline{u(-t,x)}; \quad u_0(x) \mapsto \overline{u_0(x)}.$$

In the *pseudo-conformal case* $p = p_{L^2} := 1 + \frac{4}{d}$, one also use Exercise 2.28 to verify the *pseudo-conformal symmetry*

$$u(t,x) \mapsto \frac{1}{(it)^{d/2}} \overline{u(\frac{1}{t}, \frac{x}{t})} e^{i|x|^2/2t} \tag{3.11}$$

for times $t \neq 0$. This symmetry is awkward to use directly (at least when $t_0 = 0$) because of the singularity at $t = 0$; one typically uses the time translation and time reversal symmetries to move the singularity elsewhere (e.g. to the time $t = -1$).

Similarly, the NLW (3.2) enjoys the scaling symmetry

$$\begin{aligned} u(t,x) &\mapsto \lambda^{-2/(p-1)} u(\frac{t}{\lambda}, \frac{x}{\lambda}); \\ u_0(x) &\mapsto \lambda^{-2/(p-1)} u_0(\frac{x}{\lambda}); \\ u_1(x) &\mapsto \lambda^{-2/(p-1)-1} u_1(\frac{x}{\lambda}) \end{aligned} \tag{3.12}$$

for any dilation factor $\lambda > 0$ (thus time and space have equal dimension), and (if the solution exists globally in time) the Lorentz invariance

$$u(t,x) \mapsto u(\frac{t - v\cdot x}{\sqrt{1-|v|^2}}, x - x_v + \frac{x_v - vt}{\sqrt{1-|v|^2}}) \tag{3.13}$$

for all sub-luminal velocities $v \in \mathbf{R}^d$ with $|v| < 1$ (cf. Exercise 2.6); note that the effect of this invariance on the initial data u_0, u_1 is rather complicated and requires solving the equation (3.2). The NLW also enjoys spacetime translation invariance, spatial rotation symmetry, phase rotation symmetry, conjugation symmetry, and time reversal symmetry. In the *conformal case* $p = p_{\dot H^{1/2}_x} = 1 + \frac{4}{d-1}$, one also has the *conformal symmetry*

$$u(t,x) \mapsto (t^2 - |x|^2)^{-(d-1)/2} u(\frac{t}{t^2-|x|^2}, \frac{x}{t^2-|x|^2})$$

inside the light cone $|t| < |x|$, thanks to Exercise 2.14.

Unlike the Galilean invariance (3.10), the Lorentz invariance (3.13) has the effect of *time dilation* - solutions which would ordinarily exhibit some special behaviour (e.g. blowup) at a time T will instead do so at a much later time, typically of the order of $T/\sqrt{1-|v|^2}$. To compensate for this one can compose the Lorentz transformation with the scaling transformation with $\lambda := 1/\sqrt{1-|v|^2}$, leading to the *normalised Lorentz invariance*

$$u(t,x) \mapsto (1-|v|^2)^{1/(p-1)} u(t - v\cdot x, \sqrt{1-|v|^2}(x-x_v) + x_v - vt). \tag{3.14}$$

Symmetries have many uses. Through Noether's theorem, they indicate what conservation laws are available (though certain symmetries, particularly discrete ones, do not necessarily yield a conservation law). They can give guidance as to what type of techniques to use to deal with a problem; for instance, if one is trying to establish wellposedness in a data class which is invariant under a certain symmetry, this suggests the use of estimates and other techniques which are also invariant under that symmetry; alternatively, one can "spend" the symmetry by normalising the solution, for instance in making the solution centred or concentrated at the origin (or some other specified location) in space, time, or frequency. If the data class is subcritical with respect to scaling, one can use the scaling symmetry to trade between time of existence and size of initial data; thus if one establishes a local wellposedness at a fixed time (say up to time $T = 1$) for data with small norm, then one can often also establish local wellposedness at a small time for large data; typically the time of existence will be proportional to some negative power of the norm of the data. Conversely, if the data class is supercritical with respect to scaling (or more generally is lower than the invariant norm associated to another symmetry), then it is likely that there is a significant obstruction to obtaining a wellposedness theory below that regularity, and one also expects the wellposedness theory at that regularity to be rather delicate. The reason for this is that if the regularity is below the invariant regularity, then one can use the symmetry to convert bad behaviour arising from large initial data at some time $t > 0$ to bad behaviour arising from small initial data at some time less than or equal to t, where "large" and "small" are measured with respect to the regularity $H^s_x(\mathbf{R}^d)$. Since large initial data would be expected to display bad behaviour very quickly, one then expects to give examples of arbitrary small initial data which displays bad behaviour arbitrarily quickly. This can often be used to contradict certain types of wellposedness that one could hypothesise for this regularity. See also Principle 3.1 below.

Let us give some sample applications of the symmetry laws. The first is a blowup result for the pseudoconformal focusing NLS (so $\mu = -1$ and $p = p_{L^2_x} = 1+\frac{4}{d}$). Recall that this equation has a soliton solution of the form $u(t,x) = e^{it\tau}Q(x)$

for any $\tau > 0$, where the ground state Q is a nonnegative Schwartz solution to the equation

$$\Delta Q + 2Q^{1+4/d} = 2\tau Q.$$

Applying the pseudoconformal transformation (2.33), we obtain the solution

$$\frac{1}{(it)^{d/2}} e^{-it/\tau} e^{i|x|^2/2t} Q(x/t). \tag{3.15}$$

This solves the NLS equation for all times $t \neq 0$, and in fact lies in every Sobolev space $H^s_x(\mathbf{R}^d)$ for such times, but blows up in a rather dramatic way as $t \to 0$. Thus the pseudoconformal focusing NLS can lead to blowup even from very smooth decaying initial data, though we will later see that this is due to the initial datum being "large" in an $L^2_x(\mathbf{R}^d)$ sense. This blowup occurs despite the solution being bounded in L^2_x, and despite the conservation of the L^2_x norm. Thus for PDE, a positive definite conservation law is not always sufficient to prevent blowup from occuring, in marked contrast to the situation for ODE; note that the solution (3.15) demonstrates rather vividly the lack of compactness of bounded subsets of $L^2_x(\mathbf{R}^d)$.

Now consider a general NLS. Applying the pseudoconformal transformation in Exercise 2.28, one no longer expects to recover the original equation; instead, the transformed field $v(t, x)$ will now obey the equation

$$i\partial_t v + \Delta v = t^{\frac{d}{2}(p - p_{L^2_x})} \mu |v|^{p-1} v \tag{3.16}$$

for $t \neq 0$, where $p_{L^2_x} := 1 + \frac{4}{d}$ is the pseudoconformal power. We will analyze this equation (3.16) in more detail later, but for now let us just extract a special class of solutions to (3.16) (and hence to NLS), by considering solutions v which are independent of the spatial variable and thus simply solve the ODE

$$i\partial_t v = t^{\frac{d}{2}(p - p_{L^2_x})} \mu |v|^{p-1} v. \tag{3.17}$$

This ODE can be solved explicitly as

$$v(t, x) = \alpha \exp\left(-\frac{i\mu|\alpha|^{p-1}}{q} t^q\right)$$

for any $\alpha \in \mathbf{C}$ and with $q := \frac{d}{2}(p - p_{L^2_x}) + 1$, though in the *critical-range* case $q = 0$ (so $p = 1 + \frac{2}{d}$) we have instead the solution

$$v(t, x) = \alpha \exp(-i|\alpha|^{p-1} \log |t|).$$

We can of course invert the pseudoconformal transformation and obtain explicit solutions to the original NLS for $t \neq 0$, namely

$$u(t, x) = \frac{1}{(it)^{d/2}} \overline{\alpha} \exp\left(\frac{i|x|^2}{2t} + \frac{i\mu|\alpha|^{p-1}}{qt^q}\right) \tag{3.18}$$

when $q \neq 0$ and

$$u(t, x) = \frac{1}{(it)^{d/2}} \overline{\alpha} \exp\left(\frac{i|x|^2}{2t} + i\mu|\alpha|^{p-1} \log |t|\right) \tag{3.19}$$

when $q = 0$. Of course when $\mu = 0$ we recover the explicit solution $u(t, x) = \frac{1}{(it)^{d/2}} \overline{\alpha} \exp(\frac{i|x|^2}{2t})$ to the linear Schrödinger equation (essentially the fundamental solution[8] for that equation). Comparison of these solutions yields the following heuristic: as $t \to \pm\infty$, the nonlinear Schrödinger equation should resemble the linear Schrödinger equation when in the *short-range* case $q > 0$ (so $p > 1 + \frac{2}{d}$), but not in the *long-range*[9] case $q < 0$ or critical-range case $q = 0$ (though the divergence between the two equations should consist primarily of a phase shift, which should be somehow "logarithmic" in the critical-range case). We will see some justification of this heuristic later in this chapter, though our understanding here is far from complete.

Now we observe some applications of the Galilean invariance law (3.10). Let us begin with a periodic NLS (with d, p and $\mu = \pm 1$ arbitrary). In this periodic setting we have the plane wave solutions

$$u_{\alpha,\xi}(t,x) := \alpha e^{i\xi\cdot x} e^{i|\xi|^2 t/2} e^{i\mu|\alpha|^{p-1}t} \tag{3.20}$$

for any $\xi \in 2\pi\mathbf{Z}^d$ and $\alpha \in \mathbf{C}$; one can view this as the Galilean transform of the constant-in-space solutions $\alpha e^{i\mu|\alpha|^{p-1}t}$. Suppose one fixes the frequency parameter ξ to be large, and considers two distinct solutions $u_{\alpha,\xi}, u_{\alpha',\xi}$ of the above type with $|\alpha| \sim |\alpha'|$. At time zero we have

$$\|u_{\alpha,\xi}(0)\|_{H^s_x(\mathbf{T}^d)}, \|u_{\alpha',\xi}(0)\|_{H^s_x(\mathbf{T}^d)} \sim |\alpha||\xi|^s;$$
$$\|u_{\alpha,\xi}(0) - u_{\alpha',\xi}(0)\|_{H^s_x(\mathbf{T}^d)} \sim |\alpha - \alpha'||\xi|^s$$

while at any later time t we have

$$\|u_{\alpha,\xi}(t)\|_{H^s_x(\mathbf{T}^d)}, \|u_{\alpha',\xi}(t)\|_{H^s_x(\mathbf{T}^d)} \sim |\alpha||\xi|^s;$$
$$\|u_{\alpha,\xi}(t) - u_{\alpha',\xi}(t)\|_{H^s_x} \sim |\alpha e^{i\mu|\alpha|^{p-1}t} - \alpha' e^{i\mu|\alpha'|^{p-1}t}||\xi|^s.$$

Thus the H^s_x norms of the solutions $u_{\alpha,\xi}$ and $u_{\alpha',\xi}$ do not change much in time, but the H^s_x *separation* of these solutions can change due to a phase decoherence effect. Indeed we see that if $|\alpha| \neq |\alpha'|$, then there exists a time $t \sim |\alpha|^{1-p}$ for which the two phases become completely decohered, and $\|u_{\alpha,\xi}(t) - u_{\alpha',\xi}(t)\|_{H^s_x(\mathbf{T}^d)} \sim |\alpha||\xi|^s$. If s is negative, then by taking α to be large and $|\xi|$ to be comparable to $(|\alpha|/\varepsilon)^{-1/s}$, we can construct for any $\delta, \varepsilon > 0$, a pair of solutions $u_{\alpha,\xi}, u_{\alpha',\xi}$ to NLS of $H^s_x(\mathbf{T}^d)$ norm $O(\varepsilon)$ and $H^s_x(\mathbf{T}^d)$ norm separation $O(\delta)$ at time zero, such that at some later time $t = O(\varepsilon)$ the $H^s_x(\mathbf{T}^d)$ norm separation has grown to be as large as $O(\varepsilon)$. This shows that for negative s, a pair of solutions can separate in $H^s_x(\mathbf{T}^d)$ norm arbitrarily quickly; more precisely, the solution map $u_0 \mapsto u$ is not uniformly continuous from H^s_x to $C^0_t H^s_x([0,T] \times \mathbf{T}^d)$ even for arbitrarily small T and for arbitrarily small balls in $H^s_x(\mathbf{T}^d)$. This is a negative result that rules out certain types of strong wellposedness results for the periodic NLS for negative Sobolev regularities.

[8]Indeed, one could view (3.18), (3.19) as the "nonlinear fundamental solution" for NLS. However these solutions are nowhere near as useful as the fundamental solution is for the linear equation, since we no longer have the principle of superposition in the nonlinear case and so we cannot build general solutions by superimposing translates of the fundamental solution. Nevertheless, these explicit solutions provide some useful intuition as to the asymptotic behaviour of the equation for general data.

[9]The terminology here signifies the long-term strength of the nonlinearity and can be justified heuristically as follows. One can view the nonlinearity in NLS as a potential term with time-dependent potential $\mu|u|^{p-1}$. Assuming that the nonlinear evolution decays at the same rate as the linear one, dispersive estimates suggest that $|u|$ should decay like $t^{-d/2}$. Thus we expect in the short-range case we expect the potential to be integrable in time, which suggests by Gronwall's inequality that the long-term effect of the nonlinearity is bounded.

One can run a similar argument for nonperiodic focusing NLS, by starting with the ground state solution $e^{it\tau}Q(x)$, rescaling it by λ and then applying a Galilean transform to obtain the moving soliton solution

$$u_{v,\lambda}(t,x) \mapsto \lambda^{-2/(p-1)} e^{i(x\cdot v+it|v|^2/2+it\tau/\lambda^2} Q((x-vt)/\lambda) \tag{3.21}$$

for any $v \in \mathbf{R}^d$ and $\lambda > 0$; one can use these solutions to show that the solution map to NLS (if it exists at all) is not uniformly continuous in H^s_x for certain low s; see Exercise 3.5. A similar result is also known for the defocusing case, replacing the soliton solutions with another family of solutions that can be viewed as truncated versions of the plane wave solutions (3.5); see Section 3.8.

Among all the symmetries, the scale invariance (3.9), (3.12) is particularly important, as it predicts a relationship between time of existence and regularity of initial data. Associated to this invariance is the *critical regularity* $s_c := \frac{d}{2} - \frac{2}{p-1}$. Note that the scaling (3.9) preserves the homogeneous Sobolev norm $\|u_0\|_{\dot H^{s_c}(\mathbf{R}^d)}$, and similarly (3.12) preserves $\|u_0\|_{\dot H^{s_c}(\mathbf{R}^d)} + \|u_1\|_{\dot H^{s_c-1}(\mathbf{R}^d)}$. The relationship between scaling and the inhomogeneous counterparts to these Sobolev norms is a little more complicated, of course. We refer to regularities $s > s_c$ above the critical norm as *subcritical*, and regularities $s < s_c$ below the critical norm as *supercritical*. The reason for this inversion of notation is that higher regularity data has better behaviour, and thus we expect subcritical solutions to have less pathological behaviour than critical solutions, which in turn should be better behaved than supercritical solutions. The other scalings also have their own associated regularities; the Galilean symmetry and pseudoconformal symmetry preserve the $L^2_x(\mathbf{R}^d)$ norm, whereas the Lorentz symmetry and conformal symmetries are heuristically associated to the $\dot H^{1/2}(\mathbf{R}^d) \times \dot H^{-1/2}(\mathbf{R}^d)$ norm (see Exercise 2.23).

In general, the relationship between the regularity H^s_x of the initial data, the scale-invariant regularity H^{s_c} of the equation, the frequencies of the solution, and the evolution of the solution tends to follow the following informal principles[10]:

PRINCIPLE 3.1 (Scaling heuristic). *Let u be a solution to either the NLS* (3.1) *or NLW* (3.2)*, with initial position u_0 in H^s_x (and initial velocity u_1 in H^{s-1}, in the case of the NLW).*

(a) *In the* subcritical *case $s > s_c$, we expect the high frequencies of the solution to evolve linearly for all time (unless a stronger obstruction than scaling exists). The low frequencies of the solution will evolve linearly for short times, but nonlinearly for long times.*
(b) *In the* critical *case $s = s_c$, we expect the high frequencies to evolve linearly for all time if their H^{s_c} norm is small, but to quickly develop nonlinear behaviour when the norm is large. (Again, we are assuming that no stronger obstruction to linear behaviour than scaling exists.) The low frequencies of the solution will evolve linearly for all time if their H^{s_c} norm is small, but*

[10]This principle should be taken with a grain of salt. On the one hand, it gives a good prediction for those equations in which the scaling symmetry is in some sense "dominant", and for which the worst types of initial data are given by bump functions. On the other hand, there are other situations in which other features of the equation (such as Galilean or Lorentz symmetries, or resonances) dominate, in which case instability can occur even when the scaling heuristic predicts good behaviour. Conversely, some features of the equation, such as conservation laws or monotonicity formulae, can provide more stability than the scaling heuristic predicts.

will eventually develop nonlinear behaviour (after a long period of time) when the norm is large.

(c) *In the* supercritical *case $s < s_c$, the high frequencies are very unstable and will develop nonlinear behaviour very quickly. The low frequencies are in principle more stable and linear, though in practice they can be quickly disrupted by the unstable behaviour in the high frequencies. (This relatively good behaviour of the low frequencies is sometimes enough to obtain a* weak solution *to the equation, however, by using viscosity methods to suppress the high frequencies; see Exercise 3.56.)*

Let us now briefly give a heuristic discussion that lends some support to Principle 3.1. Let $N > 0$ be a frequency; frequencies $N \gg 1$ correspond to high frequencies, while frequencies $N \ll 1$ correspond to low frequencies. A model example of an initial datum u_0 of frequency $\sim N$ is a function which is supported on a ball B of radius $1/N$, does not oscillate too much on this ball, and reaches an amplitude A on this ball. (For the NLW, one would also need to similarly specify an initial velocity.) We will assume that this "rescaled bump function" example is the "worst" type of initial data in the given class (i.e. bounded or small functions in H^s_x or H^{s-1}); this assumption corresponds to the caveat given in the above principle that no stronger obstructions to linear behaviour exist than the scaling one. The L^2_x norm of such a datum is roughly $\sim AN^{-d/2}$, and more generally (from the Fourier representation of the H^s_x norm) we expect the H^s_x norm of this datum to be $\sim AN^{s-d/2}$; thus if u_0 is bounded in H^s_x then $A = O(N^{d/2-s})$, and if u_0 is small in H^s_x then $A \ll N^{d/2-s}$. Now, both the NLS (3.1) and the NLW (3.2) contain a linear term Δu and a nonlinear term $\mu |u|^{p-1}u$. On the ball B (at least for times close to 0), the linear term has magnitude $\sim N^2 A$, while the nonlinear term has amplitude $\sim A^p$. If $N^2 A \gg A^p$, we thus expect the linear term to dominate, and the solution should behave linearly (cf. Principle 1.37). If $A^p \gg N^2 A$, we expect the nonlinear term to dominate and so one eventually expects nonlinear (and unstable) behaviour. The time in which this nonlinear behaviour becomes apparent can be predicted by comparing u against its time derivative $\partial_t u$ or its second time derivative $\partial_t^2 u$. For instance, suppose we have an NLS in which the nonlinear behaviour dominates, thus $\partial_t u$ will be dominated by the nonlinear term $\mu |u|^{p-1}u$, which has amplitude $\sim A^p$. Since u itself has amplitude $\sim A$, we expect the nonlinear behaviour to significantly affect the initial datum after time $\sim A/A^p$. Using these heuristics, one can give informal justification for all three components (a), (b), (c) if Principle 3.1; see Exercise 3.4.

A particular interesting case is when the scale-invariant regularity coincides with one of the other special regularities, such as the $\dot H^1_x$ norm (associated to the energy or Hamiltonian), the $\dot H^{1/2}_x$ norm (associated to the momentum in NLS and to the symplectic structure, Lorentz invariance, and conformal invariance in NLW), and the L^2_x norm (associated to the Galilean invariance, pseudoconformal invariance, and mass in NLS, and being the limiting regularity in NLW to even make sense of (3.2) distributionally); see Table 1. Thus we isolate as special cases the $\dot H^1_x$-*critical* (or *energy-critical*) case $s_c = 1$ (thus $d \geq 3$ and $p = 1 + \frac{4}{d-2}$), the $\dot H^{1/2}_x$-*critical* case $s_c = 1/2$ (thus $d \geq 2$ and $p = 1 + \frac{4}{d-1}$) and the L^2_x-*critical* case $s_c = 0$ (thus $d \geq 1$ and $p = 1 + \frac{4}{d}$). One can also discuss the $\dot H^1_x$-subcritical case $s_c < 1$ and the $\dot H^1_x$-supercritical case $s_c > 1$, etc. Another relevant regularity

Table 1. The critical exponents for small dimension. The cases when a critical exponent corresponds to an algebraic equation (i.e. p is equal to 3 or 5) are of particular interest.

Dimension	L^2_x-critical	$\dot H^{1/2}_x$-critical	$\dot H^1_x$-critical
1	5	∞	–
2	3	5	∞
3	7/3	3	5
4	2	7/3	3
5	9/5	2	7/3
6	5/3	9/5	2

in the case of NLW is the *Lorentz regularity* $s_l := \frac{d+1}{4} - \frac{1}{p-1} = \frac{s_c}{2} + \frac{1}{4}$, which is the regularity which is heuristically associated to the normalised Lorentz invariance (3.14), and is halfway between the scale-invariant regularity s_c and the conformal regularity $\frac{1}{2}$.

Exercise 3.4. Use the heuristic analysis of bump function initial data, as described in this section, to give some informal justification to Principle 3.1. (Be prepared to make a large number of hand-waving assumptions. The important thing here is to develop the numerology of exponents; rigorous support for these heuristics will be have to wait until later in this chapter.) In the subcritical case, develop a heuristic relationship between the H^s_x norm of the initial data and the predicted time T in which the linear behaviour dominates. (One should get $T \sim \|u_0\|_{H^s}^{2/(s-s_c)}$ for NLS and $T \sim (\|u_0\|_{H^s} + \|u_1\|_{H^{s-1}})^{1/(s-s_c)}$ for the NLW.)

Exercise 3.5. [**BKPSV**] Let d, p be arbitrary, let $\mu = +1$, and let $s < 0$ or $s < s_c := \frac{d}{2} - \frac{2}{p-1}$. Using the solutions (3.21), show that for any $\varepsilon, \delta > 0$ there exists a pair of classical solutions u, u' to (3.1) with $H^s_x(\mathbf{R}^d)$ norm $O(\varepsilon)$ and $H^s_x(\mathbf{R}^d)$ norm separation $O(\delta)$ at time zero, such that at some later time $t = O(\varepsilon)$ the $H^s_x(\mathbf{R}^d)$ norm separation has grown to be as large as $O(\varepsilon)$. This shows that there is no uniform wellposedness at this regularity, at least for the focusing regularity.

3.2. What is a solution?

For every complex problem, there is a solution that is simple, neat, and wrong. (H.L. Mencken, "The Divine Afflatus")

Before we begin the analysis of our model problems (3.1), (3.2), let us pause to address a rather fundamental question, namely what it actually means for a field u to be a solution of either of these two Cauchy problems. This question may sound philosophical in nature, but the properties associated to making a solution concept "strong" are well worth establishing rigorously, as they become important in establishing many of the further properties of the equation, such as the global existence and asymptotics of (classical) solutions.

The question of defining exactly what a solution is is more subtle than it may first appear, especially at low regularities. The reason for this is that in order for a solution to a PDE to actually be useful for applications, it is not merely enough that it exist and solve the equation in some weak sense[11] (e.g. in the sense of distributions), though this is certainly a minimal requirement; one also often desires additional properties on the solution, which do not automatically follow from the fact that the equation is solved weakly. We informally describe some of the most important of these properties[12] as follows.

- Existence: Is the solution guaranteed to exist (locally, at least) for all initial data in a certain class (e.g. H^s_x)?
- Uniqueness: Is the solution the unique object in a certain solution class (e.g. $C^0_t H^s_x(I \times \mathbf{R}^d)$) which solves the equation in a suitable sense (e.g. in a distributional sense)? Is this solution concept compatible with other notions of solution (i.e. if two solutions to the same equation exist in two different senses, are they equal to each other)?
- Continuous dependence on the data: Do small perturbations of the initial datum (in some norm) lead to small perturbations in the solution (in some other norm)? In other words, is the solution map continuous? One can also ask for stronger continuity properties such as uniform continuity, Lipschitz continuity, or analyticity.
- Bounds: If the initial datum is in some class, say H^s_x, can one control the solution in some other class, e.g. $C^0_t H^s_x(I \times \mathbf{R}^d)$? In particular, does one have *persistence of regularity*: is the solution always as smooth as the initial datum (as measured in an H^s_x sense)?
- Lifespan estimates: Is there a lower bound on the lifespan of the solution in terms of the initial data (or in terms of some norm of the initial data, such as an $H^s_x(\mathbf{R}^d)$ norm)? Equivalently, is there a *blowup criterion* that gives necessary conditions for the lifespan to shrink to zero? In some cases one has *global existence*, which case the lifespan is infinite.
- Approximability by smooth solutions: if the solution is rough, can it be written as the limit (in some topology) of smoother solutions? If the initial datum is approximated by a sequence of smooth initial data, do the corresponding solutions necessarily converge to the original solution, and in what sense?
- Stability: If one perturbs the *equation* (thus considering fields which only solve the original equation *approximately*), to what extent can these near-solutions be approximated by the *exact* solution with the same (or a nearby) initial datum?

[11]This is in marked contrast with the theory of *linear* differential equations, in which distributional solutions are very tractable, and can mostly be manipulated as if they were classical solutions, in large part because they can be expressed as the weak limit of classical solutions. Since weak convergence is often not preserved under basic nonlinear operations such as multiplication of two functions, one generally requires in nonlinear applications that a solution be a *strong* limit of classical solutions, which usually leads to the requirement that one work with *wellposed solutions*; see below.

[12]For elliptic PDE, another important property that one often desires is that the solution u is a minimiser, or at least a critical point, of the Lagrangian associated to the PDE, with respect to various classes of perturbation. One could insist on something similar for nonlinear wave and Schrödinger equations, but this has not proven to be as fruitful a property for these equations as in the elliptic case, in large part because of the highly non-convex nature of the Lagrangians involved. However, the Lagrangian formulation is (formally) linked to important properties such as conservation laws and monotonicity formulae, which *are* very desirable properties for a solution to obey.

- Structures: Do the conservation laws of the equation, which can be rigorously justified for classical (i.e. smooth and decaying) solutions, continue to hold for the solution class being studied? Similarly for monotonicity formulae, symmetries of the equation, etc.

Thus, instead of having a single unified concept of a solution class, one has instead a multi-dimensional continuum of such classes, ranging from very weak solution classes (in which the equation solves the equation in a weak sense, or is perhaps a weak limit of smoother solutions or near-solutions, but not much else), to very strong solution classes, in which one has many or all of the desirable properties listed above. Generally speaking, it is fairly easy to show existence of solution in a weak solution class by various limiting arguments (e.g. iteration or weak compactness arguments), but non-trivial effort is then required to upgrade those solutions to lie in stronger solution classes.

In this section we shall discuss five notions of solution, which in decreasing order of strength are *classical solution*, *wellposed H^s_x solution*, *strong H^s_x solution*, *weak H^s_x solution*, and *distributional solution* respectively. In fact, in this monograph we shall largely work with wellposed and classical solutions, in order to avoid a number of subtleties involving the weaker notions of solution.

To fix the discussion let us just work with the NLS equation (3.1), and fix our initial data class to be a Sobolev space $H^s_x(\mathbf{R}^d)$. The strongest notion of a solution is that of a *classical solution*. These can (broadly speaking) be defined as solutions which have so much differentiability and decay that there is no difficulty interpreting the problem (3.1) in a classical sense (i.e. without requiring the theory of weak derivatives). Furthermore, one has enough regularity and decay available[13] to justify all the various formal manipulations associated to the equation, such as conservation laws, monotonicity formulae, and so forth. The amount of regularity required to do all this can be quite large; for instance, in order to justify conservation of the Hamiltonian for NLS safely, one requires as much as three orders of differentiability on the solution, as well as some additional uniformity and decay conditions. Because of this, one occasionally runs into issues generating a classical solution theory when the nonlinearity $\mu|u|^{p-1}u$ is not very smooth (which can happen when p is close to 1); in such cases one may need to regularise the nonlinearity before discussing classical solutions. However this issue does not arise for the algebraic equations, in which p is an odd integer.

It is also easy to establish uniqueness for classical solutions (essentially because the proof of Theorem 1.14 carries over without difficulty). Here are two typical such results, one for NLS and one for NLW.

PROPOSITION 3.2 (Uniqueness for classical NLS solutions). *Let I be a time interval containing t_0, and let $u, u' \in C^2_{t,x}(I \times \mathbf{R}^d \to \mathbf{C})$ be two classical solutions to* (3.1) *with the same initial datum u_0 for some fixed μ and p. Assume also that we have the mild decay hypothesis $u, u' \in L^\infty_t L^q_x(I \times \mathbf{R}^d)$ for $q = 2, \infty$. Then $u = u'$.*

PROOF. By time translation symmetry we can take $t_0 = 0$. By time reversal symmetry we may assume that I lies in the upper time axis $[0, +\infty)$. Let us write $u' = u + v$. Then $v \in C^2_{t,x}(I \times \mathbf{R}^d \to \mathbf{C})$, $v(0) = 0$, and v obeys the difference

[13] This is a somewhat vague definition, but in practice we will always apply limiting arguments to generalise classical solutions to a wider class of *wellposed solutions*, and so the exact notion of a classical solution will not be important as long as it is dense in the class of wellposed solutions.

equation

$$i\partial_t v + \Delta v = \mu(|u+v|^{p-1}(u+v) - |u|^{p-1}u).$$

Since v and $|u+v|^{p-1}(u+v)-|u|^{p-1}u$ lie in $L_t^\infty L_x^2(I\times\mathbf{R}^d)$, we may invoke Duhamel's formula (2.13) and conclude

$$v(t) = -i\mu\int_0^t e^{i(t-s)\Delta/2}(|u+v|^{p-1}(u+v) - |u|^{p-1}u)(s)\ ds$$

for all $t \in I$. By Minkowski's inequality, and the unitarity of $e^{i(t-s)\Delta}$, we conclude

$$\|v(t)\|_{L_x^2(\mathbf{R}^d)} \leq \int_0^t \|(|u+v|^{p-1}(u+v) - |u|^{p-1}u)(s)\|_{L_x^2(\mathbf{R}^d)}\ ds.$$

Since u and v are in $L_t^\infty L_x^\infty(I \times \mathbf{R}^d)$, and the function $z \mapsto |z|^{p-1}z$ is locally Lipschitz, we have the bound

$$\begin{aligned}\|(|u+v|^{p-1}(u+v) - |u|^{p-1}u)(s)\|_{L_x^2(\mathbf{R}^d)} \lesssim_p (&\|u\|^p_{L_t^\infty L_x^\infty(I\times\mathbf{R}^d)} + \|v\|^p_{L_t^\infty L_x^\infty(I\times\mathbf{R}^d)}) \\ &\times \|v(s)\|_{L_x^2(\mathbf{R}^d)}.\end{aligned}$$

Applying Gronwall's inequality (Theorem 1.10) we conclude that $\|v(t)\|_{L_x^2(\mathbf{R}^d)} = 0$ for all $t \in I$, and hence $u = u'$ as desired. $\square$

Note that Exercise 2.24 shows that some sort of decay condition is necessary in order to establish uniqueness, even when no nonlinearity is present. For NLW one also has uniqueness of classical solutions, and moreover one can even localise the uniqueness by exploiting finite speed of propagation:

Proposition 3.3 (Uniqueness and finite speed of propagation for classical NLW solutions). *Let $t_0 = 0$. Let I be a time interval containing 0, and let $u, u' \in C^2_{t,x,\mathrm{loc}}(I \times \mathbf{R}^d \to \mathbf{C})$ be two C^2 solutions to (3.2) such that the initial data $u[0] = (u(0), \partial_t u(0))$ and $u'[0] = (u'(0), \partial_t u'(0))$ agree on the ball $\{x \in \mathbf{R}^d : |x - x_0| \leq R\}$. Then we have $u(t,x) = u'(t,x)$ for all $t \in I$ and $x \in \mathbf{R}^d$ with $|x - x_0| \leq R - |t|$.*

Proof. By spatial translation invariance we may take $x_0 = 0$; by time reversal symmetry we may restrict attention to times $0 \leq t \leq R$. By shrinking I if necessary we may take I to be compact. Write $u' = u + v$, then $v \in C^2_{t,x}(I \times \mathbf{R}^d \to \mathbf{C})$, $v[0]$ vanishes on the ball $\{x \in \mathbf{R}^d : |x| \leq R\}$, and v solves the difference equation

$$\Box v = F$$

where $F := \mu(|u+v|^{p-1}(u+v) - |u|^{p-1}u)$. Now let us define the local energy $E[t]$ for $0 \leq t \leq R$ by

$$E(t) := \int_{|x|\leq R-t} \mathrm{T}_{00}(t,x)\ dx$$

where $\mathrm{T}_{00} := \frac{1}{2}|\partial_t v|^2 + \frac{1}{2}|\nabla_x v|^2$ is the linear energy density, thus $E(0) = 0$. A computation (which is justified when v is $C^2_{t,x}$) shows that

$$\partial_t \mathrm{T}_{00} + \partial_j \mathrm{T}_{0j} = -\mathrm{Re}(\overline{F}\partial_t v)$$

where $\mathrm{T}_{0j} := -\mathrm{Re}(\overline{\partial_j v}\partial_t v)$ is the energy current. From Stokes' theorem (and the fact that v is $C^2_{t,x}$) we conclude

$$\partial_t E(t) = -\int_{|x|\leq R-t} \mathrm{Re}(\overline{F}\partial_t v)(t,x)\ dx + \int_{|x|=R-t} \mathrm{T}_{0j} n_j - \mathrm{T}_{00}\ dS$$

where dS is the surface element and $n_j := x_j/|x|$ is the outward normal. From Cauchy-Schwarz we see that $|\mathrm{T}_{0j}n_j| \leq \mathrm{T}_{00}$, thus we have

$$\partial_t E(t) \leq \int_{|x|\leq R-t} |F(t,x)||\partial_t v|(t,x)\ dx.$$

Now since u and v will be bounded on the compact region $\{(t,x) \in I \times \mathbf{R}^d : 0 \leq t \leq R; |x| \leq R-t\}$, we see that $F = O_{u,v}(|v(t,x)|)$. Applying Cauchy-Schwarz we have

$$\partial_t E(t) \lesssim_{u,v} (\int_{|x|\leq R-t} |v(t,x)|^2\ dx)^{1/2} (\int_{|x|\leq R-t} |\partial_t v(t,x)|^2\ dx)^{1/2}.$$

By definition of energy we have $(\int_{|x|\leq R-t} |\partial_t v(t,x)|^2\ dx)^{1/2} \leq E(t)^{1/2}$. Writing $v(t,x) = \int_0^t \partial_t v(s,x)\ ds$ and using Minkowski's inequality and the fact that $v(0,x) = 0$ when $|x| \leq R$, we also see that

$$(\int_{|x|\leq R-t} |v(t,x)|^2\ dx)^{1/2} \leq \int_0^t E(s)^{1/2}\ ds.$$

Dividing out by $E(t)^{1/2}$, we conclude that

$$\partial_t E(t)^{1/2} \lesssim_{u,v} \int_0^t E(s)^{1/2}\ ds$$

which after integration in t (and recalling that $E(0) = 0$) yields

$$E(t)^{1/2} \lesssim_{u,v} t \int_0^t E(s)^{1/2}\ ds.$$

Applying Gronwall's inequality (Theorem 1.10) we conclude $E(t) = 0$ for all $0 \leq t \leq R$, and the claim easily follows. $\square$

The classical theory is generally sufficient for very smooth initial data (and very smooth nonlinearities $u \mapsto \mu|u|^{p-1}u$), but for rougher data one and nonlinearities must adopt a different approach. Because the differential formulation of the problem (3.1) requires so much differentiability, it is often better to work instead with the integral formulation of the equation,

$$u(t) = e^{i(t-t_0)\Delta/2}u_0 - i\mu \int_{t_0}^t e^{i(t-t')\Delta/2}(|u(t')|^{p-1}u(t'))\ dt'; \tag{3.22}$$

for NLS and

$$\begin{aligned} u(t) = \cos((t-t_0)\sqrt{-\Delta})u_0 + \frac{\sin((t-t_0)\sqrt{-\Delta})}{\sqrt{-\Delta}}u_1 \\ - \mu \int_{t_0}^t \frac{\sin((t-t_0)\sqrt{-\Delta})}{\sqrt{-\Delta}}(|u|^{p-1}u(t'))\ dt' \end{aligned} \tag{3.23}$$

for NLW; these equations can make sense even when u is a tempered distribution which lies locally in $L^p_t L^p_x$. We refer to such solutions as *distributional solutions* to the equation. When u has sufficient smoothness and regularity, these solutions coincide with classical solutions, but are more general in the case when u is rough.

Typically, the initial datum u_0 will also lie in a Sobolev space such as $H^s_x(\mathbf{R}^d)$. Recall (from the Fourier transform) that if $u_0 \in H^s_x(\mathbf{R}^d)$, then $e^{it\Delta/2}u_0 \in C^0_t H^s_x(\mathbf{R} \times \mathbf{R}^d) \cap L^\infty_t H^s_x(\mathbf{R} \times \mathbf{R}^d)$. Inspired by this, we distinguish two subclasses of distributional solution:

- A *strong H^s_x solution* to (3.22) on a time interval I is a distributional solution which also lies in $C^0_{t,\mathrm{loc}} H^s_x(I \times \mathbf{R}^d)$.
- A *weak H^s_x solution* to (3.22) on a time interval I is a distributional solution which also lies in $L^\infty_{t,\mathrm{loc}} H^s_x(I \times \mathbf{R}^d)$.

Similarly, we can define a strong $H^s_x \times H^{s-1}_x$ solution to (3.23) to be a distributional solution which also lies in $C^0_{t,\mathrm{loc}} H^s_x \cap C^1_{t,\mathrm{loc}} H^{s-1}_x$, while a weak solution lies in $L^\infty_{t,\mathrm{loc}} H^s_x$ with one (weak) time derivative in $L^\infty_{t,\mathrm{loc}} H^{s-1}_x$.

These definitions correspond to the notions of strong and weak solutions for ODE discussed in Section 1.1, though unfortunately in the PDE setting it is usually not known whether these notions are equivalent to each other. Generally speaking, the category of strong H^s_x solutions is the broadest category of solution in which we can hope to have a good existence and uniqueness theory; for weak H^s_x solutions one typically can hope to have existence but not uniqueness. In some cases it is possible to use the formula (3.22) to show that all weak solutions are automatically strong (as in Lemma 1.3) but this generally only happens when s is large (and one also needs the nonlinearity to be fairly smooth); see for instance Exercise 3.12. As a rule of thumb, perturbative methods such as Duhamel iteration tend to yield strong solutions, whereas weak compactness methods such as viscosity methods tend to only generate weak solutions (see Exercise 3.56).

With strong H^s_x solutions, $u(t)$ and $e^{i(t-t_0)\Delta/2}u_0$ varies continuously in t and so one can make sense of (3.22) for *all* times $t \in I$ (as opposed to almost every time t, or in a weak distributional sense). In particular a strong H^s_x solution obeys the initial condition $u(t_0) = u_0$ in the usual classical sense. Also, the notion of a strong solution is stable under time translation or time reversal, and one can glue together two strong solutions with overlapping intervals of existence; see Exercises 3.10, 3.11.

Of course, with such a low level of regularity it is not obvious at all how to use the equation (3.22) to justify other desirable properties of a solution, such as conservation laws or uniqueness, even when the solution is known to be a strong H^s_x solution. To do this one often needs to strengthen the notion of a strong H^s_x solution even further, by adding some additional properties of the solution map $u_0 \mapsto u$. One particularly successful such strengthening is the notion of a *wellposed solution*.

Definition 3.4 (Wellposedness). We say that the problem (3.1) is *locally wellposed* in $H^s_x(\mathbf{R}^d)$ if for any $u_0^* \in \mathbf{R}^d$ there exists a time $T > 0$ and an open ball B in $H^s_x(\mathbf{R}^d)$ containing u_0^*, and a subset X of $C^0_t H^s_x([-T,T] \times \mathbf{R}^d)$, such that for each $u_0 \in B$ there exists a strong unique solution $u \in X$ to the integral equation (3.22), and furthermore the map $u_0 \mapsto u$ is continuous from B (with the H^s_x topology) to X (with the $C^0_t H^s_x([-T,T] \times \mathbf{R}^d)$). We refer to this strong solution u as the *H^s_x-wellposed solution* to the Cauchy problem (3.1) with the specified initial datum u_0. If we can take $X = C^0_t H^s_x([-T,T] \times \mathbf{R}^d)$ then we say that the wellposedness is *unconditional*; if we can take T arbitrarily large[14] we say the wellposedness is *global* rather than local. If the time T depends only on the H^s_x norm of the initial

[14]This is strictly weaker than as asking for $T = +\infty$, which would be a *uniformly global* wellposedness assertion which would imply, among other things, that the H^s_x norm of $u(t)$ stays bounded as $t \to \infty$ (i.e. u lies in $C^0_t H^s_x(\mathbf{R} \times \mathbf{R}^d)$ rather than just $C^0_{t,\mathrm{loc}} H^s_x(\mathbf{R} \times \mathbf{R}^d)$). Obtaining such uniformly global bounds is possible for certain defocusing equations, and is a subset of the scattering theory developed in Section 3.6.

datum we say the wellposedness is in the *subcritical sense*, otherwise it is in the *critical sense*. We say that the wellposedness is *uniform* if the solution map $u_0 \mapsto u$ is uniformly continuous from B to X; similarly we define the notion of Lipschitz wellposedness, C^k wellposedness for $k = 1, 2, \ldots$, and analytic wellposedness.

REMARK 3.5. One can of course adapt this definition to other equations. For the nonlinear wave equation (3.2), the initial data class is $H^s_x(\mathbf{R}^d) \times H^{s-1}_x(\mathbf{R}^d)$ instead of H^s_x, and the solution should lie in $C^0_t H^s_x \cap C^1_t H^{s-1}_x$ instead of $C^0_t H^s_x$, but otherwise the definition is the same. One can also easily replace the Sobolev space H^s_x with other variants such as the homogeneous Sobolev space $\dot{H}^s_x$, though it is advisable to stick to spaces which are preserved by the *linear* evolution, since otherwise there is very little chance that there will be any sort of wellposedness for the *nonlinear* evolution. (This is a major reason why we work with the L^2_x-based Sobolev spaces H^s_x in the first place.)

REMARK 3.6. In practice, the space X will be quite explicit; it is typically the space of all fields in $C^0_t H^s_x([-T, T] \times \mathbf{R}^d)$ which obey an additional integrability or regularity condition (i.e. they lie in some additional function space). In some cases one also imposes a smallness condition in X, though such conditions can usually be removed by additional arguments (for instance, by shrinking the time interval to ensure the smallness condition holds, and then using continuity arguments to re-extend the time interval). The space X is useful for understanding the development of singularities; typically, a solution needs to leave the space X in order for a singularity to develop.

Wellposed solutions are highly compatible with classical solutions. If the initial datum is smooth, then the wellposed solution and classical solution usually coincide; this usually follows from the uniqueness theory, as well as *persistence of regularity* results (which we shall discuss in the next section). If the initial datum is rough, then by approximating this datum by smooth data and taking advantage of the continuity properties of the solution one can usually represent the wellposed solution as the strong limit of classical solutions[15] in the $C^0_t H^s_x$ topology (and often in other topologies also). Note that this shows that the wellposed solution is *canonical* - it is the unique limit of the classical solutions generated by any sequence of smooth data converging to the initial datum, and so two wellposed classes of solutions corresponding to different regularities (or different spaces X) will automatically coincide on their common domain of initial data. Furthermore, wellposed solutions are often able to enjoy the conservation laws and other formal identities which would normally be reserved for classical solutions, by taking appropriate limits. In some cases one needs to regularise the nonlinearity in addition to the initial datum; in such situations the continuity of the solution map is not quite sufficient, and one needs to supplement it with some *stability* properties of the solution, so that near-solutions to the equation can be well approximated by genuine solutions. Such stability properties are of independent interest, both for theoretical reasons (such as understanding the asymptotic behaviour of solutions), and for physical reasons (because they help justify the heuristic assumptions that one used to arrive at that model). We shall see some examples of these properties in Section 3.7.

[15]In some cases one has to regularise the nonlinearity $\mu|u|^{p-1}u$ by smoothing it out at zero or tempering its growth at infinity, in order to obtain good classical solutions; we will ignore these technicalities.

Another common trick is to use the method of *a priori estimates* to obtain control on wellposed solutions. Suppose one wants to show that all wellposed solutions in a certain class and of a certain size are bounded in some norm Y by some constant M. Since one can approximate wellposed solutions by classical solutions, it typically suffices (using tools such as Fatou's lemma) to obtain the desired bound for classical solutions only. The Y norm then typically depends continuously on the time interval I, and so by using a continuity argument in time one can assume as a bootstrap hypothesis that the solution is bounded in Y by a larger constant such as $2M$. This reduces matters to establishing an *a priori* estimate; the desired conclusion is the same, namely that the Y norm is bounded by M, but now we can make the *a priori* assumptions that the solution is smooth, and is already bounded in Y by $2M$. These hypotheses can be immensely useful; the former hypothesis allows one to make all formal computations rigorous, and the latter hypothesis is often crucial in order to obtain control of nonlinear terms. Also, the method of a priori estimates can also exploit various delicate cancellations (such as *energy cancellations*) arising from the structure of the equation, which are not picked up in some other methods such as iteration methods (because the iterates do not solve the exact equation and so do not exhibit these cancellations).

One common way to construct wellposed solutions is to use iterative methods, such as Proposition 1.38. Such methods tend to yield a fairly strong type of wellposedness, and can reduce the task of constructing solutions to that of verifying a single multilinear or nonlinear estimate. However, when the regularity of the data is extremely low, or equation behaves in an extremely nonlinear fashion, then such methods can break down; indeed there are examples known where solutions still exist, but one does not have the strong type of wellposedness implied by a iterative argument (see for instance the discussion on the Benjamin-Ono equation in Section 4.4, or of the wave map equation in Chapter 6). In such situations one needs to either augment the iterative argument (using for instance some sort of gauge transformation), or else use a completely different approach. One such approach is the *viscosity method* (also known as the penalisation, weak compactness, or regularisation method). In this approach, one approximates the equation (3.1) by a smoother equation, in which the nonlinearity is smoothed out and bounded, and an additional dissipation term is added to ensure global existence (forward in time, at least). This gives a sequence of approximate solutions, which one can demonstrate to be uniformly bounded in some norm (e.g. the energy norm); the establishment of such *a priori* control on the regularised solutions is usually the most difficult task. One can then use weak compactness to extract a weak limit of these approximate solutions (see for instance Exercise 3.56). This procedure typically produces a weak solution to the original equation without much difficulty, but it is often significantly harder to upgrade this solution to a strong solution or to establish wellposedness properties such as uniqueness, continuous dependence on the data, or persistence of regularity; also, the conservation laws are often not preserved by weak limits (though one can often obtain monotonicity of the conserved quantity, at least, by tools such as Fatou's lemma), and it often requires a non-trivial amount of additional effort to establish such laws[16].

[16]To give an example, the notorious global regularity problem for the Navier-Stokes equations remains open, despite the construction of global weak solutions by Leray over seventy years ago, in 1934!

REMARK 3.7. It is of interest to search for other ways to build solutions beyond the two standard methods of iteration and regularisation. One variant of the iteration method which is occasionally useful is the *Nash-Moser iteration method*, which is a PDE version of Newton's method for finding roots of equations. The iterates in this method tend to lose regularity with each iteration, but this is counteracted by the extremely rapid convergence of the iteration scheme. For other types of PDE (notably elliptic PDE), variational and topological methods have been very effective in constructing solutions, but so far these methods have not been particularly successful when applied to nonlinear dispersive or wave equations (though the induction on energy method, which we discuss in Section 5.4, can be thought of as a type of variational approach, while the continuity method from Section 1.3 is a crude example of a topological approach). Another speculative possibility is that probabilistic constructions of solutions, valid for *almost all* initial data rather than all initial data, may eventually be more powerful than the current deterministic methods, especially for supercritical equations where the deterministic methods appear to be useless. This may require utilizing ideas from thermodynamics, such as the use of invariant Gibbs measures and similar devices.

EXERCISE 3.6 (Preservation of reality). Show that if a classical solution to a NLW is real-valued at one time t_0, then it is real-valued for all other times for which the classical solution exists. (Use uniqueness and conjugation invariance.)

EXERCISE 3.7 (Preservation of symmetry). Let I be a time interval and let $t_0 \in I$. Suppose $u \in C^2_{t,x,\mathrm{loc}}(I \times \mathbf{R}^d \to \mathbf{C})$ is a classical solution to a NLS (resp. NLW) such that $u(t_0)$ (resp. $u[t_0]$) is spherically symmetric. For NLS, we furthermore require that u obey the boundedness decay conditions in Proposition 3.2. Prove that $u(t)$ is in fact spherically symmetric for all times $t \in I$.

EXERCISE 3.8 (Descent of NLS). Suppose that a periodic NLS on a torus $\mathbf{T}^{d+1}$ is locally wellposed in $H^s_x(\mathbf{T}^{d+1})$ in either the subcritical or critical sense. Show that the same NLS, but placed on the torus $\mathbf{T}^d$ of one smaller dimension, is also locally wellposed in $H^s_x(\mathbf{T}^d)$ in the same sense. The same statement also holds for global wellposedness, and with NLS replaced by NLW (but of course we replace H^s_x by $H^s_x \times H^{s-1}_x$ in that case).

EXERCISE 3.9 (Localised blowup for focusing NLW). Show that for any defocusing NLW there exists smooth compactly supported initial data (u_0, u_1) for which the Cauchy problem (3.2) does not admit a global classical solution. (Hint: take the initial data for (3.6) and truncate it smoothly to be compactly supported. Now argue by contradiction using Proposition 3.3.)

EXERCISE 3.10 (Time shifting of strong solutions). Let I be a time interval containing t_0, and let u be a strong H^s_x solution to (3.1) with initial datum $u(t_0) = u_0$. Let t_1 be any other time in I, and let $u_1 := u(t_1)$. Show that u is also a strong H^s_x solution to (3.1) with initial datum $u(t_1) = u_1$. Thus the notion of a strong solution is independent of the initial time. Obtain a similar result for the NLW (3.2). Also, show that the field $\tilde{u}(t,x) := \overline{u(-t,x)}$ is a strong H^s_x solution to (3.1) on the interval $-I$ with initial datum $u(-t_0) = \overline{u_0}$. (These results can fail for weak solutions; see Exercise 3.15.)

EXERCISE 3.11 (Gluing of strong solutions). Let I, I' be intervals which intersect at a single time t_0. Suppose that u, u' are strong H^s_x solutions to (3.1) on

$I \times \mathbf{R}^d$ and $I' \times \mathbf{R}^d$ respectively with initial data $u(t_0) = u'(t_0) = u_0$. Show that the combined field $\tilde{u}$ on $(I \cup I') \times \mathbf{R}^d$ is also a strong solution to (3.1). Obtain the similar result for the NLW equation where u and u' have matching initial positions *and* initial velocities. This exercise, combined with Exercise 3.10, shows that there is no difficulty gluing together strong solutions on adjacent time intervals to create a unified strong solution.

EXERCISE 3.12. Let p be an odd integer and $s > d/2$. Show that every weak H^s_x solution to (3.1) is also a strong H^s_x solution. (You will need the fact that H^s_x is an algebra; see Lemma A.8.)

EXERCISE 3.13 (Local uniqueness implies global uniqueness). Fix p, d, μ, s and suppose that one knows that for any time t_0 and initial datum $u_0 \in H^s_x(\mathbf{R}^d)$, there exists an open time interval I containing t_0 such that there is at most one strong H^s_x solution to (3.1) on $I \times \mathbf{R}^d$ (i.e. one has local uniqueness of strong solutions). Show that this automatically implies global uniqueness of strong solutions, or more precisely for any time interval J containing t_0 that there is at most one strong H^s_x solution to (3.1) on $J \times \mathbf{R}^d$. (Hint: prove by contradiction and use a continuity method.)

3.3. Local existence theory

The greatest challenge to any thinker is stating the problem in a way that will allow a solution. (Bertrand Russell)

We are now ready to construct solutions to NLS and NLW, and analyze the strength of such solutions, in the senses discussed in Section 3.2. We will not attempt to give the most complete results here, but instead give a sample of results which illustrate the basic iteration method[17]. The underlying idea of this method is simple - select spaces $\mathcal{S}$ and $\mathcal{N}$ in which to hold the solution u and the nonlinearity $\mu|u|^{p-1}u$ respectively, at which point the problem reduces to that of establishing linear and nonlinear estimates in $\mathcal{S}$ and $\mathcal{N}$. The selection of these spaces, however, is something of an art rather than a science; there are some standard spaces that work well in many situations, and one can analyze individual iterates to suggest what spaces are likely to work, and which ones will not; however there is certainly no "universal iteration space" that can cover all cases, and in one usually needs to tailor the precise spaces to the equation at hand[18].

[17] This method seems to be the best method for solving NLS and NLW, at least in the subcritical and critical settings, with the Strichartz estimates (possibly with some Besov-type augmentations) being the ideal tool to close the iteration. For the less semilinear equations studied in later chapters, which contain derivatives in the nonlinearity, the iteration method often requires more ingenious choices of spaces and estimates, as well as some additional tricks such as gauge transformations, and thus face some nontrivial competition from other methods such as viscosity methods or methods based on exploiting energy cancellation to obtain a priori estimates, which have their own strengths and weaknesses.

[18] One common way to proceed here is to compute the first few nonlinear iterates in the Duhamel iteration scheme, and see what spaces one can estimate them in; if one can place them all in a common space then this suggests what choices of $\mathcal{S}$ and $\mathcal{N}$ to use. Another way is to work in reverse, starting with the quantity $\|u\|_{C^0_t H^s_x}$ (which one needs to control to obtain wellposedness) and estimating it in terms of other norms of u using the Duhamel formula; in doing so it will become apparent what types of norms need to be controlled in order to have a chance of closing the iteration. Typically, one needs to control u in those spaces in which the

A systematic study of the wellposedness theory for NLS can be found in [**Caz2**], and for NLW in [**Sog**]. A basic heuristic for NLS is that one has local wellposedness in $H^s_x(\mathbf{R}^d)$ if and only if $s \geq \max(s_c, 0)$, where $s_c := \frac{d}{2} - \frac{2}{p-1}$ is the scale-invariant regularity (and 0 is the Galilean-invariant regularity); the corresponding heuristic for NLW is that one has local wellposedness in $H^s_x(\mathbf{R}^d)$ if and only if $s \geq \max(s_c, s_l, 0)$, where $s_l := \frac{d+1}{4} - \frac{1}{p-1}$ is the regularity associated to the Lorentz invariance. This heuristic is only partially accurate (wellposedness can break down or become weaker when the nonlinearity becomes very rough compared to the regularity s, and in the case of the NLW there are still some very low regularities and exponents for which the problem is not fully resolved, see [**Tao**].

To simplify the notation let us use the time translation invariance to fix the initial time t_0 to equal zero.

Let us begin with classical solutions to NLS. It turns out that to construct classical solutions it is more convenient to work in Sobolev spaces $H^s_x(\mathbf{R}^d)$ or weighted Sobolev spaces $H^{k,k}_x(\mathbf{R}^d)$ (for suitably high values of s, k) than in more classical spaces such as $C^k_x(\mathbf{R}^d)$; the main reason for this is that the linear propagator $e^{it\Delta/2}$ preserves $H^s_x(\mathbf{R}^d)$ and are locally bounded on $H^{k,k}_x(\mathbf{R}^d)$ (see Exercise 2.52) but does not preserve $C^k_x(\mathbf{R}^d)$ (cf. Exercise 2.35). To avoid some technicalities, let us restrict attention for now to algebraic nonlinearities, so that p is an odd integer[19].

PROPOSITION 3.8 (Classical NLS solutions). *Let $p > 1$ be an odd integer, let $k > d/2$ be an integer, and let $\mu = \pm 1$. Then the NLS (3.1) is unconditionally locally wellposed in $H^{k,k}_x(\mathbf{R}^d)$ in the subcritical sense. More specifically, for any $R > 0$ there exists a $T = T(k, d, p, R) > 0$ such that for all u_0 in the ball $B_R := \{u_0 \in H^{k,k}_x(\mathbf{R}^d) : \|u_0\|_{H^{k,k}_x(\mathbf{R}^d)} < R\}$ there exists a unique solution $u \in C^0_t H^{k,k}_x([-T,T] \times \mathbf{R}^d)$ to (3.1). Furthermore the map $u_0 \mapsto u$ from B_R to $C^0_t H^{k,k}_x([-T,T] \times \mathbf{R}^d)$ is Lipschitz continuous.*

The same statements hold if $H^{k,k}_x$ is replaced by H^s_x for any $s > d/2$ (not necessarily an integer).

REMARK 3.9. This proposition implies that for a Schwartz initial datum $u_0 \in \mathcal{S}_x(\mathbf{R}^d)$ and an odd integer p one has a maximal Schwartz solution $u \in C^\infty_{t,\mathrm{loc}} \mathcal{S}_x(I \times \mathbf{R}^d)$ to any given algebraic NLS for some open interval I containing 0, which is unique by Proposition 3.2. Note that one can use the equation (3.1) to trade regularity in space for regularity in time (at a two-for-one conversion ratio), and so solutions which are Schwartz in space will also be smooth in time.

---

linear solution is already known to be controlled in. We will use some of these heuristics when studying the existence problem for other PDE in later chapters.

[19]When the nonlinearity is rough, it is often necessary to regularise it, for instance by replacing $|u|^{p-1}u$ by $(\varepsilon^2 + |u|^2)^{(p-1)/2}u$ for some $\varepsilon > 0$ and then setting $\varepsilon \to 0$, in order to have a concept of a smooth solution that one can use to approximate rough solutions to the original equation; see for instance Exercise 3.55. In some cases one can use Schauder estimates (Lemma A.9) as a substitute for product estimates such as (3.24). As these technical issues are rather dull, we shall try to avoid them as much as possible.

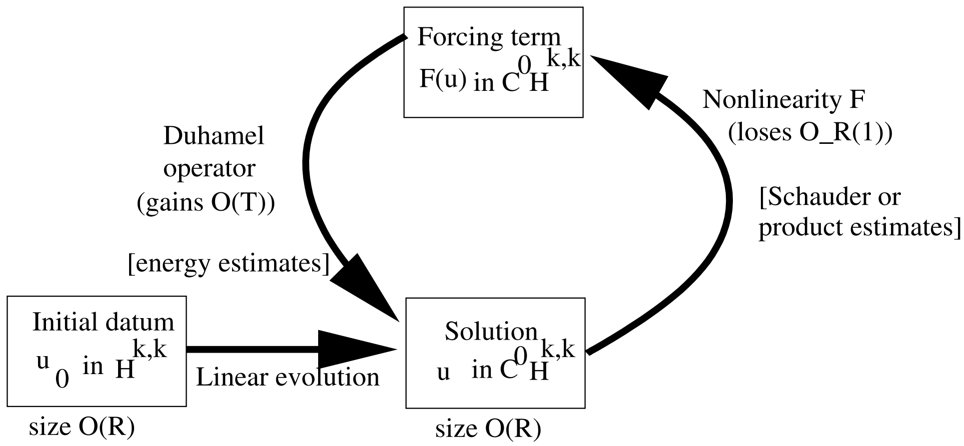

FIGURE 1. The classical energy method iteration scheme; the iteration closes if T is sufficiently small depending on R. One can also replace $H^{k,k}_x$ by H^s_x for $s > d/2$ without difficulty.

PROOF. The key observations[20] are Exercise 2.52, and the fact that the space $H^{k,k}_x(\mathbf{R}^d)$ is a Banach space algebra:

$$\|fg\|_{H^{k,k}_x(\mathbf{R}^d)} \lesssim_{k,d} \|f\|_{H^{k,k}_x(\mathbf{R}^d)} \|g\|_{H^{k,k}_x(\mathbf{R}^d)}. \tag{3.24}$$

We leave this estimate (a variant of (A.18)) as an exercise.

Let us now fix R, and let $0 < T < 1$ be a small time to be chosen later. We shall use Proposition 1.38 with $\mathcal{S} = \mathcal{N} = C^0_t H^{k,k}_x([-T,T] \times \mathbf{R}^d)$, with linear operator $D : \mathcal{N} \to \mathcal{S}$ set equal to the Duhamel operator

$$DF(t) := -i \int_0^t e^{i(t-s)\Delta/2} F(s)\, ds$$

and the nonlinear operator $N : \mathcal{S} \to \mathcal{N}$ set equal to

$$Nu(t) := \mu |u(t)|^{p-1} u(t).$$

(See Figure 1.) From Minkowski's inequality and Exercise (2.52) we verify the bound (1.51) with $C_0 = O_{k,d}(T)$. From the algebra property we see that

$$\|Nu\|_{\mathcal{N}} \lesssim_{k,d,p,R} \|u\|_{\mathcal{S}}; \quad \|Nu - Nv\|_{\mathcal{N}} \lesssim_{k,d,p,R} \|u - v\|_{\mathcal{S}}$$

whenever $u, v \in \mathcal{S}$ are such that $\|u\|_S, \|v\|_S \lesssim_{k,d} R$. If we choose T sufficiently small depending on k, d, p, R, we can thus apply Proposition 1.38 and conclude that for all $u_{\mathrm{lin}} \in \mathcal{S}$ with $\|u_{\mathrm{lin}}\|_{\mathcal{S}} \lesssim R$ there exists a unique solution $u \in \mathcal{S}$ to (1.50) with $\|u\|_{\mathcal{S}} \lesssim R$. Applying this in particular to $u_{\mathrm{lin}} := e^{it\Delta/2} u_0$ (and using Exercise 2.52) we obtain a solution to (3.1) (in the Duhamel integral form), with the map $u_0 \mapsto u$ being Lipschitz continuous from the ball in $H^{k,k}_x(\mathbf{R}^d)$ of radius $O(R)$ to $C^0_t H^{k,k}_x([0,T] \times \mathbf{R}^d)$.

The above argument establishes uniqueness so long as we restrict the $\mathcal{S}$ norm of solutions $\mathcal{S}$ to be $O(R)$. But since the $H^{k,k}$ norm of u_0) is at most R at time

[20] Indeed, this argument is quite abstract, and applies to any Banach algebra which is preserved by the linear flow. This is known as the *semigroup method* and has been extensively developed, see for instance [**Kat7**].

zero, one can in fact obtain unconditional uniqueness in $\mathcal{S}$ by a standard continuity argument. Specifically, let $u \in \mathcal{S}$ be the solution constructed by the above method, then we have $\|u\|_{\mathcal{S}} \leq C_1 R$ for some absolute constant C_1. Let $u^* \in \mathcal{S}$ be another solution. For any $0 \leq t \leq T$ let $\mathbf{H}(t)$ be the property that $\|u\|_{C_t^0 H_x^{k,k}([-t,t]\times\mathbf{R}^d)} \leq 2C_1 R$, and let $\mathbf{C}(t)$ be the property that $\|u\|_{C_t^0 H_x^{k,k}([-t,t]\times\mathbf{R}^d)} \leq C_1 R$. Then the assumptions (b),(c),(d) of Proposition 1.21 are clear, and property (a) follows from the uniqueness theory already established (if T is suitably small). This gives the unconditional uniqueness.

The same argument works with $H_x^{k,k}$ replaced by H_x^s since one still has the crucial algebra property

$$\|fg\|_{H_x^s(\mathbf{R}^d)} \lesssim_{k,d} \|f\|_{H_x^s(\mathbf{R}^d)} \|g\|_{H_x^s(\mathbf{R}^d)};$$

see Lemma A.8. $\square$

REMARK 3.10. This argument was completely insensitive to the sign μ of the nonlinearity; this is a typical feature of the local existence theory. The global existence theory, however, will be much more sensitive to the sign of the nonlinearity.

The above result shows that one has unconditional local wellposedness in $H_x^s(\mathbf{R}^d)$ for an algebraic NLS equation for any $s > d/2$. This shows (using the argument in the proof of Theorem 1.17) that given any $u_0 \in H_x^s(\mathbf{R}^d)$, there exists a unique maximal interval of existence I and a unique solution $u \in C_t^0 H_x^s(I \times \mathbf{R}^d)$. The size of this interval can only shrink to zero if the $H_x^s(\mathbf{R}^d)$ norm of the data goes to infinity. Hence if I has a finite endpoint T, then the $H_x^s(\mathbf{R}^d)$ norm of $u(t)$ will go to infinity as t approaches T. Thus the maximal interval is necessarily open, as one cannot possibly continue a solution in $C_t^0 H_x^s$ at a point where the H_x^s norm is going to infinity. An identical result also holds in the periodic case $\mathbf{T}^n$.

To rephrase the above discussion, if a solution to NLS (with algebraic nonlinearity) is initially in $H_x^s(\mathbf{R}^d)$ with $s > d/2$, then it can be continued in a unique continuous manner in $H_x^s(\mathbf{R}^d)$ so long as the $H_x^s(\mathbf{R}^d)$ stays bounded. Let us informally call a norm X a *controlling norm*[21] for this equation if the boundedness of this X norm is enough to ensure continuation of smooth solutions. Thus we now know that any sufficiently high regularity Sobolev norm is a controlling norm for any algebraic NLS. It is of interest to obtain controlling norms which are as low regularity as possible. As a rule of thumb, any reasonable norm which is subcritical, or which is critical and involves some integration in time, has a chance of being a controlling norm. For instance, we have

PROPOSITION 3.11 (Persistence of regularity). *Let I be a time interval containing $t_0 = 0$, let $s \geq 0$, and let $u \in C_t^0 H_x^s(I \times \mathbf{R}^d)$ be a strong H_x^s solution to an algebraic NLS equation. If the quantity $\|u\|_{L_t^{p-1} L_x^\infty(I\times\mathbf{R}^d)}$ is finite, then $u(t)$ is uniformly bounded in H_x^s, indeed we have*

$$\|u\|_{L_t^\infty H_x^s(I\times\mathbf{R}^d)} \leq \|u(0)\|_{H_x^s} \exp(C_{p,s,d} \|u\|^p_{L_t^{p-1} L_x^\infty(I\times\mathbf{R}^d)}). \tag{3.25}$$

In particular, if I has finite length $|I|$, then we have

$$\|u\|_{L_t^\infty H_x^s(I\times\mathbf{R}^d)} \leq \|u(0)\|_{H_x^s} \exp(C_{p,s,d} |I|^{p/(p-1)} \|u\|^p_{L_t^\infty L_x^\infty(I\times\mathbf{R}^d)}).$$

---

[21]We shall be somewhat vague as to whether X is a spatial norm or a spacetime norm. Both types of norms are useful types of controlling norms.

PROOF. We use the energy method. By time reversal symmetry we may take $I = [0, T]$ for some $T > 0$. From the Duhamel formula

$$u(t) = e^{it\Delta/2}u(0) - i\mu \int_0^t e^{i(t-t')\Delta/2}|u(t')|^{p-1}u(t')\, dt'$$

and the unitary nature of $e^{it\Delta/2}$ on $H^s_x(\mathbf{R}^d)$, we conclude from Minkowski's inequality that

$$\|u(t)\|_{H^s_x(\mathbf{R}^d)} \leq \|u(0)\|_{H^s_x(\mathbf{R}^d)} + \int_0^t \||u(t')|^{p-1}u(t')\|_{H^s_x(\mathbf{R}^d)}\, dt'.$$

We expand $|u(t')|^{p-1}u(t')$ as a polynomial of degree p in $u(t')$ and its complex conjugate $\overline{u(t')}$. Applying Lemma A.8 repeatedly we have

$$\||u(t')|^{p-1}u(t')\|_{H^s_x(\mathbf{R}^d)} \lesssim_{p,s,d} \|u(t')\|_{H^s_x}\|u(t')\|^{p-1}_{L^\infty_x}.$$

The claim now follows from Gronwall's inequality. □

REMARK 3.12. In the converse direction, any field in $C^0_t H^s_x$ will be locally in $L^\infty_t L^\infty_x$ and hence in $L^{p-1}_t L^\infty_x$ by Sobolev embedding. Thus one can continue a solution in H^s_x for $s > d/2$ if and only if the $L^{p-1}_t L^\infty_x$ norm remains locally finite; in particular, if the solution blows up (fails to remain smooth) at some time T_*, then the solution must become unbounded near the blowup time T_* (which justifies the terminology of "blowup"). Since these blowup criteria are independent of H^s_x, we thus observe if an initial datum u_0 lies both in a lower regularity Sobolev space $H^{s_1}_x$ and a higher regularity Sobolev space $H^{s_2}_x$, where $s_2 > s_1 > d/2$, then the solution can be continued in one regularity for precisely the same amount of time as it can be continued in another; thus it is not possible to develop a singularity which causes the $H^{s_2}_x$ norm to blow up while the $H^{s_1}_x$ norm remains bounded. This phenomenon (known as *persistence of regularity* - if a solution map preserves rough regularities, then it also preserves smooth regularities) is typical of all regularities for which one has a strong wellposedness theory, but can fail[22] for regularities that are excessively low (see Exercise 3.15). Note that from the time reversal symmetry (and uniqueness) we also see that the regularity cannot spontaneously increase: if a solution lies in $C^0_t H^{s_1}_x(I \times \mathbf{R}^d)$ and is *not* in $H^{s_2}_x$ at some initial time t_0, then it will also not be in $H^{s_2}_x$ for any later (or earlier) time. Thus regularity is neither created nor destroyed in the Sobolev scale, so long as the solution persists. This is in contrast to dissipative equations such as the heat equation, which is smoothing when evolved forwards in time and illposed when evolved backwards in time.

REMARK 3.13. Observe that the $L^{p-1}_t L^\infty_x$ norm that controls the persistence of regularity here is invariant under the scaling (3.9). This is closely related to the fact that no factor of $|I|$ appears in (3.25). It has the consequence that the bound (3.25) holds even for unbounded intervals I, and thus shows that one can keep the H^s_x norm of a solution $u(t)$ bounded even as $t \to \pm\infty$, provided that one can somehow keep the global $L^{p-1}_t L^\infty_x$ norm bounded also. This result is an instance of a general principle, that scale-invariant global spacetime integrability bounds imply

[22]More precisely, persistence of regularity can fail when there is no control of critical or subcritical type on the solution. Note that in Proposition 3.11, the regularity H^s that was being controlled could be subcritical, critical, or supercritical; the important thing is that the controlling norm $L^{p-1}_t L^\infty_x$ is critical. More generally, it is permissible in a scale invariant argument to use *one* quantity which is subcritical or supercritical, so long as all the bounds are linear in that quantity.

good asymptotic behaviour at infinity; this philosophy will be particularly apparent in the scattering theory in Section 3.6.

REMARK 3.14. The scale-invariance of the controlling norm is a general phenomenon; controlling norms are either critical (invariant with respect to scaling) or subcritical (they only become scale-invariant if multiplied by some positive power of the length $|I|$ of the time interval, as is the case for instance with the $L_t^\infty L_x^\infty$ norm here). All other things being equal, it is preferable to use a critical controlling norms than a subcritical one (provided of course that a critical controlling norm can be located in the first place) as they are generally smaller, and can yield global control on solutions rather than just local control. Norms which are supercritical, on the other hand, cannot possibly be controlling norms (this would lead to absurd results, such as the spacetime bounds for large time intervals being *smaller* than the bounds for small time intervals). The most famous example of this is the three-dimensional Navier-Stokes equations, which enjoy boundedness of kinetic energy but for which global existence of smooth solutions is a major unsolved problem, in large part because the kinetic energy turns out to be a supercritical quantity in three spatial dimensions and thus cannot be a controlling norm. In practice, possession of a bound on a supercritical quantity has proven to be of little use in the global regularity theory, unless combined with additional information such as a bound on a subcritical quantity (so that one can interpolate between the two to obtain critical controlling quantities). More recently, techniques have been developed to combine supercritical control with existing *critical* control, to obtain even better critical control; in particular, in the energy-critical defocusing NLS, the mass and momentum conservation laws (which are supercritical in this case) can be used to limit the concentration behaviour of energy towards higher frequencies and thus yield control of other critical quantities such as certain spacetime Lebesgue norms. See Chapter 5.

We now turn from classical solutions to less regular solutions, in particular considering solutions in H_x^s for $s \leq d/2$. In this case, we no longer expect the solution to lie in $L_x^\infty(\mathbf{R}^d)$ for all time, since H_x^s no longer embeds into L_x^∞. However, the Strichartz estimates in Theorem 2.3 suggest that one can still lie in *time-averaged* L_x^∞ spaces such as $L_t^{p-1}L_x^\infty(\mathbf{R}^d)$ for regularities lower than $d/2$; intuitively, this reflects the fact that while an H_x^s function can focus much of its "energy" at one spatial point to create a large L_x^∞ norm (cf. Proposition A.4), the dispersive effects of the Schrödinger evolution imply that this focus cannot be maintained for more than a short period of time. Of course, this is only a heuristic, because the Strichartz estimates only apply directly to the *linear* Schrödinger evolution rather than the nonlinear one, however it does suggest that some sort of iterative argument, using the Strichartz estimates to treat the nonlinear equation as a perturbation of the linear one, can work.

To do this, it is convenient to create a single space S^s which captures all the Strichartz norms at a certain regularity H_x^s simultaneously. We illustrate this first with the L_x^2 theory. We introduce the *Strichartz space* $S^0(I \times \mathbf{R}^d)$ for any time interval I, defined as the closure of the Schwartz functions under the norm[23]

$$\|u\|_{S^0(I\times\mathbf{R}^d)} := \sup_{(q,r)\text{ admissible}} \|u\|_{L^q_t L^r_x(I\times\mathbf{R}^d)},$$

where admissibility was defined in Theorem 2.3. In particular the S^0 norm controls the $C^0_t L^2_x$ norm. This norm is a Banach space and has a dual $N^0(I\times\mathbf{R}^d) := S^0(I\times\mathbf{R}^d)^*$; by construction we see that

$$\|F\|_{N^0(I\times\mathbf{R}^d)} \leq \|F\|_{L^{q'}_t L^{r'}_x(I\times\mathbf{R}^d)}$$

whenever the right-hand side is finite. The Strichartz estimates in Proposition 2.3 can then be combined into a unified estimate

$$\|u\|_{S^0(I\times\mathbf{R}^d)} \lesssim_d \|u(t_0)\|_{L^2_x(I\times\mathbf{R}^d)} + \|F\|_{N^0(I\times\mathbf{R}^d)} \tag{3.26}$$

whenever $t_0 \in I$ and $iu_t + \frac{1}{2}\Delta u = F$. Because of this estimate, one often expects to place L^2_x solutions of NLS in the space S^0, at least provided that one has some hope of placing the nonlinearity $F = \mu|u|^{p-1}u$ in the companion space N^0. A typical application of this estimate is

Proposition 3.15 (Subcritical L^2_x NLS solutions). [**Tsu**] *Let p be an L^2_x-subcritical exponent (so $1 < p < 1 + \frac{4}{d}$) and let $\mu = \pm 1$. Then the NLS (3.1) is locally wellposed in $L^2_x(\mathbf{R}^d)$ in the subcritical sense. More specifically, for any $R > 0$ there exists a $T = T(k,d,p,R) > 0$ such that for all u_0 in the ball $B_R := \{u_0 \in L^2_x(\mathbf{R}^d) : \|u_0\|_{L^2_x(\mathbf{R}^d)} < R\}$ there exists a unique strong L^2_x solution u to (3.1) in the space $S^0([-T,T]\times\mathbf{R}^d) \subset C^0_t L^2_x([-T,T]\times\mathbf{R}^d)$. Furthermore the map $u_0 \mapsto u$ from B_R to $S^0([-T,T]\times\mathbf{R}^d)$ is Lipschitz continuous.*

Remark 3.16. One can weaken the space $S^0([-T,T]\times\mathbf{R}^d)$ somewhat and still obtain uniqueness (see [**CWeis**], [**CWeis2**]). However, it is not known if one can replace S^0 by $C^0_t L^2_x$ and thus obtain an unconditional wellposedness result. In the next section we shall extend this local existence result to a global existence result.

Proof. We modify the proof of Proposition 3.8. Again we fix R and choose $T > 0$ later. We will apply Proposition 1.38 for a suitable choice of norms $\mathcal{S}, \mathcal{N}$ and some $\varepsilon > 0$; a specific instance of our scheme in the case $d = 1, p = 3$ is described in Figure 2. One such choice is to set $\mathcal{S} = S^0([-T,T]\times\mathbf{R}^d)$ and $\mathcal{N} = N^0([-T,T]\times\mathbf{R}^d)$. In order to place the u_{lin} in $B_{\varepsilon/2}$, we see from the Strichartz estimate (3.26) that we need to take $\varepsilon = C_1 R$ for some large constant $C_1 > 0$ (depending only on d). The estimate (1.51) also follows from (3.26) (for some large $C_0 > 0$ depending on d), so it remains to verify (1.52). In other words, we need to show that

$$\||u|^{p-1}u - |v|^{p-1}v\|_{N^0([-T,T]\times\mathbf{R}^d)} \leq \frac{1}{2C_0}\|u - v\|_{S^0([-T,T]\times\mathbf{R}^d)}$$

whenever $\|u\|_{S^0([-T,T]\times\mathbf{R}^d)}, \|v\|_{S^0([-T,T]\times\mathbf{R}^d)} \leq C_1 R$. It is convenient to introduce the exponent pair (q,r) by solving the equations

$$\frac{2}{q} + \frac{d}{r} = \frac{d}{2}; \quad \frac{p}{r} = \frac{1}{r'}.$$

[23]In the case $d = 2$ case, the set of admissible exponents is not compact, and so one has to truncate the supremum, for instance restricting $q \geq 2 + \varepsilon$ for some $\varepsilon > 0$, in order for the Strichartz constants to be uniform in the exponent. Also, in some endpoint applications it is more convenient to strengthen the norms S^0, N^0 to a certain Besov-space version of themselves. We ignore these technicalities to simplify the exposition.

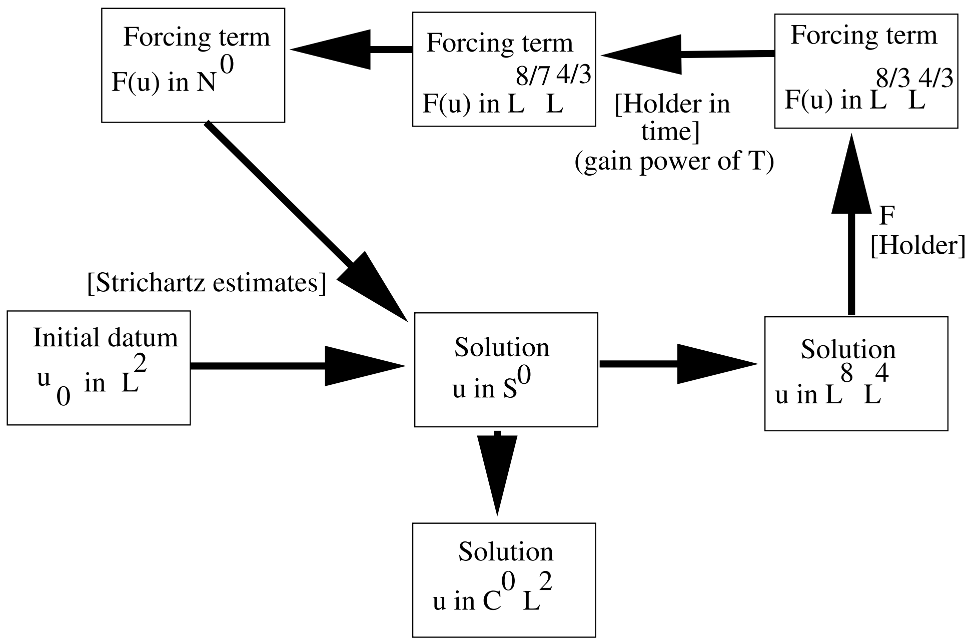

FIGURE 2. An iteration scheme in $L^2_x(\mathbf{R})$ for the the one-dimensional cubic NLS $d = 1, p = 3$ (which is L^2_x-subcritical), so that $q = 8$ and $r = 4$. In all the iteration schemes presented here, the sign μ of the nonlinearity is irrelevant. The subcritical nature of the equation allows for a gain of a power of T at some stage. This is not the simplest iteration scheme available for this equation, but it is rather representative. Note that the fact that u was a strong solution (i.e. $u \in C^0_t L^2_x$) is a byproduct of the argument but is not otherwise used in an essential way in the proof.

One can easily check using the hypothesis $1 < p < 1 + \frac{4}{n}$ that we have $2 < q < r < \infty$, so in particular (q, r) is admissible. In particular, we can estimate the N^0 norm by the $L^{q'}_t L^{r'}_x$ norm. Since $q < r$, we see that $\frac{p}{q} > \frac{1}{q'}$, so we may replace the $L^{q'}_t$ norm by the $L^{q/p}_t$ norm by paying a factor of T^α for some $\alpha > 0$. If we then use the elementary estimate

$$||u|^{p-1}u - |v|^{p-1}v| \lesssim_p |u - v|(|u|^{p-1} + |v|^{p-1}) \tag{3.27}$$

and Hölder's inequality, we conclude

$$\begin{aligned}
&\||u|^{p-1}u - |v|^{p-1}v\|_{N^0([-T,T]\times\mathbf{R}^d)}\\
&\lesssim_p T^\alpha \|u - v\|_{L^q_t L^r_x([-T,T]\times\mathbf{R}^d)}(\|u\|_{L^q_t L^r_x([-T,T]\times\mathbf{R}^d)} + \|v\|_{L^q_t L^r_x([-T,T]\times\mathbf{R}^d)})^{p-1}\\
&\lesssim_{p,C_1,R} T^\alpha \|u - v\|_{L^q_t L^r_x([-T,T]\times\mathbf{R}^d)}\\
&\leq T^\alpha \|u - v\|_{S^0([-T,T]\times\mathbf{R}^d)}
\end{aligned}$$

Thus we obtain (1.52) if T is chosen sufficiently small depending on p, C_1, R. We can then apply Proposition 1.38 to construct a solution in $\mathcal{S}$ to (3.1) with norm at most $C_1R/2$, which is unique among all solutions with norm at most C_1R, and the map $u_0 \mapsto u$ will be Lipschitz continuous. The requirement that the norm be at most C_1R can be dropped from the uniqueness conclusion by using a continuity argument as in Proposition 3.8. $\square$

In the critical case $p = 1 + \frac{4}{d}$ one still has wellposedness, but in the critical sense (so that the time of existence T depends on the profile of the datum and not just on the norm). More precisely, we have

PROPOSITION 3.17 (Critical L^2_x NLS solutions). [**Tsu**] *Let p be the L^2_x-critical exponent $p = 1 + \frac{4}{d}$ and let $\mu = \pm 1$. Then the NLS (3.1) is locally wellposed in $L^2_x(\mathbf{R}^d)$ in the critical sense. More specifically, given any $R > 0$ there exists $\varepsilon_0 = \varepsilon_0(R, d) > 0$, such that whenever $u_* \in L^2_x(\mathbf{R}^d)$ has norm at most R, and I is a time interval containing 0 such that*

$$\|e^{it\Delta/2}u_*\|_{L^{2(n+2)/n}_{t,x}(I\times\mathbf{R}^d)} \leq \varepsilon_0$$

then for any u_0 in the ball $B := \{u_0 \in L^2_x(\mathbf{R}^d) : \|u_0 - u_*\|_{L^2_x(\mathbf{R}^d)} \leq \varepsilon_0\}$ there exists a unique strong L^2_x solution $u \in S^0(I \times \mathbf{R}^d)$ to (3.1), and furthermore the map $u_0 \mapsto u$ is Lipschitz from B to $S^0(I \times \mathbf{R}^d)$ (of course, the Lipschitz constant will depend on u_*).

This proposition is proven similarly to Proposition 3.15 and is left to Exercise 3.18. Note that if the initial datum is sufficiently small in L^2_x norm, then this Proposition, combined with Strichartz estimates, yields global existence in time. If the initial datum is instead large, the Proposition combined with Strichartz estimates will still give local existence, because the global $L^{2(n+2)/n}_{t,x}$ norm of $e^{it\Delta/2}u_*$ will be finite, and hence can be localised to be small by choosing a sufficiently small time interval.

REMARK 3.18. This proposition and the preceding one should be compared against Principle 3.1. It turns out that the L^2_x theory becomes bad for supercritical powers $p > 1 + \frac{4}{d}$; see Section 3.8 for further discussion and results.

Similar results hold for other regularities, such as H^1_x. Here it is convenient to use the norms

$$\|u\|_{S^1(I\times\mathbf{R}^d)} := \|u\|_{S^0(I\times\mathbf{R}^d)} + \|\nabla u\|_{S^0(I\times\mathbf{R}^d)}$$

and

$$\|u\|_{N^1(I\times\mathbf{R}^d)} := \|u\|_{N^0(I\times\mathbf{R}^d)} + \|\nabla u\|_{N^0(I\times\mathbf{R}^d)}$$

Note that as the Schrödinger equation commutes with derivatives, we see from (3.26) that

$$\|u\|_{S^1(I\times\mathbf{R}^d)} \lesssim_d \|u(t_0)\|_{H^1_x(I\times\mathbf{R}^d)} + \|F\|_{N^1(I\times\mathbf{R}^d)}. \tag{3.28}$$

Let us give two sample results in dimension $d = 3$, in which the H^1_x-critical exponent is the quintic one $p = 5$:

PROPOSITION 3.19 ($H^1_x(\mathbf{R}^3)$ subcritical NLS solutions). *Let $\mu = \pm 1$. If $2 \leq p < 5$, then the NLS (3.1) is locally wellposed in $H^1_x(\mathbf{R}^3)$ in the subcritical sense.*

REMARK 3.20. For this Proposition and the next, the reader may wish to refer back to the Strichartz "game board" for Schrödinger equations on $H^1_x(\mathbf{R}^3)$ from Figure 1 of Chapter 2, and see how the various "moves" of Leibnitz, Hölder,

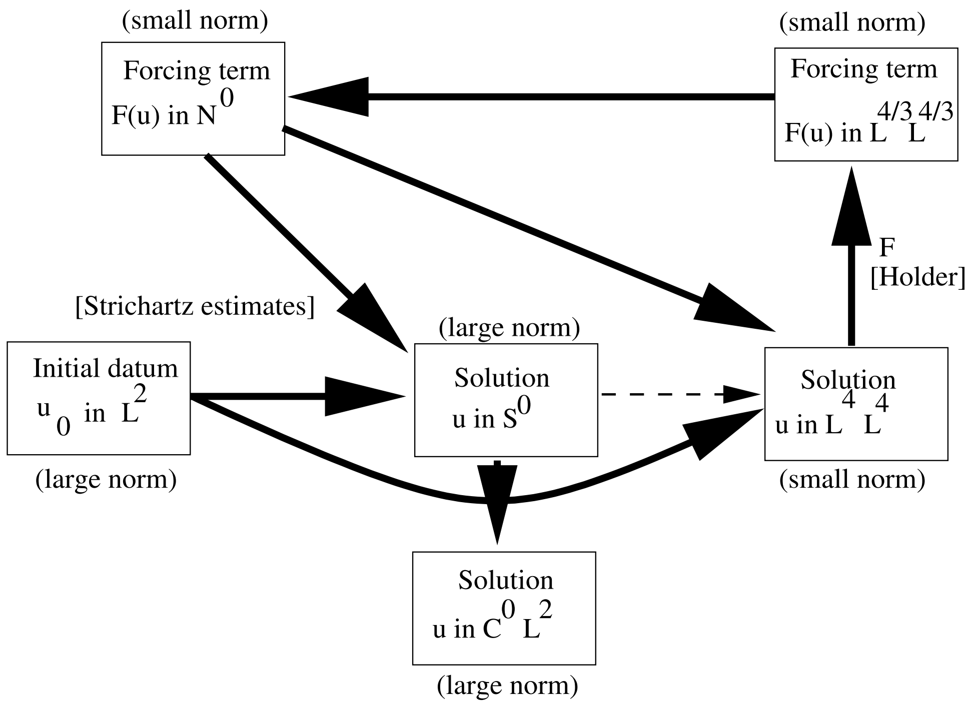

FIGURE 3. An iteration scheme in $L^2_x(\mathbf{R}^2)$ for the two-dimensional cubic NLS $d = 2, p = 3$ (which is L^2_x-critical). For small data one can simplify the scheme somewhat, but for large data it is important that the S^0 is allowed to be large, while the $L^4_{t,x}$ component of the S^0 kept small; thus the main loop of the iteration should involve the $L^4_{t,x}$ norm more than once in order to close the argument, because no gain of a power of T is available in the critical setting. This also makes the $L^4_{t,x}$ norm a *controlling norm* for the evolution.

Sobolev, and Strichartz affect the "game pieces" u, ∇u, $F(u)$, etc. on this board. (The objective of the iteration "game" is to construct a set of assumptions (thus placing game pieces in various spaces with various norm bounds) on the solution, such that it is possible to apply a legal sequence of moves and end up with all the game pieces returning to the same spaces but with better estimates.)

PROOF. (Sketch) We apply Proposition 1.38 with $\mathcal{S} = S^1([-T,T] \times \mathbf{R}^3)$, $\mathcal{N} = N^1([-T,T] \times \mathbf{R}^3)$. By arguing as in Proposition 3.15, we will be done as soon as we show that

$$\| |u|^{p-1}u - |v|^{p-1}v \|_{N^1([-T,T]\times\mathbf{R}^3)} \lesssim_{p,R} T^\alpha \|u - v\|_{S^1_x([-T,T]\times\mathbf{R}^3)}$$

for some $\alpha = \alpha(p) > 0$, whenever the S^1_x norms of u, v are $O(R)$ for some $R > 0$. Let us omit the domain $[-T,T] \times \mathbf{R}^3$ from the notation for brevity. Choosing the admissible exponents $(10, 30/13)$ for the S^1_x norm and $(2, 6)$ for the N^1 norm, it

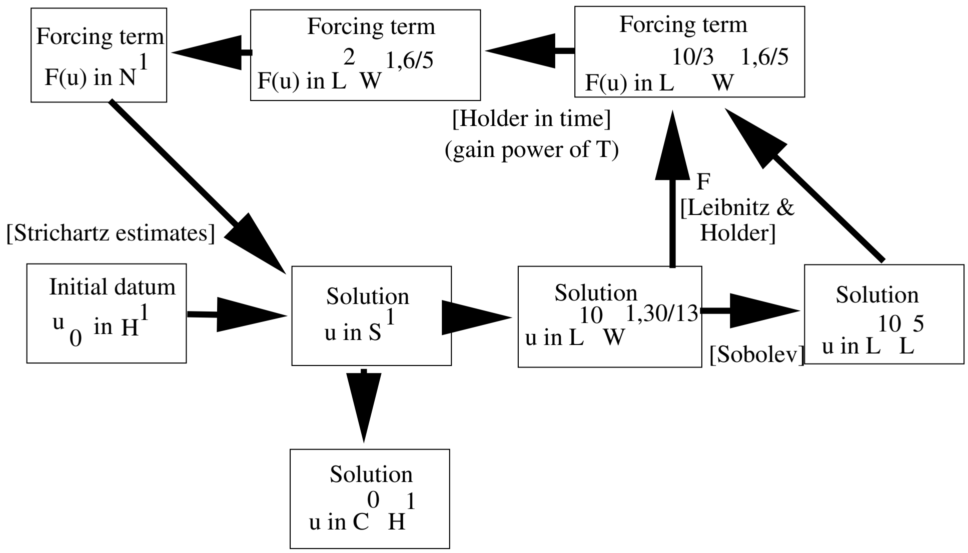

FIGURE 4. An iteration scheme in $H^1_x(\mathbf{R}^3)$ for the three-dimensional cubic NLS $d = 3, p = 3$ (which is H^1_x-subcritical). This is similar to Figure 2; besides the changes in numerology, the main new feature is the appearance of the Leibnitz rule and (non-endpoint) Sobolev embedding to handle the additional derivatives.

suffices to show that

$$\|\nabla^k(|u|^{p-1}u - |v|^{p-1}v)\|_{L^2_t L^{6/5}_x} \lesssim_{p,R} T^\alpha \|u - v\|_{L^{10}_t W^{1,30/13}_x}$$

for $k = 0, 1$. Let us just deal with the higher order case $k = 1$, which is the more difficult of the two[24]. Observe that the gradient of $|u|^{p-1}u$ can be written as $F_1(u)\nabla u + F_2(u)\overline{\nabla u}$, where $F_1, F_2 : \mathbf{C} \to \mathbf{C}$ are functions which grow like $F_j(z) = O_p(|z|^{p-1})$, and which have the Lipschitz-type bound $\nabla F_j(z) = O_p(|z|^{p-2})$. Let us just consider the contribution of the F_1 terms to the above expression, so we need to show

$$\|F_1(u)\nabla u - F_1(v)\nabla v\|_{L^2_t L^{6/5}_x} \lesssim_{p,R} T^\alpha \|u - v\|_{L^{10}_t W^{1,30/13}_x}.$$

From the Lipschitz bound (and the hypothesis $p \geq 2$) we have

$$F_1(u)\nabla u - F_1(v)\nabla v = O_p((|u|^{p-1} + |v|^{p-1})\nabla(u - v)) + O_p((|u|^{p-2} + |v|^{p-2})(u - v)\nabla u).$$

We estimate the $L^2_t L^{6/5}_x$ norm by the $L^{10/p}_t L^{6/5}_x$ norm, gaining a power of T (here we use the fact that $p < 5$), and use Hölder to estimate

$$\begin{aligned} \|F_1(u)\nabla u - F_1(v)\nabla v\|_{L^2_t L^{6/5}_x} \lesssim T^\alpha (&\|u\|^{p-1}_{L^{10}_t L^{5(p-1)/2}_x} \|\nabla(u - v)\|_{L^{10}_t L^{30/13}_x} \\ &+ \|u\|^{p-2}_{L^{10}_t L^{5(p-1)/2}_x} \|u - v\|_{L^{10}_t L^{5/2p}_x} \|\nabla u\|_{L^{10}_t L^{30/13}_x}). \end{aligned}$$

---

[24] A general principle in the local-in-time theory is that the highest order terms are always the most difficult to estimate, so that once those are dealt with the lower order terms should be treatable by a modification of the argument.

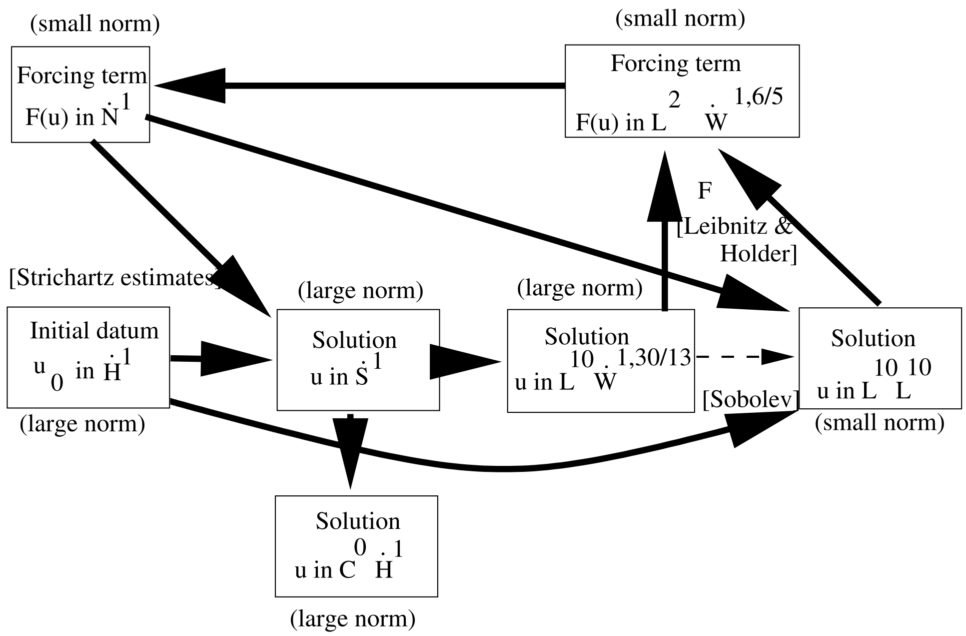

FIGURE 5. An iteration scheme in H^1_x for the three-dimensional quintic NLS $d = 3, p = 5$ (which is H^1_x-critical). This rather tricky scheme is similar to Figure 4 but with homogeneous norms, and with certain norms identified as being small to compensate for the lack of a Hölder in time (cf. Figure 3). It will be important that the smallness in $L^{10}_{t,x}$ is exploited more than once in order to close the iteration properly.

From Sobolev embedding and the hypothesis $2 \leq p < 5$ we have

$$\|u\|_{L^{10}_t L^{5(p-1)/2}_x} \lesssim \|u\|_{L^{10}_t W^{1,30/13}_x}$$

and the claim then follows from the hypothesised bounds on u, v. (See Figure 4 for an illustration of the scheme in the case $p = 3$.) $\square$

There is a version of this argument available in the limit $p = 5$:

PROPOSITION 3.21 ($\dot H^1_x(\mathbf{R}^3)$ critical NLS solutions). *Let $\mu = \pm 1$ and $p = 5$. Then the NLS (3.1) is locally wellposed in $\dot H^1_x(\mathbf{R}^3)$ in the critical sense. More precisely, given any $R > 0$ there exists $\varepsilon_0 = \varepsilon_0(R) > 0$, such that whenever $u_* \in \dot H^1_x(\mathbf{R}^3)$ has norm at most R, and I is a time interval containing 0 such that*

$$\|e^{it\Delta/2} u_*\|_{L^{10}_{t,x}(I \times \mathbf{R}^3)} \leq \varepsilon_0$$

then for any u_0 in the ball $B := \{u_0 \in \dot H^1_x(\mathbf{R}^3) : \|u_0 - u_*\|_{\dot H^1_x(\mathbf{R}^3)} \leq \varepsilon_0\}$ there exists a unique strong $\dot H^1_x$ solution $u \in \dot S^1(I \times \mathbf{R}^3)$ to (3.1), and furthermore the map $u_0 \mapsto u$ is Lipschitz from B to $\dot S^1(I \times \mathbf{R}^3)$. Here $\|u\|_{\dot S^1} := \|\nabla u\|_{S^0}$ is the homogeneous counterpart to the S^1 norm.

We leave the proof of this to the exercises. Note that the $L^{10}_{t,x}$ norm is controlled (via Sobolev embedding) by the $L^{10}_t \dot W^{1,30/13}_x$ norm, which in turn is controlled by the $\dot{S}^1$ norm. From Strichartz estimates we thus conclude that

$$\|e^{it\Delta/2}u_*\|_{L^{10}_{t,x}(\mathbf{R}\times\mathbf{R}^3)} \lesssim \|u_*\|_{\dot{H}^1_x(\mathbf{R}^3)},$$

and thus Proposition 3.21 implies global wellposedness for quintic NLS on $\mathbf{R}^3$ with small energy, and local wellposedness (in the critical sense) for large energy. Again, this proposition and the preceding one should be compared against Principle 3.1.

In the supercritical case $p > 5$, the H^1_x perturbation theory breaks down completely; again, see Section 3.8. However in the defocusing case with $5 < p < 6$ one can still construct global *weak* H^1_x solutions by a weak compactness method; see Exercise 3.56.

Similar wellposedness results exist for the NLW equation, and for the periodic NLS equation; we leave this to the exercises. One can briefly summarise (and oversimplify) the known results for local wellposedness as follows. For NLS, one has local wellposedness in $H^s_x(\mathbf{R}^d)$ as long as $s \geq 0$ and the nonlinearity[25] is H^s_x-subcritical or H^s_x-critical, though in the latter case the wellposedness is in the critical sense (the time of existence depends on the profile of the datum rather than the norm, but for small norm one has global existence). See [**CWeis2**], [**Caz2**]. For NLW, one has a similar result but with an additional constraint[26] $s \geq s_l$, where s_l is the Lorentz regularity $s_l := \frac{d+1}{4} - \frac{1}{p-1}$; this constraint is only relevant in the $H^{1/2}_x$-subcritical cases $p < 1 + \frac{4}{d-1}$; see [**LSog**], [**Sog**]. For periodic NLS, the situation is much less well understood, due to the lack of sharp Strichartz estimates in this setting; see [**Bou**]. (The local theory for periodic NLW is essentially identical to non-periodic NLW; see Exercise 3.24.)

Exercise 3.14. Prove (3.24). (Hint: use the Leibnitz rule, Hölder, Sobolev, and Gagliardo-Nirenberg, controlling the lower order terms before moving on to the higher ones. A Littlewood-Paley approach is possible but somewhat lengthy because of the need to continually commute the Littlewood-Paley operators with weights such as $\langle x\rangle^k$.)

Exercise 3.15. Using the Fourier transform, show that the solution (3.15) to the pseudoconformal focusing NLS blows up in H^s_x for any $s > 0$ as $t \to 0$, but stays bounded in L^2_x, and even goes to zero in H^s_x for $s < 0$. (This reflects the fact that this equation is locally wellposed in the subcritical sense in H^s_x for $s > 0$, is locally wellposed in the critical sense in L^2_x, and is illposed in H^s_x for $s < 0$.) Using this, show that when $s < 0$, one no longer has uniqueness for weak H^s_x solutions, and that Exercise 3.10 also breaks down for weak H^s_x solutions.

Exercise 3.16. Let $u \in \dot{S}^1(I \times \mathbf{R}^3)$ be an H^1_x-wellposed solution to quintic NLS (so $p = 5$ and $d = 3$), and suppose that $u(t_0) \in H^k(\mathbf{R}^3)$ for some $t_0 \in I$ and some integer $k \geq 0$. Show that $u(t) \in H^k(\mathbf{R}^3)$ for all $t \in I$, and in fact

$$\|u(t_0)\|_{H^k(\mathbf{R}^3)} \lesssim_{\|u\|_{\dot{S}^1(I\times\mathbf{R}^3)}} \|u(t_0)\|_{H^k(\mathbf{R}^3)}.$$

(Hint: Let $M := \|u\|_{\dot{S}^1(I\times\mathbf{R}^3)}$. Subdivide I into time intervals I_j where the $L^{10}_{t,x}$ norm on I_j is small compared with M. Then use Strichartz estimates and continuity arguments to establish S^k control on each I_j.)

[25] There is also a technical smoothness condition required on the nonlinearity in the non-algebraic case; see [**CWeis2**], [**Caz2**].

[26] Again, we need a smoothness condition on the nonlinearity. Also this result is not fully established in high dimension $n \geq 4$ when s is very close to zero, for technical reasons; see [**Tao**].

EXERCISE 3.17 (Unconditional uniqueness). Let $u, v \in C_t^0 \dot{H}_x^1(I \times \mathbf{R}^3)$ be strong H_x^1 solutions to quintic NLS (so $p = 5$ and $d = 3$) with $u(t_0) = v(t_0)$ for some $t_0 \in I$. Show that $u = v$. (Hint: Let J be a small time interval containing t_0, and use Strichartz estimates to control $\|u - v\|_{L_t^2 L_x^6(J \times \mathbf{R}^3)}$ in terms of itself and $\|u - v\|_{L_t^\infty L_x^6(J \times \mathbf{R}^3)}$. Then use the continuity of $u - v$ in $\dot{H}_x^1$ and hence in L_x^6 to close the argument. To extend from J back to I, use the continuity method.) In particular, this shows that the H_x^1-wellposed solution given by Proposition 3.21 is the only strong H_x^1 solution. One in fact has unconditional uniqueness for strong H_x^1 solutions for all H_x^1-critical and H_x^1-subcritical equations; see [**Kat9**], [**Caz2**], [**TV**].

EXERCISE 3.18 (L_x^2-critical wellposedness). Prove Theorem 3.17. (Hint: there are now two norms one wishes to place the solution u in: the S^0 norm, and the $L_{t,x}^{2(n+2)/n}$ norm. The solution will be small in the latter norm but can be large in the former norm. To account for this, one either has to apply Proposition 1.38 with an artificial norm such as

$$\|u\|_{\mathcal{S}} := \delta \|u\|_{S^0} + \|u\|_{L_{t,x}^{2(n+2)/n}}$$

for some small δ, or else use work with the iteration scheme directly and establish bounds on all the various norms at each stage of the iteration. See also Figure 3.)

EXERCISE 3.19 (H_x^1-critical wellposedness). Prove Proposition 3.21. (You may find Figure 5 to be helpful.)

EXERCISE 3.20. Show that the cubic NLS on the circle $\mathbf{T}$ is locally wellposed in $L_x^2(\mathbf{T})$ in the subcritical sense. (Hint: use Exercise 2.74.) Also, show persistence of regularity, or more precisely if the initial datum lies in $H_x^k(\mathbf{T})$ for some positive integer k, then the local $L_x^2(\mathbf{T})$ solution constructed by the iteration method is in fact a strong $H_x^k(\mathbf{T})$ solution.

EXERCISE 3.21 (Classical local wellposedness of NLW). Show that an algebraic NLW is unconditionally locally wellposed in $H_x^s \times H_x^{s-1}$ for $s > d/2$ in the subcritical sense, thus for each $R > 0$ there exists a $T = T(k, d, p, R) > 0$ such that for all initial data (u_0, u_1) in the ball $B_R := \{(u_0, u_1) \in H_x^s(\mathbf{R}^d) \times H_x^{s-1}(\mathbf{R}^d) : \|u_0\|_{H_x^s(\mathbf{R}^d)} + \|u_1\|_{H_x^s(\mathbf{R}^d)} < R\}$ there exists a unique classical solution $u \in C_t^0 H_x^s([-T, T] \times \mathbf{R}^d) \cap C_t^1 H^{s-1}([-T, T] \times \mathbf{R}^d)$ to (3.1). Furthermore the map $(u_0, u_1) \mapsto u$ is Lipschitz continuous. (Hint: adapt the proof of Proposition 3.8, and use (2.29).) Show also that the solution can be continued in time as long as u stays bounded in spacetime.

EXERCISE 3.22 ($H_x^1(\mathbf{R}^3)$ subcritical NLW solutions). Let $\mu = \pm 1$ and $2 \leq p < 5$. Show that the NLW (3.2) is locally wellposed in $H_x^1(\mathbf{R}^3) \times L_x^2(\mathbf{R}^3)$ in the subcritical sense. (Hint: there are many schemes available. The simple scheme Figure 6, that does not use Strichartz estimates and which dates back to [**Jor**], only works up to $p \leq 3$; for higher p one needs to use spaces that involve some integration in time. You may also wish to review Figure 2 from Chapter 2, and peek at Figure 4 from Chapter 5.) For the critical case $p = 5$, see Exercise 5.1.

EXERCISE 3.23. Let $d \geq 3$, $\mu = \pm 1$, and let $p := 1 + \frac{4}{d-2}$ be the $\dot{H}_x^1$-critical power. Show that for any $u_0 \in H_x^1(\mathbf{R}^d)$ with sufficiently small norm, there exists a unique global solution $u \in S^1(\mathbf{R} \times \mathbf{R}^d)$ to the NLS (3.1) with the specified initial

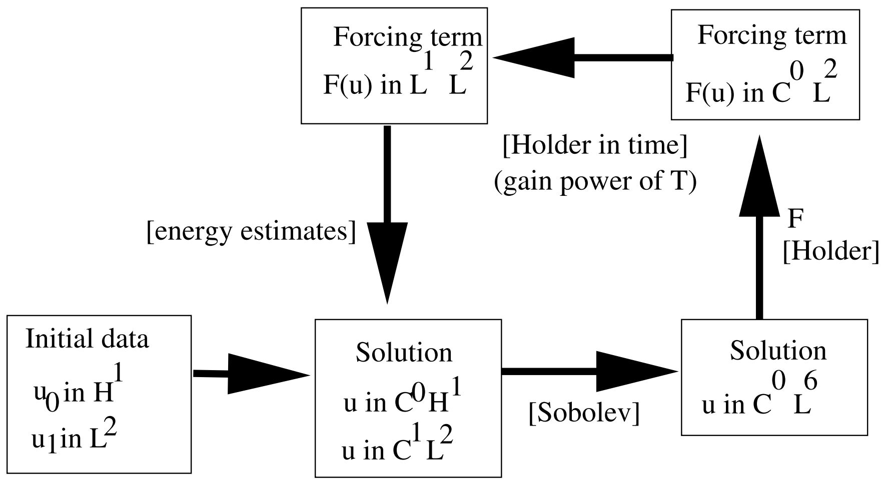

FIGURE 6. A simple iteration scheme in $H^1_x \times L^2_x$, based on the energy estimate, for the three-dimensional cubic NLW $d = 3, p = 3$. For higher powers p, Strichartz estimates and spaces are needed.

datum. (Hint: for $3 \leq d \leq 6$, one can modify the proof of Proposition 3.21 to accomodate the inhomogeneous Sobolev and Strichartz norms. For $d > 6$, we have $p < 2$ and there is now a difficulty obtaining a contraction mapping. However, one can still construct iterates, show that they are bounded in $S^1(\mathbf{R}\times\mathbf{R}^d)$, and converge locally in time in $S^0(I \times \mathbf{R}^d)$. A variant of this argument then gives uniqueness, at least locally in time, which can then be extended to be global in time by continuity arguments. See [**CWeis2**], [**TV**] for a more thorough treatment of this equation.)

EXERCISE 3.24. Suppose that an NLW on $\mathbf{R}^d$ is known to be locally wellposed in $H^s_x \times H^{s-1}_x$ in the subcritical sense for some $s \geq 0$. Assume also that one has the finite speed of propagation result, Proposition 3.3, for $H^s_x \times H^{s-1}_x$-wellposed solutions. (In practice, this hypothesis can either be deduced from the corresponding result for classical solutions by a limiting argument, or else by direct inspection of the iterates used to construct the wellposed solution.) Show that the corresponding periodic NLW on $\mathbf{T}^d$ is also locally wellposed in $H^s_x \times H^{s-1}_x$. (You may find Exercise A.18 to be useful.) This type of descent argument does not always work in the nonperiodic setting, especially for large times; indeed, it is quite possible for the behaviour of an equation for large, localized data to be better in higher dimensions than in lower ones, due to the increased dispersion available.

EXERCISE 3.25 (Analytic wellposedness). Consider an algebraic NLS, and let $s > d/2$ and $R > 0$. By the H^s_x version of Proposition 3.8, we know that there exists $T = T_{d,p,s} > 0$ such that every $u_0 \in H^s_x(\mathbf{R}^d)$ with norm at most R extends to a strong H^s_x solution $u \in C^0_t H^s_x([0,T] \times \mathbf{R}^d)$. Show that if T is small enough, the map $u_0 \mapsto u$ is in fact an real analytic map, thus there is a power series expansion $u = \sum_{k=0}^\infty \mathcal{N}_k(u_0, \ldots, u_0)$ which converges absolutely in $C^0_t H^s_x([0,T] \times \mathbf{R}^d)$, where for each k, $\mathcal{N}_k$ is a k-multilinear operator from $H^s_x(\mathbf{R}^d)$ to $C^0_t H^s_x([0,T]\times\mathbf{R}^d)$. (Hint: you can adapt the proof of the Cauchy-Kowalevski theorem, see Exercise 1.1.) One consequence of analyticity is that the solution map is also infinitely differentiable from H^s_x to $C^0_t H^s_x$.

3.4. Conservation laws and global existence

The journey of a thousand miles begins with one step. (Lao Tzu)

The wellposedness theory of the previous section allowed us to use iterative methods to construct local-in-time solutions to NLS and NLW given suitable regularity assumptions on the initial data; if the datum was also small in a critical norm (e.g. small in L^2_x norm for the L^2_x-critical NLS, or small in $\dot H^1_x$ norm for the H^1_x-critical NLS) then one obtained a global solution also. These methods were perturbative in nature (using the Duhamel formula to approximate the nonlinear evolution by the linear evolution) and thus do not work directly for large data and long times (since one expects the evolution to be genuinely nonlinear in this case). However, in these cases one can use non-perturbative tools to gain enough control on the equation to prevent the solution from blowing up. In this section we describe the most important tool for doing this, namely the conservation laws.

As before, we begin by discussing the nonlinear Schrödinger equation (3.1). In this case we have two independent conservation laws, namely energy conservation and mass/momentum conservation[27]. The latter can be most effectively described by *pseudo-stress-energy tensor* $\mathrm{T}_{\alpha\beta}$, defined for $\alpha, \beta = 0, 1, \ldots, d$ by by

$$\begin{aligned} \mathrm{T}_{00} &= |u|^2 \\ \mathrm{T}_{0j} = \mathrm{T}_{j0} &= \mathrm{Im}(\overline{u}\partial_{x_j} u) \\ \mathrm{T}_{jk} &= \mathrm{Re}(\partial_{x_j} u \overline{\partial_{x_k} u}) - \frac{1}{4}\delta_{jk}\Delta(|u|^2) + \frac{(p-1)\mu}{p+1}\delta_{jk}|u|^{p+1}. \end{aligned}$$

If the solution is sufficiently smooth, one easily verifies the local conservation laws (2.35); see Exercise 3.26. In particular, for smooth decaying solutions, this leads to conservation of total mass $M[u(t)]$, defined by

$$M[u(t)] := \int_{\mathbf{R}^d} \mathrm{T}_{00}(t,x)\ dx = \int_{\mathbf{R}^d} |u(t,x)|^2\ dx \tag{3.29}$$

and total momentum $\vec{p}[u(t)] = (p_1[u(t)], \ldots, p_d[u(t)])$, defined by

$$p_j[u(t)] := \int_{\mathbf{R}^d} \mathrm{T}_{0j}(t,x)\ dx = \int_{\mathbf{R}^d} \mathrm{Im}(\overline{u}\partial_{x_j} u)\ dx;$$

again, see Exercise 3.26. As in Section 2.4, the pseudo-stress-energy tensor also yields many other important conservation laws and montonicity formulae; we will develop some of these later in this chapter.

The above conservation laws are initially only justified for smooth decaying solutions. However, if the conservation law is controlled by an H^s_x norm, and one has a satisfactory wellposedness theory at H^s_x, then there is a standard limiting argument that allows one to extend these conservation laws to these H^s_x-well posed solutions. We illustrate this principle with the example of the one-dimensional cubic NLS:

Proposition 3.22 (Conservation law for a wellposed solution). *Let $d = 1$, $p = 3$, $\mu = \pm 1$. Let $u \in S^0(I \times \mathbf{R}^d) \subseteq C^0_t L^2_x(I \times \mathbf{R}^d)$ be a strong solution to* (3.1) *defined on some time interval I. Let the total mass $M[u(t)]$ be defined by* (3.29)*. Then $M[u(t)] = M[u(t_0)]$ for all $t_0, t \in I$.*

[27] In the special case of the one-dimensional cubic NLS ($d = 1, p = 3$), it turns out that the equation is completely integrable and there are in fact infinitely many conservation laws, see Exercise 3.36.

PROOF. Since I is the increasing union of open intervals we may assume without loss of generality that I is open. Fix $t_0 \in I$. Since $u \in C_t^0 L_x^2(I \times \mathbf{R})$, the set of times $\{t \in I : M[u(t)] = M[u(t_0)]\}$ is closed in I, and it clearly contains t_0, so it suffices by the continuity method to show that this set is open. In other words, it suffices to show that for any $t_1 \in I$, that we have $M[u(t)] = M[u(t_1)]$ for all times t in a neighbourhood of t_1.

By time translation invariance we can take $t_1 = 0$. Set $u_0 := u(0)$. We can approximate u_0 as the limit in $L_x^2(\mathbf{R}^d)$ of a sequence $u_0^{(n)} \in H_x^{100}(\mathbf{R}^d)$ (say) of smooth initial data, which will be uniformly bounded in $L_x^2(\mathbf{R}^d)$. Applying Proposition 3.15, we can obtain a time interval $[-T, T] \subset I$ independent of n, and strong solutions $u^{(n)} \in S^0([-T,T] \times \mathbf{R}^d)$ to (3.1) with initial data $u^{(n)}(0) = u_0^{(n)}$. Since $u_0^{(n)}$ converges to u_0 in L_x^2, the uniqueness and continuity conclusions of Proposition 3.15 guarantee that $u^{(n)}$ will converge to u in $S^0([-T,T] \times \mathbf{R}^d)$. Next, since the $u^{(n)}$ are uniformly bounded in $S^0([-T,T] \times \mathbf{R}^d)$, they are also uniformly bounded in $L_t^4 L_x^\infty([-T,T] \times \mathbf{R}^d)$ and hence in $L_t^2 L_x^\infty([-T,T] \times \mathbf{R}^d)$. We may thus apply Proposition 3.11 (and the remarks following that proposition) and conclude that $u^{(n)} \in C_t^0 H_x^s([-T,T] \times \mathbf{R}^d)$ for any $s > d/2$. From this, Sobolev embedding, and the equation (3.1) it is easy to see that $u^{(n)}$ is smooth on $[-T,T] \times \mathbf{R}^d$. This is enough regularity for us to apply the classical mass conservation law in Exercise 3.26 and conclude that

$$\int_{\mathbf{R}^d} |u^{(n)}(t,x)|^2 \, dx = \int_{\mathbf{R}^d} |u^{(n)}(0,x)|^2 \, dx$$

for all $t \in [-T,T]$. Since $u^{(n)}$ converges to u in $S^0([-T,T] \times \mathbf{R}^d)$, it also converges in $C_t^0 L_x^2([-T,T] \times \mathbf{R}^d)$, and hence on taking limits as $n \to \infty$ we have

$$\int_{\mathbf{R}^d} |u(t,x)|^2 \, dx = \int_{\mathbf{R}^d} |u(0,x)|^2 \, dx$$

and hence $M[u(t)] = M[u(0)]$ for all t in a neighbourhood of t, as desired. □

Note that the full power of the wellposedness theory was used here; not only the existence aspect of the theory, but also the uniqueness, persistence of regularity, and continuous dependence on the data. This basic argument - obtaining conservation laws for rough solutions by approximating them by smooth solutions - is quite general and extends to many other equations. There is an additional twist however in the case when the nonlinearity is not algebraic (i.e. p is not an odd integer), because it is often not possible in such cases to obtain an approximating solution that is sufficiently smooth that one can justify the conservation law classically. In such cases one must not only regularise the initial data, but also regularise the *equation* as well; this requires a further aspect of the wellposedness theory, namely the *stability* theory of the equation, which we will address in Section 3.7.

A conservation law can often, but not always, be combined with a local wellposedness result to obtain a global wellposedness result. Let us illustrate this with a simple example:

PROPOSITION 3.23. [**Tsu**] *Let $d = 1$, $p = 3$, $\mu = \pm 1$, and $t_0 \in \mathbf{R}$. Let $u_0 \in L_x^2(\mathbf{R}^d)$, and let I be any bounded time interval containing t_0. Then there is a unique strong solution $u \in S^0(I \times \mathbf{R}^d) \subseteq C_t^0 L_x^2(I \times \mathbf{R}^d)$ to (3.1) defined on some time interval I. Furthermore, the map $u_0 \mapsto u$ is a continuous map from L_x^2 to $S^0(I \times \mathbf{R}^d)$.*

In particular, this proposition gives a global strong solution $u \in C_t^0 L_x^2(\mathbf{R} \times \mathbf{R}^d)$ to (3.1). However this solution is only in the S^0 space locally in time. (Indeed, in the focusing case $\mu = -1$, soliton solutions to this equation exist which do not lie globally in S^0.)

Proof. we first use time translation invariance and time reversal symmetry to reduce to the case when $t_0 = 0$ and $I = [0, T]$ for some $T > 0$. (Note that by Exercise 3.11 one can easily glue a strong solution on an interval such as $[0, T]$ to a strong solution to an interval such as $[-T, 0]$, and stay in S^0.)

We give two proofs of this result; they are equivalent, but offer slightly different perspectives as to how a conservation law extends a local existence result to a global existence result. For the first proof, we divide the long time interval $[0, T]$ into shorter time steps, where on each smaller interval the perturbative theory gives a local solution. More precisely, let $M[u(0)] := \|u_0\|^2_{L^2_x(\mathbf{R}^d)}$ denote the initial mass, and observe from Proposition 3.15 that there exists a time $\tau = \tau(M[u(0)]) > 0$ such that the equation (3.1) will have a local strong solution in $S^0([t_0, t_0 + \tau] \times \mathbf{R}^d)$ whenever the initial datum $u(t_0)$ has mass less than or equal to $M[u(0)]$. We now split the time interval $[0, T]$ as a finite union of intervals $[t_n, t_{n+1}]$, where each interval has length less than or equal to τ, and $t_0 = 0$. By applying Proposition 3.15 followed by Proposition 3.23 (and Exercise 3.11), an easy induction shows that for every n we can construct a strong solution u to (3.1) in $S^0([0, t_n] \times \mathbf{R}^d)$, and thus we eventually obtain a strong solution to $S^0([0, T] \times \mathbf{R}^d)$; see Figure 7. Uniqueness follows from Proposition 3.15 and a continuity argument similar to that used to prove Exercise 3.13. The continuous dependence follows by concatenating the continuous dependence results on each of the intervals $[t_n, t_{n+1}]$, using the fact that the $S^0([t_n, t_{n+1}] \times \mathbf{R}^d)$ norm of u controls the L^2_x norm of $u(t_{n+1})$, and using the fact that the composition of continuous maps is continuous.

The second proof proceeds by contradiction; it is quicker but is more indirect (and does not give the continuous dependence as easily). We sketch it as follows. Just as the Picard existence and uniqueness theorems imply a blowup criterion for ODE (Proposition 1.17), the existence and uniqueness theory in Proposition 3.15 gives a blowup criterion for L^2_x strong solutions to (3.1), namely that these solutions will exist globally unless the L^2_x norm goes to infinity in finite time. However, Proposition 3.22 clearly implies that the L^2_x norm of a strong solution stays bounded. Thus blowup cannot occur, and one must instead have global existence. $\square$

One can combine this global existence result with persistence of regularity theory (e.g. Proposition 3.11) to show that the global solution constructed in Proposition 3.23 preserves regularity; see Exercise 3.28. In particular, with a smooth decaying initial datum we have a global smooth solution to the one-dimensional cubic NLS.

Similar arguments give global L^2_x-wellposedness for any L^2_x-subcritical equation. The situation is remarkably different when one considers the *two-dimensional* cubic NLS ($d = 2$, $p = 3$, $\mu = \pm 1$). The key difference is that whereas the one-dimensional cubic NLS was L^2_x-subcritical, the two-dimensional cubic NLS is L^2_x-critical. This is reflected in the local wellposedness theory for this equation, given by Proposition 3.17. If the initial datum has a sufficiently small L^2_x norm, then this proposition already gives a global existence result without any need for a conseration law. However, when the L^2_x norm is large, one cannot simply combine the conservation law

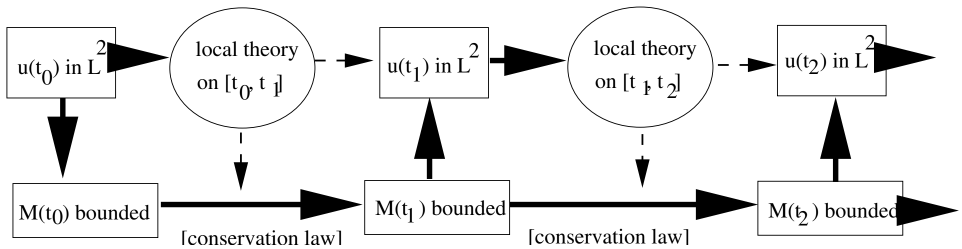

FIGURE 7. The first proof of Proposition 3.23. The nonlinear iteration in the local theory could potentially increase the L^2_x norm as one advances from one time step t_n to the next t_{n+1}, thus leading to a collapse of the lifespan $t_{n+1} - t_n$ of that theory; however if one uses instead the conservation law to control the L^2_x norm then no collapse occurs (cf. the global wellposedness of the ODE in Figure 9 from Chapter 1). Indeed the local theory plays a mostly qualitative role in the global argument, justifying the local existence of the solution as well as the conservation law, but not providing the key quantitative bounds.

with the local existence theory to obtain a global existence theory; the problem is that the time of existence given by the local wellposedness theory does not depend only on the L^2_x norm on the datum, but also on the profile of the datum itself (and more specifically on the spacetime behaviour of the free evolution of the datum). Because of this, the conservation law is insufficient by itself to make either of the arguments used in the proof of Proposition 3.23 extend to this case; iterating the local wellposedness theorem can lead to a shrinking interval of existence, which can lead to blowup in finite time. Indeed, in the focusing case $\mu = -1$, the explicit blowup solution given in (3.15) shows that even for smooth $L^2_x(\mathbf{R}^d)$ initial data one can have finite time blowup for this equation. (Note that the classical uniqueness theory shows that this failure of global existence cannot be avoided by strengthening the notion of solution.) In the defocusing case $\mu = 1$, global existence (and wellposedness) for $L^2_x(\mathbf{R}^d)$ initial data is suspected to be true for the two-dimensional cubic NLS, but this is not known, even for radially symmetric initial data, and is considered a major open problem in the field; a similar open question exists for any other L^2_x-critical defocusing NLS. However, the situation improves when the initial data is assumed to be in the energy class $H^1_x(\mathbf{R}^d)$, rather than merely in $L^2_x(\mathbf{R}^d)$, because a new conservation law becomes available, namely energy conservation[28] For a general NLS, the total energy $E[u(t)]$ takes the form

$$E[u(t)] := \int_{\mathbf{R}^d} \frac{1}{2}|\nabla u(t,x)|^2 + \frac{2\mu}{p+1}|u(t,x)|^{p+1}\, dx. \tag{3.30}$$

We refer to the first term $\frac{1}{2}\int_{\mathbf{R}^d} |\nabla u(t,x)|^2$ as the *kinetic energy* or *linear energy*, and the second term $\frac{2\mu}{p+1}|u(t,x)|^{p+1}\, dx$ as the *potential energy* or *nonlinear energy*. Note that in the defocusing case $\mu = +1$ these two terms have the same sign,

[28] In principle, momentum conservation should become useful also once one reaches the regularity $H^{1/2}_x$ or higher, thanks to Lemma A.10. However, because the momentum is a vector rather than a positive quantity, the momentum is in practice not sufficiently coercive to obtain any useful control of the solution. See however the Morawetz arguments in the next section.

whereas in the focusing case $\mu = -1$ they have opposite signs. In practice, this means that the energy conservation law is more coercive (gives more control on the solution) in the defocusing case than in the focusing case.

The following heuristic principle, related to Principle 1.37 and Principle 3.1, can be helpful in predicting behaviour of equations in the energy class:

PRINCIPLE 3.24 (Energy principle). *Suppose a solution has finite energy. If the linear energy dominates the nonlinear energy, we expect linear behaviour; if the nonlinear energy dominates the linear energy, we expect nonlinear behaviour.*

This heuristic has very little rigorous justification to back it up, yet is surprisingly accurate in many cases, as we shall see in several places in this text.

For sufficiently classical solutions one can justify conservation of energy $E[u(t)]$ by integration by parts; see Exercise 3.31. In the energy subcritical cases $s_c \leq 1$, the energy functional $u \mapsto E[u]$ is continuous in $H^1_x(\mathbf{R}^d)$ (Exercise 3.32). Combining this with the local H^1_x wellposedness theory (such as Proposition 3.19 and 3.21) as in Proposition 3.22, one can extend the energy conservation law to all H^1_x-critical and H^1_x-subcritical wellposed solutions. (In fact one has this for strong solutions also, thanks to uniqueness results such as Exercise 3.17. The high-dimensional cases $n > 6$ are a little tricky; see [**Caz2**], [**TV**].)

Let us now return to the two-dimensional cubic NLS ($d = 2, p = 3$), and see what this new conservation law gives us. The focusing and defocusing cases are now rather different (as one can already see from the blowup solution (3.15)). In the defocusing case, it is clear that energy and mass together will control the H^1_x norm of the solution:

$$\|u(t)\|^2_{H^1_x} \lesssim E[u(t)] + M[u(t)].$$

Conversely, the Gagliardo-Nirenberg inequality (Proposition A.3) shows that the energy and mass are controlled by the H^1_x norm:

$$E[u(t)] \lesssim \|u(t)\|^2_{H^1_x}(1 + \|u(t)\|^2_{L^2_x}) \lesssim \|u(t)\|^2_{H^1_x}(1 + \|u(t)\|^2_{H^1_x})$$
$$M[u(t)] = \|u(t)\|^2_{L^2_x} \lesssim \|u(t)\|^2_{H^1_x}.$$

From these bounds and the energy and mass conservation laws we see that for any H^1_x-wellposed solution, the H^1_x norm of the solution $u(t)$ at some later time t is bounded by a quantity depending only on the H^1_x norm of the initial datum. On the other hand, Proposition 3.19 shows that an H^1_x wellposed solution can be continued in time as long as the H^1_x norm does not go to infinity. Combining these two statements we obtain

PROPOSITION 3.25. *The defocusing two-dimensional cubic NLS ($d = 2, p = 3, \mu = +1$) is globally wellposed in H^1_x. Indeed for any $u_0 \in H^1_x$ and any time interval I containing t_0, the Cauchy problem* (3.1) *has an H^1_x-wellposed solution $u \in S^1(I \times \mathbf{R}^2) \subseteq C^0_t H^1_x(I \times \mathbf{R}^2)$.*

The reader should see how the scheme in Figure 7 is modified to accomodate the utilisation of two conservation laws (mass and energy) rather than just one.

REMARK 3.26. This argument is in fact quite general and works for any H^1_x-subcritical defocusing NLS; see for instance [**Caz2**] or Exercise 3.35. One also has global wellposedness in H^1_x for the H^1_x-critical equation but this is significantly more difficult and will be discussed in Chapter 5. This fits well with Principle 3.24, since in the H^1_x-critical equation, Sobolev embedding only barely manages to control the nonlinear energy in terms of the linear energy.

REMARK 3.27. In the focusing case, a problem arises because of the negative sign of the nonlinear component of the energy $E[u(t)]$. This means that while the energy is still controlled by the H^1_x norm, the H^1_x norm is not necessarily controlled by the energy. Indeed, (3.15) shows that global wellposedness fails for some H^1_x data. However this turns out to be the borderline case: for any H^1_x data with mass strictly less than that of the blowup solution in (3.15), the Gagliardo-Nirenberg inequality becomes strong enough again to control the H^1_x norm by the energy and thus regain global wellposedness. See Exercise 3.33.

The above discussion was for the two-dimensional cubic NLS, which was L^2_x-critical. For equations which are L^2_x-subcritical, it turns out that the Gagliardo-Nirenberg inequality is now so strong that the sign μ of the nonlinearity plays essentially no role in the global theory; see Exercise 3.34. For equations which are L^2_x-supercritical but H^1_x-subcritical or H^1_x-critical, the defocusing equation enjoys global existence in H^1_x as discussed above, but blowup can occur for the focusing equation unless a suitable smallness condition is met; see Exercise 3.35 and Section 3.8.

Having discussed the conservation laws for the NLS, we now turn to the NLW (3.2), which we write using the Minkowski metric as

$$\partial^\alpha \partial_\alpha u = \mu |u|^{p-1} u. \tag{3.31}$$

For this equation there is no mass conservation law, and the energy/momentum conservation laws can be unified via the *stress-energy tensor*

$$\mathrm{T}^{\alpha\beta} := \mathrm{Re}(\partial^\alpha u \overline{\partial^\beta u}) - \frac{1}{2} g^{\alpha\beta} \mathrm{Re}(\partial^\gamma u \overline{\partial_\gamma u} + \frac{2\mu}{p+1} |u|^{p+1}) \tag{3.32}$$

(compare with (2.45)). In coordinates,

$$\begin{aligned}
\mathrm{T}^{00} = \mathrm{T}_{00} &= \frac{1}{2}|\partial_t u|^2 + \frac{1}{2}|\nabla u|^2 + \frac{\mu}{p+1}|u|^{p+1} \\
\mathrm{T}^{0j} &= -\mathrm{T}_{j0} = -\mathrm{Re}(\partial_t u \overline{\partial_{x_j} u}) \\
\mathrm{T}^{jk} = \mathrm{T}_{jk} &= \mathrm{Re}(\partial_{x_j} u \overline{\partial_{x_k} u}) - \frac{\delta_{jk}}{2}(|\nabla u|^2 - |\partial_t u|^2 + \frac{2\mu}{p+1}|u|^{p+1}).
\end{aligned}$$

From (3.31) we easily verify the divergence-free property

$$\partial_\alpha \mathrm{T}^{\alpha\beta} = 0 \tag{3.33}$$

or in coordinates

$$\partial_t \mathrm{T}^{00} + \partial_{x_j} \mathrm{T}^{0j} = 0; \quad \partial_t \mathrm{T}^{0j} + \partial_{x_k} \mathrm{T}^{jk} = 0, \tag{3.34}$$

for classical solutions at least. This leads (for classical, decaying solutions) to conservation of the total energy

$$E[u[t]] := \int_{\mathbf{R}^d} \mathrm{T}^{00}(t,x)\, dx = \int_{\mathbf{R}^d} \frac{1}{2}|\partial_t u|^2 + \frac{1}{2}|\nabla u|^2 + \frac{\mu}{p+1}|u|^{p+1}\, dx \tag{3.35}$$

and total momentum[29]

$$p_j(u[t]) := \int_{\mathbf{R}^d} \mathrm{T}^{0j}(t,x)\, dx = -\int_{\mathbf{R}^d} \mathrm{Re}(\partial_t u \overline{\partial_{x_j} u})\, dx.$$

By the limiting arguments as before, these conservation laws can be extended to H^1_x-wellposed solutions, as long as the equation is either H^1_x-subcritical or H^1_x-critical.

For defocusing H^1_x-subcritical equations, the energy conservation can lead to global existence in even for large initial data. Let us illustrate this with the three-dimensional cubic NLW ($d = 3, p = 3, \mu = +1$). From Exercise 3.22 we have a local wellposedness result for initial data in $H^1_x \times L^2_x$ in the subcritical sense, which easily implies a blowup criterion, namely that the $H^1_x \times L^2_x$ wellposed solutions to this equation can be continued in time as long as the quantity $\|u(t)\|_{H^1_x} + \|\partial_t u(t)\|_{L^2_x}$ does not go to infinity in finite time. To bound this quantity, we observe from (3.35) and energy conservation that we can bound the homogeneous component of this quantity easily:

$$\|u(t)\|_{\dot H^1_x} + \|\partial_t u(t)\|_{L^2_x} \lesssim E[u[t]]^{1/2} = E[u[0]]^{1/2}.$$

To control the lower order term $\|u(t)\|_{L^2_x}$ we use the fundamental theorem of calculus and Minkowski's inequality:

$$\begin{aligned}\|u(t)\|_{L^2_x} &\leq \|u(0)\|_{L^2_x} + \left|\int_0^t \|\partial_t u(s)\|_{L^2_x}\, ds\right| \\ &\lesssim \|u(0)\|_{L^2_x} + \left|\int_0^t E[u[s]]^{1/2}\, ds\right| \\ &= \|u(0)\|_{L^2_x} + |t| E[u[0]]^{1/2}.\end{aligned}$$

Thus we can control the $H^1_x \times L^2_x$ norm of $u[t]$ by a quantity that depends mostly on the initial data:

$$\|u(t)\|_{H^1_x} + \|\partial_t u(t)\|_{L^2_x} \lesssim \|u(0)\|_{L^2_x} + (1+|t|) E[u[0]]^{1/2}.$$

In particular, if $u[0] \in H^1_x \times L^2_x$, then $E[u[0]]$ is finite and the quantity $\|u(t)\|_{H^1_x} + \|\partial_t u(t)\|_{L^2_x}$ cannot go to infinity in finite time for any $H^1_x \times L^2_x$-wellposed solution. Comparing this with the blowup criterion[30] we see that we have a global $C^0_t H^1_x \cap C^1_t L^2_x$ strong solution to this equation for any initial data in the energy space $H^1_x \times L^2_x$. Once one has this global existence result in the energy class, we also obtain it for smoother classes; see Exercise 3.38.

The above argument in fact works for any H^1_x-subcritical NLW; see [**GV2**], [**Sog**]. The case of the H^1_x-critical equation is again more delicate (because the lifespan given by the local existence theorem depends on the profile of the data as

[29] Note the presence of the time derivative, which is absent for the NLS momentum. Indeed, while the NLS momentum is naturally associated to the regularity $\dot H^{1/2}_x$, the NLW momentum is associated to the regularity $\dot H^1_x$, and thus is of the same order as the energy. Thus the relationship between momentum and energy in NLW is different from that in NLS, which turns out to be a crucial difference in the critical scattering theory, see Chapter 5 below.

[30] It is instructive to use the other approach to global existence from Proposition 3.23, namely dividing up a long time interval into short ones. Here, because the quantity $\|u(t)\|_{H^1_x} + \|\partial_t u(t)\|_{L^2_x}$ could grow linearly in t, the interval of existence guaranteed by the local theory could decay polynomially in t. However, the length of this interval will not go to zero in finite time, which is all one needs to establish global existence.

well as on the $H^1 \times L^2_x$ norm, and so the blowup criterion is more subtle), and will be treated in Section 5.1; again, this is in accordance with Principle 3.24. For focusing NLW, there are no Gagliardo-Nirenberg tricks available to control the nonlinearity (since there is no mass conservation law), and indeed for such equations large data can lead to blowup no matter what the power is; see Exercise 3.9.

EXERCISE 3.26. Verify (2.35) for the pseudo-stress-energy tensor for $C^3_{t,x,\mathrm{loc}}$ solutions to (3.1). These conservation laws may seem somewhat miraculous, but can be explained using Lagrangian methods; see [**SSul**], as well as Exercises 3.30, 3.40. Conclude that if $u \in C^3_{t,x,\mathrm{loc}}(I \times \mathbf{R}^d)$ is a solution to (3.1) which also lies in $C^0_t H^s_x(I \times \mathbf{R}^d)$ for a sufficiently large s, then we have mass conservation $M[u(t)] = M[u(t_0)]$ and momentum conservation $p_j(t) = p_j(t_0)$ for all $t, t_0 \in I$. (In order to justify the integration by parts, one needs to apply a smooth cutoff in space, and then let the cutoff go to infinity, using the H^s_x control on u to show that the error incurred by the cutoff goes to zero.) These rather restrictive regularity conditions can usually be removed by a limiting argument.

EXERCISE 3.27. Obtain the analogue of Proposition 3.22 but with $d = 2$ instead of $d = 1$. (The challenge here is that the equation is now L^2_x-critical instead of L^2_x-subcritical, and one has to use Proposition 3.17 instead of Proposition 3.15.)

EXERCISE 3.28. Let $p = 3, d = 1, \mu = \pm 1$, and let $u_0 \in H^s_x(\mathbf{R})$ for some $s \geq 0$. Show that the solution u constructed in Proposition 3.23 is a strong H^s_x solution, and we have the bound $\|u(t)\|_{H^s_x} \lesssim \exp(Ct)\|u_0\|_{H^s_x}$, where $C = C(s, \|u_0\|_{L^2_x}) > 0$ depends only on s and the initial mass.

EXERCISE 3.29. Show that the energy (3.30) is formally the Hamiltonian for the NLS (3.1) using the symplectic structure from Exercise 2.47. Also use Noether's theorem to formally connect the mass conservation law to the phase invariance of NLS, and the momentum conservation law to the translation invariance of NLS.

EXERCISE 3.30. Use Exercise 3.2 to link the pseudo-stress-energy tensor and energy conservation law for d-dimensional NLS to the stress-energy tensor for $d+1$-dimensional NLW. (In $d + 1$ dimensions, you may find it convenient to introduce a null frame $(t', x', x_1, \ldots, x_d)$ where $t' := t - x_{d+1}$ and $x' := t + x_{d+1}$, and compute the coordinates of the stress-energy tensor in that frame.)

EXERCISE 3.31. Let $u \in C^3_{t,x}(I \times \mathbf{R}^d)$ be a classical solution to (3.1). Verify the identity

$$\partial_t\Big(\frac{1}{2}|\nabla u(t,x)|^2 + \frac{2\mu}{p+1}|u(t,x)|^{p+1}\Big) =$$
$$\partial_j\Big(\frac{1}{2}\mathrm{Im}(\overline{\partial_{jk} u(t,x)}\partial_k u(t,x)) + \mu|u(t,x)|^{p-1}\mathrm{Im}(\overline{u(t,x)}\partial_j u(t,x))\Big).$$

If u is also in $C^0_t H^{k,k}_x(I \times \mathbf{R}^d)$ for some sufficiently large k, deduce the energy conservation law $E[u(t_1)] = E[u(t_0)]$ for all $t_0, t_1 \in I$ (by arguing as in Exercise 3.26).

EXERCISE 3.32. If $s_c \leq 1$, show that the energy functional $u \mapsto E[u]$ is well-defined and continuous on the space $H^1_x(\mathbf{R}^d)$. (Hint: use Sobolev embedding and an estimate similar to (3.27).) When $s_c > 1$, show that the energy is not always finite for $H^1_x(\mathbf{R}^d)$ data.

EXERCISE 3.33. [**Wei**] Let $u_0 \in H^1_x(\mathbf{R}^2)$ have mass strictly less than the mass of the blowup solution in (3.15). Show that there is a global strong H^1_x solution to the cubic defocusing two-dimensional NLS (i.e. (3.1) with $d = 2, p = 3, \mu = -1$) with initial datum u_0. (Hint: you will need the relationship between Q and $W_{\max}$ given by Lemma B.1 in order to control the H^1_x norm by the energy.) Note that this is consistent with Principle 3.24. There has been much recent analysis of the case when the mass is exactly equal to, or slightly higher than, the mass of the blowup solution (3.15): see [**Mer2**], [**Mer3**], [**MR**], [**MR2**], [**MR3**].

EXERCISE 3.34. Let u be an H^1_x-wellposed solution to the one-dimensional cubic NLS ($d = 1, p = 3, \mu = \pm 1$) with initial datum u_0; this is a global solution by Proposition 3.23 and persistence of regularity. Establish the bound $\|u(t)\|_{H^1_x(\mathbf{R})} \lesssim_{\|u_0\|_{H^1_x}(\mathbf{R})} 1$ for all times t, regardless of whether the equation is focusing or defocusing. (The point here is that in the L^2_x-subcritical equations, the Gagliardo-Nirenberg inequality allows one to control the nonlinear component of the energy by a *fractional* power of the linear component of the energy, times a factor depending only on the conserved mass; cf. Exercise 1.22 and Principle 3.24.)

EXERCISE 3.35. Consider the defocusing three-dimensional cubic NLS ($d = 3, p = 3$). Show that one has global H^1_x-wellposed solutions if the initial datum u_0 is sufficiently small in $H^1_x(\mathbf{R}^3)$ norm, and in the defocusing case one has global H^1_x-wellposedness for arbitrarily large $H^1_x(\mathbf{R}^3)$ initial data. (Again, one can rely primarily on the Gagliardo-Nirenberg inequality and the conservation laws. An alternate approach is to develop a small data global existence theory at the critical regularity $\dot H^{1/2}_x(\mathbf{R}^3)$ by perturbative arguments, and then use persistence of regularity to move from $\dot H^{1/2}_x$ to H^1_x.)

EXERCISE 3.36. Consider the one-dimensional cubic NLS ($d = 1, p = 3$). It turns out that there is a conserved quantity (at least for classical solutions to NLS) of the form

$$E_2(u) := \int_{\mathbf{R}} |\partial_{xx} u|^2 + c_1\mu|\partial_x u|^2|u|^2 + c_2\mu \mathrm{Re}((\overline{u}\partial_x u)^2)) + c_3\mu^2|u|^6\ dx$$

for certain absolute constants c_1, c_2, c_3 whose exact value is unimportant here. (The verification of this conservation law is extremely tedious if done directly, though the machinery in Section 4.2 can expedite some of the algebra; one can also proceed via the Lax pair formulation of this equation.) Assuming this, conclude the bound $\|u(t)\|_{H^2_x(\mathbf{R})} \lesssim_{\|u_0\|_{H^2_x}(\mathbf{R})} 1$ for all times t, at least for classical solutions to (3.1). (The same bound in fact holds for all H^2_x solutions, and one has a similar result with H^2_x replaced by H^k_x for any integer $k \geq 0$.)

EXERCISE 3.37. Show that the 1D cubic periodic NLS (with either sign of nonlinearity) is globally wellposed in $L^2_x(\mathbf{T})$. Also show that if the initial datum is smooth, then the solution is globally classical. (One should of course use the local theory from Exercise 3.20.)

EXERCISE 3.38. Show that for every smooth initial data $u[0]$ to the three-dimensional cubic defocusing NLW ($d = 3, p = 3, \mu = +1$), there is a unique classical solution. (In the text we have already established global wellposedness in $H^1_x \times L^2_x$ for this equation; it is now a matter of applying the persistence of regularity theory (to ensure the solution is smooth) and the finite speed of propagation and uniqueness theory (to localise the initial data to be compactly supported).)

EXERCISE 3.39. Consider a global $H^1_x \times L^2_x$-wellposed solution u to the three-dimensional cubic defocusing NLW ($d = 3, p = 3, \mu = +1$), as constructed in the text. If $(u_0, u_1) \in H^k_x \times H^{k-1}_x$ for some $k \geq 1$, show that the quantity $\|u(t)\|_{H^k_x} + \|u(t)\|_{H^{k-1}_x}$ grows at most polynomially in time t, in fact we have the bound

$$\|u(t)\|_{H^k_x} + \|u(t)\|_{H^{k-1}_x} \lesssim_{\|u_0\|_{H^k_x} + \|u_1\|_{H^{k-1}_x}} (1 + |t|)^{C_k}$$

for some $C_k > 0$. (Hint: induct on k. This result should be compared with the exponential bounds in Exercise 3.28. The difference is that for the wave equation, the energy estimate (2.28) or (2.29) gains one degree of regularity, which is not the case for the Schrödinger equation. However, in many cases it is possible to use additional smoothing estimates and almost conservation laws to recover polynomial growth of Sobolev norms for the Schrödinger equation; see [**Sta**], [**CKSTT8**], and the scattering estimates we give in Section 3.6 can eliminate this growth altogether).

EXERCISE 3.40. Show that the NLW (3.2) is the (formal) Euler-Lagrange equation for the Lagrangian $S(u, g) = \int_{\mathbf{R}^{1+d}} L(u, g)\ dg$, where $L(u, g) := g^{\alpha\beta} \partial_\alpha u \partial_\beta u + \frac{2\mu}{p+1}|u|^{p+1}$. Conclude that the stress-energy tensor given here coincides with the one constructed in Exercise 2.60.

EXERCISE 3.41 (Positivity of stress-energy tensor). Let u be a classical solution to a defocusing NLW, and let $\mathrm{T}^{\alpha\beta}$ be the associated stress energy tensor. Let x^α, y^α be forward timelike or forwardlike vectors, thus $x^0, y^0 > 0$ and $x^\alpha x_\alpha, y^\alpha y_\alpha \leq 0$. Show that $\mathrm{T}_{\alpha\beta} x^\alpha y^\beta \geq 0$. In particular we have the positivity property (2.47). (Hint: first establish this when x^α is the standard timelike vector e^0, then use Lorentz invariance to handle the case when x is a general timelike vector, then use limiting arguments to handle the lightlike case.) This positivity is related to the finite speed of propagation property for NLW but is not identical; indeed, in the focusing case $\mu = -1$, the positivity fails but one still has finite speed of propagation.

EXERCISE 3.42. Consider a H^1_x-subcritical focusing NLS. Show that the ground state $e^{it\tau}Q(x)$ has positive energy in the L^2_x-supercritical case $p > 1 + \frac{4}{d}$, zero energy in the L^2_x-critical case $p = 1 + \frac{4}{d}$, and negative energy in the L^2_x-subcritical case $p < 1 + \frac{4}{d}$. Similarly for all translates, rescalings, and Galilean transforms of the ground state. (Hint: Use Exercise B.3.)

EXERCISE 3.43 (Orbital stability of NLS). [**Wei3**] Consider a L^2_x-subcritical focusing NLS. Define the *ground state cylinder* Σ as in Exercise B.14. Show that if the initial datum $u_0 \in H^1_x(\mathbf{R}^d)$ is sufficiently close to Σ in $H^1_x(\mathbf{R}^d)$, then the global H^1_x solution u to the Cauchy problem is such that $\mathrm{dist}_{H^1_x}(u(t), \Sigma) \sim \mathrm{dist}_{H^1_x}(u_0, \Sigma)$ for all $t \in \mathbf{R}$. (Hint: first rescale so that u_0 has the same mass as Σ, then use Exercise B.14.) An earlier result of this type appears in [**CL**]. The theory of orbital stability of ground states for much more general equations has been studied extensively; see for instance [**GSS**], [**GSS2**] for a systematic approach.

3.5. Decay estimates

Things fall apart; the centre cannot hold. (WB Yeats, "The second coming")

The conservation laws of the preceding section can give global bounds on a solution $u(t)$ to NLS or NLW that are either uniform in time, or grow at some controlled rate (polynomial or exponential in time). We have already seen that such bounds can be sufficient to obtain global existence of the solution. However, one is not just interested in whether solutions exist globally in time; one is also interested in the asymptotic behaviour of these solutions as $t \to \pm\infty$. The conservation laws show that these solutions stay bounded in certain norms, but this still leaves a lot of possibilities for the asymptotic development. For instance, consider the following two rather different (and informally described) modes of behaviour:

- (Linear-type behaviour) The solution u behaves like a solution to the linear equation; thus the nonlinear effects become asymptotically negligible. In particular, we expect the solution to obey the same type of dispersive and Strichartz estimates as the linear equation, thus for instance we may expect the L^∞_x or other L^p_x norms of the solution to go to zero as $t \to \pm\infty$. We also expect Sobolev norms such as $\|u(t)\|_{H^s_x}$ to stabilise as $t \to \pm\infty$, as opposed to growing polynomially or exponentially in time.
- (Soliton-type behaviour) The solution refuses to disperse, and for every time t the solution has a significant presence at some location $x(t)$ depending on t, for instance the local mass $\int_{|x-x(t)|\leq R} |u(t,x)|^2\, dx$ might be bounded away from zero for some fixed R. In particular the L^∞_x or L^p_x norms of $u(t)$ will *not* go to zero as $t \to \pm\infty$. This is the type of behaviour exhibited by soliton solutions such as (3.7) (possibly after applying some symmetries such as Galilean or Lorentz invariance). One can also consider more complex behaviour when for each fixed time t, the solution has significant presence at multiple points $x_1(t), \ldots, x_k(t)$; this is the case for *multi-soliton* solutions, which are essentially a nonlinear superposition of several separated solitons.

There is some evidence (both theoretical and numerical) that for "most"[31] global solutions to an NLS or NLW, that the asymptotic behaviour eventually decouples into the above two extremes: most solutions should eventually resolve into a "localised" component which behaves like a soliton or multi-soliton solution, plus a "radiation" component which disperses like a linear solution. Making this rather vaguely worded *soliton resolution conjecture* a rigorous theorem is a major open problem in the field, and somewhat out of reach of current technology except in special cases (e.g. small perturbations of a special solution such as a soliton, multisoliton, or the vacuum state 0, or the one-dimensional cubic NLS, which is completely integrable). However, significantly more is known in the defocusing case $\mu = -1$. In many defocusing cases it is known that soliton-type behaviour is excluded, and all solutions in fact disperse like a linear equation. These results are part of the *scattering theory* for these equations and will be discussed more fully in the next section. For now, let us just say that the question of whether a solution

[31] One of the many difficulties with establishing this conjecture is that we expect there to be a small class of exceptional solutions which exhibit more exotic (and unstable) behaviour, such as periodic "breather" solutions, or clusters of solitons which diverge from each other only logarithmically. Almost all of the known tools in the subject are *deterministic* in the sense that if they work at all, they must work for *all* data in a given class, while to settle this conjecture it may be necessary to develop more "stochastic" techniques that can exclude small classes of exceptional solutions.

disperses or not is intimately tied to whether there is some sort of *decay estimate* for the solution in various norms, such as an L^p_x norm for some $p > 2$; in many cases, knowing that such an L^p_x norm goes to zero as $t \to \infty$ (either in a classical sense, or in some time-averaged sense) is sufficient to establish that the solution scatters to a linear solution, while conversely estimates such as Strichartz estimates assure us that the L^p_x norms of such solutions do indeed go to zero. (This should also be compared with Principle 3.24.)

Thus it is of interest to obtain decay estimates on solutions to defocusing equations. The conservation laws establish boundedness in L^2_x-based spaces such as L^2_x and H^1_x, but do not yield any decay estimates in higher L^p_x norms. There are a number of known ways to establish such a decay estimate; in this section we shall discuss three such, namely the Morawetz inequality approach, the (pseudo)conformal identity approach, and the vector fields approach.

We begin with the Morawetz inequality approach. This method is based on monotonicity formulae, as discussed in Section 1.5. It is here that we can begin to usefully exploit the momentum conservation laws. As momentum is a vector, these laws are not of the coercive type needed to obtain uniform bounds on a solution as in the preceding section, but the vector structure does permit[32] one to construct various quantities based on the momentum density which are monotone in time, and so the fundamental theorem of calculus will provide some spacetime bounds that force some decay in the solution.

In the linear setting, Morawetz inequalities for the NLS and NLW were already introduced in Section 2.4 and Section 2.5, using the pseudo-stress-energy and stress-energy tensors respectively. The NLS and NLW also have such tensors, and in the defocusing case the sign of the nonlinearity will be favourable for preserving the desired monotonicity. In the case of the NLS, we can repeat the derivation of (2.37) (taking into account the new nonlinear term in the T_{jj} components of the pseudo-stress-energy tensor) and obtain the identity

$$
\begin{aligned}
\partial_t^2 \int_{\mathbf{R}^d} a(x)|u(t,x)|^2\, dx &= \partial_t \int_{\mathbf{R}^d} \partial_{x_j} a(x) \mathrm{Im}(\overline{u}(t,x)\partial_{x_j} u(t,x))\, dx \\
&= \int_{\mathbf{R}^d} (\partial_{x_j}\partial_{x_k} a(x)) \mathrm{Re}(\partial_{x_j} u \overline{\partial_{x_k} u})\, dx \\
&\quad + \frac{(p-1)\mu}{p+1} \int_{\mathbf{R}^d} |u(t,x)|^{p+1} \Delta a(x)\, dx \\
&\quad - \frac{1}{4} \int_{\mathbf{R}^d} |u(t,x)|^2 \Delta^2 a(x)\, dx,
\end{aligned}
\tag{3.36}
$$

at least for smooth a of polynomial growth, and for classical, decaying solutions u to (3.1). One can specialise this to $a = |x|$ and $d \geq 3$ (justifying this as in Exercise

[32] Indeed, it is not possible for an (autonomous) quantity based on the mass or energy density to have a non-trivial monotonicity in time, as this would conflict with time reversal symmetry. The momentum density is not subject to this problem, since reversing time also reverses the sign of the momentum density.

2.57) to conclude that

$$\begin{aligned}\partial_t \int_{\mathbf{R}^d} \operatorname{Im}(\overline{u}(t,x)\frac{x}{|x|}\cdot \nabla u(t,x))\ dx = &\int_{\mathbf{R}^d} \frac{|\nabla\!\!\!/ u(t,x)|^2}{|x|}\ dx \\ &+ \frac{2(p-1)\mu}{p+1}\int_{\mathbf{R}^d} \frac{|u(t,x)|^{p+1}}{|x|}\ dx \\ &- \frac{1}{4}\int_{\mathbf{R}^d} |u(t,x)|^2 \Delta^2 a(x)\ dx.\end{aligned} \tag{3.37}$$

When $d \geq 3$ the tempered distribution $\Delta^2 a$ is non-negative. In particular, in the defocusing case this time derivative is non-negative, and we have the monotonicity formula

$$\partial_t \int_{\mathbf{R}^d} \operatorname{Im}(\overline{u}(t,x)\frac{x}{|x|}\cdot \nabla u(t,x))\ dx \gtrsim_p \int_{\mathbf{R}^d} \frac{|u(t,x)|^{p+1}}{|x|}\ dx.$$

Integrating this along a time interval $[t_0, t_1]$ and using Lemma A.10, we obtain the *Morawetz inequality*

$$\int_{t_0}^{t_1}\int_{\mathbf{R}^d} \frac{|u(t,x)|^{p+1}}{|x|}\ dx \lesssim_{p,d} \sup_{t=t_0,t_1} \|u(t)\|^2_{\dot H^{1/2}(\mathbf{R}^d)} \tag{3.38}$$

for any classical solution to defocusing NLS on $[t_0,t_1]\times \mathbf{R}^d$. In practice, the requirement that this solution is classical can be dropped by the usual limiting arguments[33], provided that one is working with a wellposed solution at a regularity strong enough to control both sides of (3.38).

Suppose we are working with an H^1_x-wellposed solution, with a defocusing equation which is H^1_x-subcritical or H^1_x-critical (this turns out to be sufficient to justify the bound (3.38)). As we saw in the preceding section, the conservation laws of mass and energy allow one in this case to show that the solution is global, and bound the H^1_x norm (and hence $\dot H^{1/2}_x$ norm) of $u(t)$ by a quantity depending on the H^1_x norm of the initial datum u_0. Applying this to (3.38) and letting the time interval $[t_0,t_1]$ go to infinity, we obtain the global spacetime bound

$$\int_{\mathbf{R}}\int_{\mathbf{R}^d} \frac{|u(t,x)|^{p+1}}{|x|}\ dxdt \lesssim_{p,d,\|u_0\|_{H^1_x}} 1, \tag{3.39}$$

first observed in [**LStr**] (and inspired by a similar result in [**Mor**] for nonlinear wave equations). This is a decay estimate, as it shows that the quantity $\int_{\mathbf{R}^d} \frac{|u(t,x)|^{p+1}}{|x|}\ dx$ must go to zero, at least in some time-averaged sense. Because the weight $\frac{1}{|x|}$ is large at the origin, this means (roughly speaking) that the solution cannot maintain a significant presence near the origin for extended periods of time. This is a nonlinear effect caused by the defocusing nature of the nonlinearity; it fails utterly in the focusing case $\mu = -1$ (as one can see by inspecting the soliton solution (3.7)), and also behaves strangely in the linear case $\mu = 0$ (see Exercise 3.44). It is especially useful for spherically symmetric solutions, as such solutions already decay away from the origin (Exercise A.19). However, this estimate is not as effective for general solutions, which can be located arbitrarily in space. This problem can be alleviated to some extent by exploiting spatial translation invariance. For sake of simplicity we discuss the three-dimensional case $d = 3$, in which the formulae are

[33] In the case where the NLS is not algebraic, one also needs to regularise the nonlinearity in order to create an approximating sequence of classical solutions, and exploit some stability theory as in Section 3.7; we ignore this rather tedious detail.

cleanest (for higher dimensions, see [**Vis**], [**TVZ**], and for lower dimensions, see [**Gri6**]). By translating (3.37) by y we obtain

$$
\begin{aligned}
\partial_t \int_{\mathbf{R}^3} \mathrm{Im}(\overline{u}(t,x)\frac{x-y}{|x-y|} \cdot \nabla u(t,x))\ dx = \int_{\mathbf{R}^3} \frac{|\nabla\!\!\!/_y u(t,x)|^2}{|x-y|}\ dx \\
+ \frac{2(p-1)\mu}{p+1} \int_{\mathbf{R}^3} \frac{|u(t,x)|^{p+1}}{|x-y|}\ dx \\
+ \pi |u(t,y)|^2
\end{aligned}
\tag{3.40}
$$

where $\nabla\!\!\!/_y$ is the angular component of the gradient using y as the origin. This estimate can then be used to obtain a translated version of (3.39) which prevents the solution u from concentrating at the point y for long periods of time. The freedom afforded by this additional parameter y can be exploited by integrating (3.40) against a suitable weight in y. It turns out that the best weight to achieve this with is the mass density $\mathrm{T}_{00}(t,y) = |u(t,y)|^2$. This gives

$$
\begin{aligned}
&\partial_t \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} |u(t,y)|^2 \mathrm{Im}(\overline{u}(t,x)\frac{x-y}{|x-y|} \cdot \nabla u(t,x))\ dxdy \\
&\quad = \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} |u(t,y)|^2 |\nabla\!\!\!/_y u(t,x)|^2\ \frac{dxdy}{|x-y|} \\
&\quad + \frac{2(p-1)\mu}{p+1} \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} |u(t,y)|^2 |u(t,x)|^{p+1}\ \frac{dxdy}{|x-y|} \\
&\quad + \pi \int_{\mathbf{R}^3} |u(t,y)|^4\ dy \\
&\quad + \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} (\partial_t |u(t,y)|^2) \mathrm{Im}(\overline{u}(t,x)\frac{x-y}{|x-y|} \cdot \nabla u(t,x))\ dxdy.
\end{aligned}
\tag{3.41}
$$

To deal with the final term of (3.41), we rewrite it in terms of the pseudo-stress-energy tensor as

$$
\int_{\mathbf{R}^3} \int_{\mathbf{R}^3} (\partial_t \mathrm{T}_{00}(t,y)) \frac{x_k - y_k}{|x-y|} \mathrm{T}_{0k}(t,x)\ dxdy
$$

and then use (2.35) and integration by parts to write this as

$$
- \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} \mathrm{T}_{0j}(t,y)(1 - \frac{(x_j - y_j)(x_k - y_k)}{|x-y|^2}) \mathrm{T}_{0k}(t,x)\ \frac{dxdy}{|x-y|};
$$

if u is smooth and decaying then there is no difficulty justifying the integration by parts. However, an application of Cauchy-Schwarz shows that

$$
\begin{aligned}
&|\mathrm{T}_{0j}(t,y)(1 - \frac{(x_j - y_j)(x_k - y_k)}{|x-y|^2}) \mathrm{T}_{0k}(t,x)| \\
&\quad \leq |u(t,x)||\nabla\!\!\!/_y u(t,x)||u(t,y)||\nabla\!\!\!/_x u(t,y)| \\
&\quad \leq \frac{1}{2}(|u(t,y)|^2 |\nabla\!\!\!/_y u(t,x)|^2 + |u(t,x)|^2 |\nabla\!\!\!/_x u(t,y)|^2
\end{aligned}
$$

whenever $x \neq y$; this can be seen for instance by rotating $x - y$ to be a multiple of the unit vector e_1 and working in coordinates. From this pointwise bound and symmetry we thus see that we can bound the last term in (3.41) by the first. If we are in the defocusing or linear cases $\mu \geq 0$, we can also drop the second term as being non-negative, and we conclude

$$\partial_t \int_{\mathbf{R}^3} \int_{\mathbf{R}^3} |u(t,y)|^2 \mathrm{Im}(\overline{u}(t,x) \frac{x-y}{|x-y|} \cdot \nabla u(t,x))\ dxdy \geq \pi \int_{\mathbf{R}^3} |u(t,x)|^4\ dx.$$

But by (a translated version of) Lemma A.10 we have the pointwise bound

$$|\int_{\mathbf{R}^3} \int_{\mathbf{R}^3} |u(t,y)|^2 \mathrm{Im}(\overline{u}(t,x) \frac{x-y}{|x-y|} \cdot \nabla u(t,x))\ dxdy| \lesssim \|u(t)\|_{L^2_x}^2 \|u(t)\|_{\dot{H}^{1/2}_x}^2.$$

From the fundamental theorem of calculus, we thus obtain the *interaction Morawetz inequality*

$$\int_{t_0}^{t_1} \int_{\mathbf{R}^3} |u(t,x)|^4\ dxdt \lesssim \sup_{t=t_0,t_1} \|u(t)\|_{L^2_x}^2 \|u(t)\|_{\dot{H}^{1/2}_x}^2 \tag{3.42}$$

whenever u is a classical solution to a defocusing or linear NLS on $[t_0,t_1] \times \mathbf{R}^3$; this was first observed in [**CKSTT10**], and should also be compared with Example 1.34. There is no difficulty applying a limiting argument to extend this inequality to H^1_x-wellposed solutions when the NLS is H^1_x-subcritical or H^1_x-critical. Using the energy and mass conservation laws, we see in particular that we have the spacetime bound

$$\int_{t_0}^{t_1} \int_{\mathbf{R}^3} |u(t,x)|^4\ dxdt \lesssim_{\|u_0\|_{H^1_x}} 1 \tag{3.43}$$

in this case, where u_0 is the initial datum of this solution. This bound resembles (3.39), but is a linear phenomenon rather than a nonlinear one (it holds true even when $\mu = 0$), and does not involve the weight $\frac{1}{|x|}$ and so is not tied to any particular spatial origin[34]. This makes it a more useful decay estimate for controlling solutions to NLS when there is no assumption of spherical symmetry.

The ordinary Morawetz estimates for NLS have a counterpart for NLW; see Exercise 3.46. However, it seems difficult to locate a useful analogue of the interaction Morawetz inequality for the NLW; the somewhat miraculous positivity properties of the time derivative of the interaction functional do not seem to be present in the wave equation setting, even if one drops the nonlinearity. Fortunately, for these equations the ordinary Morawetz estimate is already quite powerful, especially when combined with finite speed of propagation.

The Morawetz inequalities are based on the monotonicity formulae method. Another way to obtain decay is to find a conserved (or almost conserved) quantity which is non-autonomous (depending explicitly on t). Often, such laws arise by conjugating an autonomous conservation law with a symmetry (or approximate symmetry) of the equation. For instance, for the free Schrödinger equation we have already seen that the pseudoconformal symmetry from Exercise 2.28 conjugates the usual energy conservation law to the conservation of the pseudoconformal energy (2.33). Turning now to (classical) solutions u of the nonlinear Schrödinger equation,

[34]While the original Morawetz inequality controls the extent to which a solution can concentrate near a fixed point y, the trick of integrating that inequality against the mass density means that the interaction Morawetz inequality controls the extent to which the solution concentrates against *itself*. In this perspective, the L^4_x quantity $\int_{\mathbf{R}^3} |u(t,x)|^4\ dx$ can be rewritten as $\int_{\mathbf{R}^3} \int_{\mathbf{R}^3} |u(t,x)|^2 |u(t,y)|^2 \delta(x-y)\ dxdy$ and thought of as a self-interaction of the mass density. The interaction inequality can also be thought of as an ordinary Morawetz inequality for a six-dimensional (or "two-particle") Schrödinger equation; see Exercise 3.45.

recall that the pseudoconformal transform v of u, as defined in Exercise 2.28, obeys the modified equation (3.16). In particular, in the L^2_x-critical case $p = p_{L^2_x}$ the pseudoconformal transformation is a symmetry of the equation; we have already used this fact to construct the blowup solution (3.15). But this transform is still useful even in the non-L^2_x-critical case; one can view the equation (3.16) as an NLS in which the degree of focusing or defocusing, as quantified by $t^{\frac{d}{2}(p-p_{L^2_x})}\mu$, is now time-dependent. In analogy with the usual NLS, we can define a (non-autonomous) energy $E[v(t), t]$ for $t \neq 0$ by

$$E_[v(t), t] := \int_{\mathbf{R}^d} \frac{1}{2}|\nabla v(t)|^2 + 2t^{\frac{d}{2}(p-p_{L^2_x})}\mu \frac{|v(t)|^{p+1}}{p+1}\ dx.$$

Because (3.16) is not time translation invariant in general, we do not expect this energy to be perfectly conserved in time (except when $p = p_{L^2_x}$). Nevertheless, it should be "almost" conserved in that its time derivative should be small. Indeed, a computation (essentially the one in Exercise 3.31) shows that

$$\partial_t E[v(t), t] = \frac{d}{2}(p - p_{L^2_x})t^{\frac{d}{2}(p-p_{L^2_x})-1}\mu \int_{\mathbf{R}^d} \frac{|v(t)|^{p+1}}{p+1}\ dx;$$

in other words, the only source for the time variation of the energy of v is the explicit dependence on the time variable t.

This formula can be used to control the time evolution of the energy of v, which in turn gives control on the original solution by means of the easily verified identity

$$E[v(t), t] = E_{\text{pc}}[u(1/t), 1/t] \tag{3.44}$$

where E_{pc} is the *pseudoconformal energy*

$$E_{\text{pc}}[u(t), t] := \frac{1}{2}\int_{\mathbf{R}^d} |(x + it\nabla)u(t)|^2\ dx + 2\frac{\mu t^2}{p+1}\int_{\mathbf{R}^d} |u(t)|^{p+1}, \tag{3.45}$$

which is defined for all t including 0. For instance, in the L^2_x-critical case $p = p_{L^2_x}$, the quantity $E[v(t), t]$ is conserved in t, and hence the pseudoconformal energy $E_{\text{pc}}[u(t), t]$ is also conserved in t. (This was only established for $t \neq 0$, but can be verified for $t = 0$ also, either by a limiting argument or by establishing the conservation of $E_{\text{pc}}[u(t), t]$ directly; see Exercise 3.47.) In the L^2_x-critical defocusing case ($p = p_{L^2_x}, \mu = +1$), we obtain in particular that

$$2\frac{t^2}{p+1}\int_{\mathbf{R}^d} |u(t)|^{p+1} \leq E_{\text{pc}}[u(t), t] = E_{\text{pc}}[u(0), 0] = \frac{1}{2}\|xu(0)\|^2_{L^2_x}$$

which leads to the decay bound

$$\|u(t)\|_{L^{2(d+2)/2}_x} \lesssim_d |t|^{-\frac{d}{d+2}}\|xu(0)\|^{\frac{d}{d+2}}_{L^2_x}$$

for all times $t \neq 0$ for which the (classical) solution exists; this bound can be extended to more general classes of solution by the usual limiting arguments. Some further examples of decay estimates of this type are given in the exercises; they can give quite strong decay for a wide range of powers, but have the drawback that they require some spatial decay on the initial datum (in this case, one needs $xu(0)$ to lie in L^2_x).

For the NLW, the analogue of the pseudoconformal energy is the *conformal energy* $Q[u(t), t]$, which was already introduced in (2.54) for the linear wave equation and is defined the same way for the NLW (using the nonlinear stress-energy tensor defined in (3.32), of course). For classical solutions to NLW one can use the properties of the stress energy tensor to verify the identity

$$\partial_t Q[u(t), t] = -\mu t(d-1)\frac{p-p_c}{p+1}\int_{\mathbf{R}^d} |u(t,x)|^{p+1}\, dx$$

where $p_c := 1 + \frac{4}{d-1}$ is the conformal (or H^1_x-critical) exponent. This can be utilised to obtain decay estimates in analogy with the pseudoconformal energy and the NLS, especially for the conformal power $p = p_c$.

The pseudoconformal and conformal energy methods provide decay of the solution in an L^p_x sense. In some cases one wishes to also establish decay of the solution in an L^∞ sense. This can be done via Sobolev embedding but requires one to control quantities that involve more than one derivative. One way to do this (assuming a sufficiently small, smooth, and localised initial datum) is by the vector field method, which we introduced in Section 2.5. For technical reasons it is a little simpler to work with a derivative nonlinear wave equation rather than a semilinear NLW; for sake of illustration and concreteness we shall work with classical solutions to the (rather artificial) three-dimensional scalar equation $\Box u = (\partial_t u)^3$. We recall the Killing vector fields X^α defined in Section 2.5. We use this to define the higher order energies $E_n(t)$ for any $n = 0, 1, \ldots$ as

$$E_n(t) = \sum_{m=0}^n \sum_{K_1,\ldots,K_m} \|K_1 \ldots K_m u(t)\|^2_{\dot H^1} + \|\partial_t K_1 \ldots K_m u(t)\|^2_{L^2_x}$$

where $K_1, \ldots, K_m$ range over the vector fields ∂_t, ∂_{x_j}, $x_j\partial_{x_k} - x_k\partial_{x_j}$, or $t\partial_{x_j} + x_j\partial_t$. Since the Killing vector fields commute with $\Box$, we have

$$\Box(K_1 \ldots K_m u) = -K_1 \ldots K_m((\partial_t u)^3)$$

and hence by the energy estimate (2.28) we have

$$E_n(t) \lesssim E_n(0) + \sum_{m=0}^n \sum_{K_1,\ldots,K_m} \int_0^t \|K_1 \ldots K_m((\partial_t u)^3)(t')\|_{L^2_x}\, dt'$$

for all $t \geq 0$. Let us apply this with $n = 5$ (this is far more regularity than strictly necessary, but serves to illustrate the general idea). We have at most five Killing vector fields $K_1, \ldots, K_m$ applied to $(\partial_t u)^3$ on the right-hand side. Using the Leibnitz rule repeatedly, we can distribute these derivatives and observe that at most one of the factors $\partial_t u$ will receive more than two of these Killing vector fields. We place that factor in L^2_x and the other two in L^∞, and obtain

$$\begin{aligned}\|K_1 \ldots K_m((\partial_t u)^3)(t')\|_{L^2_x} \lesssim &\big(\sup_{m'\leq 5}\sup_{K_1,\ldots,K_{m'}} \|K_1 \ldots K_{m'}\partial_t u(t')\|_{L^2_x}\big)\\ &\big(\sup_{m'\leq 2}\sup_{K_1,\ldots,K_{m'}} \|K_1 \ldots K_{m'}\partial_t u(t')\|_{L^\infty_x}\big)^2.\end{aligned}$$

Now one observes that the commutator of ∂_t or ∂_{x_j} with one of the vector fields in the list ∂_t, ∂_{x_j}, $x_j\partial_{x_k} - x_k\partial_{x_j}$, or $t\partial_{x_j} + x_j\partial_t$ is a linear combination of the vector fields ∂_t and ∂_{x_j}. Using that fact repeatedly, we can bound

$$\|K_1 \ldots K_{m'}\partial_t u(t')\|_{L^2_x} \lesssim E_5(t)$$

whenever $m' \leq 5$. Applying the Klainerman-Sobolev inequality (see Exercise 3.48) one obtains

$$\|K_1 \ldots K_{m'}\partial_t u(t')\|_{L^\infty_x(\mathbf{R}^3)} \lesssim \langle t'\rangle^{-1} E_5(t).$$

Combining all these estimates together we obtain the integral inequality

$$E_5(t) \lesssim E_5(0) + \int_0^t \langle t' \rangle^{-2} E_5(t')^3 \; dt'.$$

From this and a standard continuity argument we see that if $E_5(0) \lesssim \varepsilon$ for some suffiicently small absolute constant $\varepsilon > 0$, then we have $E_5(t) \lesssim \varepsilon$ for all $t \geq 0$ for which the classical solution u exists. Applying the Klainerman-Sobolev inequality, this leads to decay bounds such as $\|\nabla_{t,x} u(t)\|_{L^\infty_x} \lesssim \langle t \rangle^{-1}$. See [**Sog**] for a more detailed and general treatment of this vector fields approach.

EXERCISE 3.44. Let $d = 3$. Show that the estimate (3.39) continues to hold for the *linear* equation $\mu = 0$ when $5/3 < p < 7$, but fails for $p < 5/3$ or $p > 7$. (Hint: first obtain bounds for $\int_{\mathbf{R}} \int_{|x| \leq R} |u(t,x)|^{p+1} \; dxdt$ for R a power of two by using Strichartz estimates, and then sum in R. For the negative results, start with a bump function initial datum (or a Gaussian) and rescale it as in Exercise 2.42.) The estimate is also true at the endpoints $p = 5/3$ and $p = 7$ but requires a Lorentz space refinement of the Strichartz estimates, observed in [**KTao**].

EXERCISE 3.45. For simplicity let us work with a global classical solution $u : \mathbf{R} \times \mathbf{R}^3 \to \mathbf{C}$ to the three-dimensional linear Schrödinger equation (so $d = 3$ and $\mu = 0$). Define the "two-particle" field $U : \mathbf{R} \times \mathbf{R}^6 \to \mathbf{C}$ by $U(t,x,y) := u(t,x)u(t,y)$. Show that U solves the six-dimensional linear Schrödinger equation. Apply (2.37) to the solution U with the weight $a(x,y) := |x - y|$ (using limiting arguments as necessary to deal with the fact that a is not smooth) and deduce another proof of (3.42) in the linear case $\mu = 0$. How does the argument change when one places a defocusing nonlinearity in the equation?

EXERCISE 3.46 (Morawetz inequality for the wave equation). [**Mor**] Let $u : I \times \mathbf{R}^3 \to \mathbf{C}$ be a classical solution to a three-dimensional defocusing NLW (thus $d = 3$ and $\mu = 1$), and let $\mathrm{T}^{\alpha\beta}$ be the associated stress-energy tensor. Using (3.34) and the identity

$$\mathrm{T}^{jk} = \mathrm{Re}(\partial_{x_j} u \overline{\partial_{x_k} u}) - \frac{\delta_{jk}}{4} \Box(|u|^2) + \frac{p\delta_{jk}}{2(p+1)} |u|^{p+1}$$

for the spatial component of the stress-energy tensor, establish the identity

$$\partial_t \int_{\mathbf{R}^3} \frac{x_j}{|x|} \mathrm{T}^{0j} \; dx = \int_{\mathbf{R}^3} \frac{|\nabla\!\!\!/ u|^2}{|x|} + \frac{p}{p+1} \frac{|u|^{p+1}}{|x|} - \frac{1}{2|x|} \Box(|u|^2).$$

Integrate this in time and use the Hardy inequality (Lemma A.2) to establish the *Morawetz inequality*

$$\int_I \int_{\mathbf{R}^3} \frac{|\nabla\!\!\!/ u|^2}{|x|} \; dxdt + \int_I \int_{\mathbf{R}^3} \frac{|u|^{p+1}}{|x|} \; dxdt + \int_I |u(t,0)|^2 \; dt \lesssim_p E[u]$$

where $E[u] = E[u[t]]$ is the conserved energy; compare this with (3.38).

EXERCISE 3.47. Let u be a classical solution to a NLS. Verify the identity

$$E_{\mathrm{pc}}[u(t),t] = t^2 E[u(t)] - t \int_{\mathbf{R}^d} x_j \mathrm{T}_{0j}(t,x) \; dx + \int_{\mathbf{R}^d} \frac{1}{2} |x|^2 \mathrm{T}_{00}(t,x) \; dx$$

which connects the pseudoconformal energy to the ordinary energy and the pseudo-stress-energy tensor. Use this to verify the evolution law

$$\partial_t E_{\mathrm{pc}}[u(t),t] = -\frac{d\mu t(p - p_{L^2_x})}{p+1} \int_{\mathbf{R}^d} |u(t,x)|^{p+1} \; dx$$

directly, without recourse to the pseudoconformal transformation. From this and Gronwall's inequality, deduce the estimate

$$\|u(t)\|_{L_x^{p+1}}^{p+1} \lesssim_{d,p} t^{-2}\|xu(0)\|_{L_x^2}^2,$$

in the defocusing, L_x^2-supercritical case $\mu = +1, p > p_d$ and all $t \neq 0$, as well as the estimate

$$\|u(t)\|_{L_x^{p+1}}^{p+1} \lesssim_{d,p,t_0} t^{-d(p-1)/2}(\|xu(0)\|_{L_x^2}^2 + \int_0^{t_0}\int_{\mathbf{R}^d} |u(t,x)|^{p+1}\, dxdt),$$

in the defocusing, L_x^2-subcritical case $\mu = +1, p < p_d$ and all $t \geq t_0 > 0$.

EXERCISE 3.48. Let $f \in C_t^\infty \mathcal{S}_x(I \times \mathbf{R}^3)$ for some time interval I. By repeating the arguments used to deduce (2.57) from (2.56), derive the *Klainerman-Sobolev inequality*

$$\|\nabla_{t,x} f(t)\|_{L_x^\infty(\mathbf{R}^3)} \lesssim \langle t\rangle^{-1} \sum_{m\leq 3}\sum_{K_1,\ldots,K_m} \|\nabla_{t,x} K_1 \ldots K_m f(t)\|_{L_x^2(\mathbf{R}^3)} \tag{3.46}$$

for all $t \in I$, where $K_1, \ldots, K_m$ ranges over all the vector fields ∂_t, ∂_{x_j}, $x_j\partial_{x_k} - x_k\partial_{x_j}$, or $t\partial_{x_j} + x_j\partial_t$.

3.6. Scattering theory

To know the road ahead, ask those coming back. (attributed to Confucius)

The decay estimates of the preceding section give asymptotic control for global solutions to NLS or NLW. It turns out in many cases, these estimates can be bootstrapped to provide quite strong control on these solutions, in particular establishing that they *scatter* to a linear solution. Intuitively, the reason for this is that if $u(t)$ decays to zero as $t \to \pm\infty$, then the nonlinearity $\mu|u(t)|^{p-1}u(t)$ decays even faster, and so the nonlinear component of the NLS or NLW equation will vanish asymptotically (in a relative sense), and thus (by Principle 1.37) we expect the evolution to behave linearly as $t \to \pm\infty$. The main tool for making these heuristics rigorous is the Duhamel formula ((3.22) or (3.23)), applied for various values of t and t_0; the arguments often bear some similarity with the Duhamel iteration arguments used to establish local existence, though with some subtle differences. For instance, in the local theory, large exponents p are more difficult to deal with than small exponents (because they exacerbate the large values of the solution, which are the main source of difficulty in closing a local iteration argument), but in the asymptotic theory, the small exponents tend to be the most difficult[35] (because they do not attenuate the small values of the solution as much as the large exponents, and so the nonlinearity does not decay as fast asymptotically). Since one needs to combine the local and asymptotic theories to understand scattering, it should thus be unsurprising that most scattering results only hold for exponents p that are neither too large nor too small. (For instance, recall from our discussion of the exact solutions (3.18), (3.19) that we do not expect scattering results when $p \leq 1 + \frac{2}{d}$.) It is sometimes useful to ensure that one's arguments are as scale-invariant as possible, as this can allow one to treat both the local theory and the asymptotic theory in a unified manner.

[35] For similar reasons, the asymptotic theory sometimes gets *easier* when there are derivatives in the nonlinearity, despite the fact that these derivatives can make the local theory significantly harder.

We begin by discussing the scattering theory for NLS in the energy class H^1_x. To reduce the number of cases slightly we shall only consider scattering from $t = 0$ to $t = +\infty$ or vice versa; one can certainly consider scattering back and forth between $t = 0$ and $t = -\infty$, or between $t = -\infty$ and $t = +\infty$, but the theory is more or less the same in each of these cases. We will also assume that the nonlinearity is either H^1_x-subcritical or H^1_x-critical, so that we have a good H^1_x-wellposedness theory (locally in time, at least).

A solution to the linear Schrödinger equation in this class takes the form $e^{it\Delta/2}u_+$ for some $u_+ \in H^1_x$. We say that a global strong H^1_x solution u to the nonlinear equation (3.1) with initial datum $u(0) = u_0$ *scatters in* H^1_x to a solution $e^{it\Delta/2}u_+$ to the linear equation as $t \to +\infty$ if we have

$$\|u(t) - e^{it\Delta/2}u_+\|_{H^1_x} \to 0 \text{ as } t \to +\infty,$$

or equivalently (by using the unitarity of $e^{it\Delta/2}$)

$$\|e^{-it\Delta/2}u(t) - u_+\|_{H^1_x} \to 0 \text{ as } t \to +\infty.$$

In other words, we require the functions[36] $e^{-it\Delta/2}u(t)$ converge in H^1_x as $t \to +\infty$. From the Duhamel formula (3.22) we have

$$e^{-it\Delta/2}u(t) = u_0 - i\mu \int_0^t e^{-it'\Delta/2}(|u(t')|^{p-1}u(t'))\ dt'$$

and so u scatters in H^1_x as $t \to +\infty$ if and only if the improper integral

$$\int_0^\infty e^{-it\Delta/2}(|u(t)|^{p-1}u(t))\ dt$$

is conditionally convergent in H^1_x, in which case the asymptotic state u_+ is given by the formula

$$u_+ = u_0 - i\mu \int_0^\infty e^{-it\Delta/2}(|u(t)|^{p-1}u(t))\ dt. \tag{3.47}$$

Thus one can view the asymptotic state u_+ as a nonlinear perturbation of the initial state u_0. If we compare (3.47) with (3.22) and eliminate u_0 we obtain the identity

$$u(t) = e^{it\Delta/2}u_+ + i\mu \int_t^\infty e^{i(t-t')\Delta/2}(|u(t')|^{p-1}u(t'))\ dt' \tag{3.48}$$

which can be viewed as the limiting case $t_0 = +\infty$ of (3.22).

Suppose that for every asymptotic state $u_+ \in H^1_x$ there exists a unique initial datum $u_0 \in H^1_x$ whose corresponding H^1_x-wellposed solution is global and scatters to $e^{it\Delta/2}u_+$ as $t \to +\infty$. Then we can define the *wave operator* $\Omega_+ : H^1_x \to H^1_x$ by $\Omega_+ u_+ := u_0$. Note that the uniqueness aspect of the H^1_x-wellposedness theory ensures that the wave operator is injective. If it is also surjective - in other words, if *every* H^1_x-wellposed solution is global and scatters in H^1_x as $t \to +\infty$, we say that we also have *asymptotic completeness*.

In general, the existence of wave operators is relatively easy to establish (as long as the power p is not too small or too large, and especially if a smallness condition is assumed), both in focusing and defocusing cases. The asymptotic completeness,

[36] It is instructive to write $e^{-it\Delta/2}u(t) = S_{\mathrm{lin}}(t)^{-1}S(t)u_0$, where $S(t) : u_0 \mapsto u(t)$ is the propagator for the nonlinear Schrödinger equation, and $S_{\mathrm{lin}}(t)$ is the corresponding linear propagator. Thus scattering is an assertion that the "gap" between $S(t)$ and $S_{\mathrm{lin}}(t)$ converges to something bounded in H^1_x as $t \to \infty$.

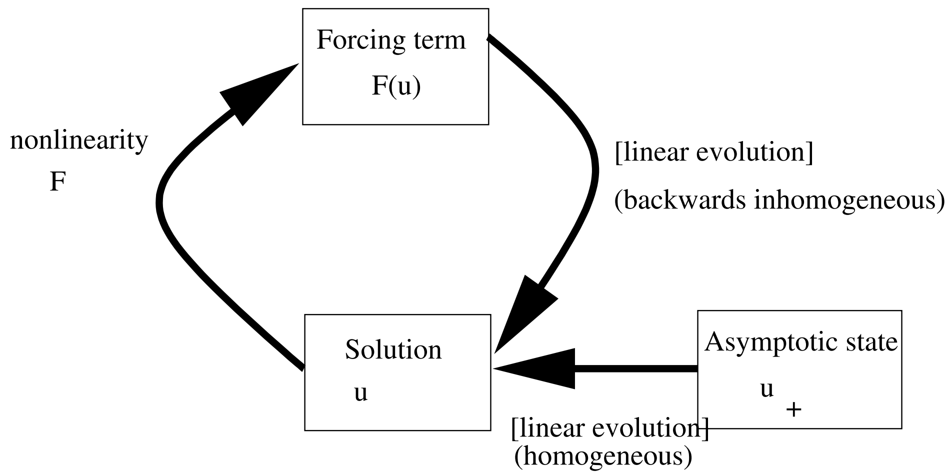

FIGURE 8. The iteration scheme used to construct a solution from an asymptotic state at late times; it is essentially a backwards-in-time version of the local existence scheme, but on an unbounded time interval.

however, is a bit harder, is restricted to the defocusing case (since soliton solutions clearly do not scatter to linear solutions), and requires the decay estimates. We will not attempt a complete theory here, but just illustrate with a single example, namely the cubic defocusing three-dimensional NLS ($d = 3, p = 3, \mu = +1$). Note that global wellposedness for H^1_x for this equation (in the subcritical sense) was already established in Exercise 3.35.

PROPOSITION 3.28 (Existence of wave operators). *Let $d = 3$, $p = 3$, and $\mu = +1$. Then the wave operator $\Omega_+ : H^1_x \to H^1_x$ exists and is continuous.*

PROOF. (Sketch) To construct the wave operator Ω_+, we need to evolve a state at $t = +\infty$ to $t = 0$. We shall factor this problem into two sub-problems; first we shall solve the "asymptotic problem", getting from $t = +\infty$ to some finite time $t = T > 0$, and then we will solve the "local problem" of getting from $t = T$ to $t = 0$. The latter problem will be an immediate consequence of the global wellposedness problem, so we focus on the former. We shall use the same Duhamel iteration method used to prove Proposition 3.19, but with (3.48) being used instead of the usual Duhamel formula (3.22). Fix $u_+ \in H^1_x$; we will assume the bound $\|u_+\|_{H^1_x} \leq A$ for some $A > 0$. From the Strichartz estimate (3.28) we have

$$\|e^{it\Delta/2}u_+\|_{S^1(\mathbf{R}\times\mathbf{R}^3)} \lesssim_A 1.$$

We would like to make this norm not only bounded, but small, by restricting the time variable. This is not possible at present because the S^1 norm contains some components of L^∞_t type, which do not necessarily shrink upon restricting time. To fix this we shall pass from S^1 to a smaller controlling norm; a convenient choice here is the norm

$$\|u\|_{\mathcal{S}_0} := \|u\|_{L^5_{t,x}} + \|u\|_{L^{10/3}_t W^{1,10/3}_x}.$$

From Sobolev embedding we have

$$\begin{aligned}\|e^{it\Delta/2}u_+\|_{\mathcal{S}_0(\mathbf{R}\times\mathbf{R}^3)} &\lesssim \|e^{it\Delta/2}u_+\|_{L^5_t W^{1,30/11}_x(\mathbf{R}\times\mathbf{R}^3)} + \|e^{it\Delta/2}u_+\|_{L^{10/3}_t W^{1,10/3}_x(\mathbf{R}\times\mathbf{R}^3)} \\ &\lesssim \|e^{it\Delta/2}u_+\|_{S^1(\mathbf{R}\times\mathbf{R}^3)} \\ &\lesssim_A 1.\end{aligned}$$

Let $\varepsilon > 0$ be a small absolute constant to be chosen later. If we set $T = T(u_+)$ large enough, we see from monotone convergence that

$$\|e^{it\Delta/2}u_+\|_{\mathcal{S}_0([T,+\infty)\times\mathbf{R}^3)} \leq \varepsilon.$$

We now solve (3.48) in the spacetime slab $[T, +\infty) \times \mathbf{R}^3$ by iteration, keeping the iterates bounded in $S^1([T, +\infty) \times \mathbf{R}^3)$ and small in $\mathcal{S}_0$, and the nonlinearity $|u|^{p-1}u$ small in $L^{10/7}_t W^{1,10/7}_x([T, +\infty)\times\mathbf{R}^3)$. This constructs a unique solution $u \in S^1([T, +\infty)\times\mathbf{R}^3)$ to (3.48), which can be shown to be a strong H^1_x solution to (3.1) in this interval by a variant of Exercise 3.10. Using the global H^1_x-wellposedness theory, one can then extend this solution uniquely to $S^1([0, +\infty) \times \mathbf{R}^3)$, and in particular u will take some value $u_0 = u(0) \in H^1_x$ at time $t = 0$. This gives existence of the wave map; continuity can be established by concatenating the continuity given from the above iteration scheme in the interval $[T, +\infty)$ with the continuity arising from the global wellposedness in the interval $[0, T]$; note that the time T can be chosen to be uniform under small H^1_x perturbations in u_+ thanks to the Strichartz estimates. The uniqueness can be made to be unconditional (in the category of strong H^1_x solutions) by arguing as in Exercise 3.17. □

Remark 3.29. The above argument shows that (perhaps unintuitively) it is in fact *easier* to evolve from an asymptotic state at $t = +\infty$ to a large finite time $t = T$, than it is to evolve from $t = T$ down to $t = 0$, as the former does not even require energy conservation or the defocusing sign of the nonlinearity. The reason for this is that in the asymptotic regime $t \to +\infty$, the asymptotic state is so dispersed that the nonlinear effects are extremely weak; it is only at time T and below that the solution reaches sufficient levels of concentration that one must start paying more serious attention to the nonlinearity.

Now we establish asymptotic completeness. For pedagogical purposes we shall split the argument into three parts. First we begin with a conditional result, that shows that asymptotic completeness is implied by a certain spacetime bound; this is a purely perturbative argument that does not require any decay estimates. Then, we show that this rather strong spacetime bound is implied by a seemingly weaker spacetime bound. Finally, we use the decay estimates of the previous section to establish that spacetime bound.

Proposition 3.30 (Spacetime bound implies asymptotic completeness). *Let $d = 3$, $p = 3$, and $\mu = +1$. Suppose that there exists a bound of the form*

$$\|u\|_{S^1(\mathbf{R}\times\mathbf{R}^3)} \lesssim_{\|u_0\|_{H^1_x}} 1 \tag{3.49}$$

for all H^1_x-wellposed solutions to (3.1) *(thus we assume that the nonlinear equation obeys the same type of global Strichartz estimate as the linear equation). Then the wave operator Ω_+ is surjective from H^1_x to H^1_x, and the inverse Ω_+^{-1} is continuous. (In conjunction with Proposition 3.28, this implies that Ω_+ is a homeomorphism from H^1_x to itself.)*

Proof. We shall demonstrate the surjectivity here, and leave the continuity to an exercise. We need to show that for any $u_0 \in H^1_x$, the global H^1_x-wellposed solution u to (3.1) scatters in H^1_x; by the preceding discussion, this is equivalent to the conditional convergence of the integral $\int_0^\infty e^{-it\Delta/2}(|u(t)|^2 u(t))\ dt$ in H^1_x. By Strichartz estimates (e.g. (3.28)), it will suffice to show that $|u|^2 u$ lies in $N^1(\mathbf{R}\times\mathbf{R}^3)$. But from the Leibnitz rule and Hölder's inequality, followed by Sobolev embedding we have

$$\begin{aligned}
\||u|^2 u\|_{N^1(\mathbf{R}\times\mathbf{R}^3)} &\lesssim \sum_{k=0}^1 \|\nabla^k(|u|^2 u)\|_{L^{10/7}_{t,x}(\mathbf{R}\times\mathbf{R}^3)} \\
&\lesssim \sum_{k=0}^1 \||u|^2|\nabla^k u|\|_{L^{10/7}_{t,x}(\mathbf{R}\times\mathbf{R}^3)} \\
&\lesssim \|u\|^2_{L^5_{t,x}} \|u\|_{L^{10/3}_t W^{1,10/3}_x(\mathbf{R}\times\mathbf{R}^3)} \\
&\lesssim \|u\|^3_{S^1(\mathbf{R}\times\mathbf{R}^3)}
\end{aligned}$$

and the claim follows by (3.49). □

Proposition 3.31 (Weak spacetime bound implies strong spacetime bound). *Let $d = 3$, $p = 3$, and $\mu = +1$. Suppose that there exists a bound of the form*

$$\|u\|_{L^q_{t,x}(\mathbf{R}\times\mathbf{R}^3)} \lesssim_{\|u_0\|_{H^1_x}} 1 \tag{3.50}$$

for all H^1_x-wellposed solutions to (3.1) *and some fixed* $10/3 \leq q \leq 10$. *Then we have* (3.49).

Note that Sobolev embedding shows that the L^q_x norm is controlled by the S^1 norm. The S^1 norm is ostensibly a stronger norm as it also controls one derivative of the solution, but the point is that Strichartz estimates will allow one to control this stronger norm by the weaker norm (and the energy). This bootstrapping phenomenon is typical for any subcritical equation (reflecting a certain amount of "room" in the iteration argument); for critical equations, the situation is more delicate as the relevant Strichartz norm is now scale-invariant and thus can only be controlled by other scale-invariant quantities; see Exercise 3.51. One can also combine this result with persistence of regularity results such as Proposition 3.11, giving in particular the bound

$$\|u\|_{C^0_t H^s_x(\mathbf{R}\times\mathbf{R}^3)} \lesssim_{s,\|u_0\|_{H^1_x}} \|u_0\|_{H^s_x(\mathbf{R}^3)}$$

for any $s \geq 0$ for which the right-hand side is finite.

Proof. Let u be an H^1_x-wellposed solution to (3.1). We shall apply a perturbative argument; to do this, we need the solution u to be made small in some sense. This shall be accomplished by partitioning the time axis[37].

Let $\varepsilon = \varepsilon(\|u_0\|_{H^1_x}) > 0$ be a small number to be chosen later. Using (3.50), we can divide the time axis $\mathbf{R}$ into $O_{\varepsilon,q,\|u_0\|_{H^1_x}}(1)$ intervals I, such that on each such

[37]This is very similar to how one iterates a local existence result to a global one, as in Figure 7. A key difference is that the time intervals considered here can be arbitrarily large or even infinite. In practice, this means that we are no longer permitted to use Hölder in time (except perhaps on some exceptionally short intervals), as we generally cannot afford to lose a power of the length of the time interval.

interval we have

$$\|u\|_{L^q_x(I\times\mathbf{R}^3)} \leq \varepsilon. \tag{3.51}$$

Now fix one of these intervals I, say $I = [t_0, t_1]$. From (3.28) we have

$$\|u\|_{S^1(I\times\mathbf{R}^3)} \lesssim \|u(t_0)\|_{H^1_x(\mathbf{R}^3)} + \||u|^2 u\|_{N^1(I\times\mathbf{R}^3)}.$$

From energy conservation we have $\|u(t_0)\|_{H^1_x(\mathbf{R}^3)} = O_{\|u_0\|_{H^1_x}}(1)$. Now we argue as in the proof of Proposition 3.30. Estimating the N^1 norm by the $L^{10/7}_t W^{1,10/7}_x$ norm and using the Leibniz rule and Hölder inequality, we see that

$$\begin{aligned}\||u|^2 u\|_{N^1(I\times\mathbf{R}^3)} &\lesssim \sum_{k=0}^{1} \||u|^2 |\nabla^k u|\|_{L^{10/7}_{t,x}(I\times\mathbf{R}^3)} \\ &\lesssim \|u\|^2_{L^5_{t,x}(I\times\mathbf{R}^3)} \|u\|_{L^{10/3}_t W^{1,10/3}(I\times\mathbf{R}^3)} \\ &\lesssim \|u\|^2_{L^5_{t,x}(I\times\mathbf{R}^3)} \|u\|_{S^1(I\times\mathbf{R}^3)}.\end{aligned}$$

Now from the definition of S^1 and Sobolev embedding we have

$$\|u\|_{L^r_{t,x}(I\times\mathbf{R}^3)} \lesssim \|u\|_{S^1(I\times\mathbf{R}^3)}$$

for all $10/3 \leq r \leq 10$. Interpolating this with (3.51) we obtain

$$\|u\|_{L^5_{t,x}(I\times\mathbf{R}^3)} \lesssim \varepsilon^\alpha \|u\|^{1-\alpha}_{S^1(I\times\mathbf{R}^3)}$$

for some $0 < \alpha < 1$ depending on q. Combining all these estimates we obtain

$$\|u\|_{S^1(I\times\mathbf{R}^3)} \lesssim O_{\|u_0\|_{H^1_x}}(1) + \varepsilon^\alpha \|u\|^{1-\alpha}_{S^1(I\times\mathbf{R}^3)}.$$

If we choose ε sufficiently small depending on $\|u_0\|_{H^1_x}$, then standard continuity arguments (see Exercise 1.21) yields

$$\|u\|_{S^1(I\times\mathbf{R}^3)} = O_{\|u_0\|_{H^1_x}}(1).$$

Summing this over all of the intervals I we obtain (3.49) as desired. □

It thus remains to establish the spacetime bound (3.50). In the case of spherically symmetric solutions, one can combine the ordinary Morawetz inequality (3.39), which in this case gives

$$\int_{\mathbf{R}} \int_{\mathbf{R}^3} \frac{|u(t,x)|^4}{|x|}\, dxdt \lesssim_{\|u_0\|_{H^1_x}} 1,$$

with the radial Sobolev inequality (Exercise A.19), which when combined with the conservation of mass and energy give

$$\|u(t,x)|x|\|_{L^\infty_t L^\infty_x(\mathbf{R}\times\mathbf{R}^3)} \lesssim_{\|u_0\|_{H^1_x}} 1.$$

Multiplying the two gives

$$\|u\|_{L^5_{t,x}(\mathbf{R}\times\mathbf{R}^3)} \lesssim_{\|u_0\|_{H^1_x}} 1,$$

which is of the desired form (3.50) with $q = 5$. Note how the Morawetz inequality provides the decay near the origin, while the radial Sobolev inequality provides the decay away from the origin. In the non-radial case, we cannot run this argument so easily (though see [**Bou6**]); however the interaction Morawetz inequality (3.43) yields (3.50) immediately (with $q = 4$).

The above types of arguments are known to give scattering results in H^1_x for defocusing NLS equations which are strictly between the H^1_x-critical and L^2_x-critical powers; see [**Caz2**]. A scattering theory at the H^1_x-critical endpoint (based upon a spacetime bound such as (3.50)) has recently been established but is significantly more difficult; see Chapter 5. The H^1_x-scattering theory for the L^2_x-critical equation remains open, even in the spherically symmetric defocusing case. Similar remarks also apply to the NLW, but with the role of the L^2_x-critical exponent now played by the $\dot H^{1/2}_x$-critical (conformal) exponent.

For NLS equations below the L^2_x-critical exponent, no scattering theory is known in H^1_x, but one can extend the range of exponents for which a scattering result is known by assuming more spatial decay on the solution. For instance, one can work in the pseudoconformal space

$$\Sigma := \{u_0 \in H^1_x(\mathbf{R}^d) : xu_0 \in L^2_x(\mathbf{R}^d)\},$$

as one can now utilise the pseudoconformal decay laws for such initial data (such as those in Exercise 3.47). It turns out that the exponent p still needs to be above a certain threshold in order for that decay law to be strong enough to give scattering; more precisely, if we have a defocusing NLS with

$$\frac{2+d+\sqrt{d^2+12d+4}}{4d} < p \leq 1 + \frac{4}{d-2}$$

and $u_0 \in \Sigma$, then there is a global H^1_x-wellposed solution u, and $e^{-it\Delta/2}u(t)$ converges in Σ to some asymptotic state $u_+ \in \Sigma$. See [**Caz2**], [**TVZ**]. On the other hand, for NLS equations in which the power p is less than or equal to $1 + \frac{2}{d}$, the asymptotic effects of the nonlinearity are not negligible, and it is known that the solution does not in general scatter to a free solution; see Section 3.8.

The pseudoconformal transformation is a useful tool for analyzing the asymptotic behaviour of NLS, because it swaps the asymptotic regime $t \to +\infty$ with the local regime $t \to 0^+$ (though at the possible cost of introducing a singularity at $t = 0$). This transformation should also (heuristically at least) swap the initial datum with its Fourier transform, or something resembling its Fourier transform; see Exercise 2.28. The Fourier transform swaps H^1_x to the weighted space $H^{0,1}_x := \{u_0 : \langle x\rangle u_0 \in L^2_x\}$, and so one might expect to be able to use this transformation to somehow intertwine the H^1_x theory with an $H^{0,1}_x$ theory. A sample result is as follows.

Proposition 3.32. *Consider the two-dimensional defocusing cubic NLS (thus $d = 2, p = 3, \mu = +1$, and the equation is L^2_x-critical). Let $u_0 \in H^{0,1}_x$. Then there exists a global L^2_x-wellposed solution to* (3.1)*, and furthermore the $L^4_{t,x}(\mathbf{R} \times \mathbf{R}^3)$ norm of u_0 is finite.*

The $L^4_{t,x}$ bound is sufficient to yield a scattering result in L^2_x; see Exercise 3.54. In contrast, for H^1_x data, no scattering result is known (the Morawetz inequalities do some decay here, but it is not scale-invariant), while for L^2_x data, not even global existence is known (unless the mass is small).

Proof. We shall use an argument from [**BC**]. By time reversal symmetry and gluing arguments we may restrict attention to the time interval $[0, +\infty)$. Since u_0 lies in $H^{0,1}_x$, it also lies in L^2_x. Applying the L^2_x wellposedness theory (Proposition 3.17) we can find an L^2_x-wellposed solution $u \in S^0([0, T] \times \mathbf{R}^3)$ on some time interval $[0, T]$, with $T > 0$ depending on the profile of u_0. In particular the $L^4_{t,x}([0, T] \times \mathbf{R}^3)$ norm of u is finite. Next we apply the pseudoconformal law (which is exact in the

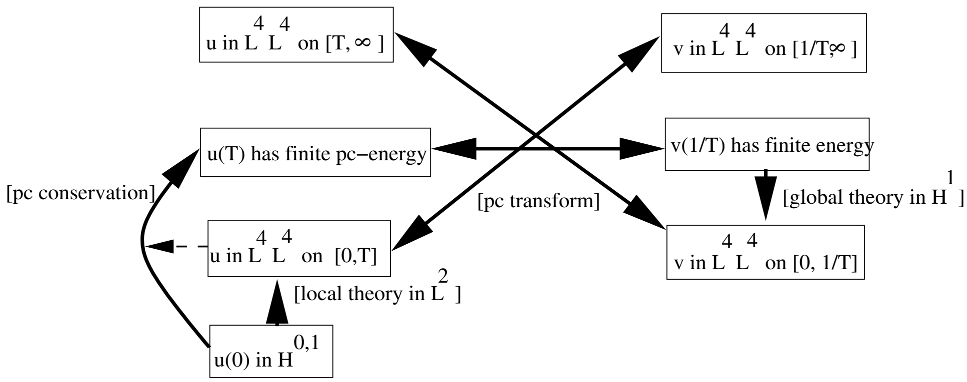

FIGURE 9. The scheme used to establish a global $L^4_{t,x}$ bound from $H^{0,1}_x$ initial data. Notice how the original and pseudoconformal viewpoints together form a "coordinate chart" for the compactified time interval $[0, +\infty]$, thus reducing a global problem to two local ones.

L^2_x-critical case, and can be justified by the usual limiting arguments) to conclude that

$$E_{\mathrm{pc}}[u(T), T] = E_{\mathrm{pc}}[u_0, 0] = \frac{1}{2} \int_{\mathbf{R}^2} |xu_0|^2 \, dx < \infty$$

since $u_0 \in H^{0,1}_x$.

We have obtained a solution from $t = 0$ to $t = T$. To go all the way to $t = +\infty$ we apply the pseudoconformal transformation (Exercise 2.28) at time $t = T$, obtaining an initial datum $v(1/T)$ at time $1/T$ by the formula

$$v(1/T, x) := \frac{1}{i/T} \overline{u(T, Tx)} e^{iT|x|^2/2}.$$

From (3.44) we see that v has finite energy:

$$\frac{1}{2} \int_{\mathbf{R}^2} |\nabla v(1/T, x)|^2 \, dx + \frac{1}{2} \int_{\mathbf{R}^2} |v(1/T, x)|^4 \, dx = E_{\mathrm{pc}}[u(T), T] < \infty.$$

Also, the pseudoconformal transformation conserves mass and hence

$$\int_{\mathbf{R}^2} |v(1/T, x)|^2 \, dx = \int_{\mathbf{R}^2} |u(T, x)|^2 \, dx = \int_{\mathbf{R}^2} |u_0(x)|^2 \, dx < \infty.$$

We thus see that $v(1/T)$ has a finite H^1_x norm. We can thus use the global H^1_x-wellposedness theory (from Exercise 3.35) backwards in time to obtain an H^1_x-wellposed solution $v \in S^1([0, 1/T] \times \mathbf{R}^2)$ to the equation (3.16), which in this case is identical to the original NLS: $i\partial_t v + \frac{1}{2}\Delta v = |v|^2 v$. In particular, $v \in L^4_{t,x}([0, 1/T] \times \mathbf{R}^2)$. We now invert the pseudoconformal transformation, which now defines the original field u on the new slab $[1/T, \infty) \times \mathbf{R}^2$. From Exercise 2.38 we see that the $L^4_{t,x}([1/T, \infty) \times \mathbf{R}^2)$ and $C^0_t L^2_x([1/T, \infty) \times \mathbf{R}^2)$ norms of u are finite. This is enough to make u an L^2_x-wellposed solution to NLS on the time interval $[1/T, \infty)$; for v classical this is an immediate consequence of Exercise 2.28, and for general $v \in S^1([0, 1/T] \times \mathbf{R}^2)$ the claim follows by a limiting argument using the H^1_x-wellposedness theory. Gluing together the two intervals $[0, 1/T]$ and $[1/T, \infty)$, we have obtained a global $L^4_{t,x}([0,+\infty) \times \mathbf{R}^2)$ solution u to (3.1) as desired. We summarise the above argument in Figure 9. □

REMARK 3.33. One can go through the above argument and extract an explicit bound on the global $L^4_{t,x}$ norm of the solution u, but it depends on the profile of the initial datum u_0 and not just on its $H^{0,1}_x$ norm (as this is what determines how small T is). Indeed, if one could obtain a bound depending only in the $H^{0,1}_x$ norm then the scaling invariance and a limiting argument would allow one to replace $H^{0,1}_x$ with L^2_x, which would lead to the (still open) result that one has global wellposedness and scattering in L^2_x for this equation. The above argument can also be generalised, linking a wellposedness theory in H^s_x to a scattering theory in $H^{0,s}$ for any $s \geq 0$ and any L^2_x-critical equation; see [**BC**].

Observe how in the above argument, the pseudoconformal transformation was used to convert an asymptotic time horizon $t = +\infty$ to a finite time horizon $t = 0$, thus allowing one to use local theory to obtain asymptotic control of the solution. There is a somewhat similar trick for wave equations known as *conformal compactification*, in which one applies a conformal transformation of Minkowski spacetime to a pre-compact Lorentzian manifold (the "Penrose diamond"). In one dimension $d = 1$, this compactification is especially simple in null coordinates $u := t+x, v := t-x$, as it is given simply by $(u, v) \mapsto (\tan^{-1} u, \tan^{-1} v)$, thus mapping $\mathbf{R}^{1+1}$ to the diamond-shaped region $\{(t, x) \in \mathbf{R}^{1+1} : |t + x|, |t - x| < \pi/2\}$. If the equation is of a suitable type (it typically must obey some sort of "null condition", or the nonlinearity must be sufficiently high order), then this transformation does not introduce any severe singularities at the boundary of the diamond (cf. (3.16) in the case $p < p_{L^2_x}$), and one can use local theory on the Penrose diamond to obtain a transformed solution on the entire diamond (if the datum is sufficiently small, smooth, and decaying), and then by inverting the conformal compactification one obtains a global solution in Minkowski space. Typically the transformed solution extends to the boundary of the Penrose diamond and beyond, which often leads to scattering-type behaviour for the original solution. See [**Chr**].

Several of the above methods can also be used to establish various scattering results for NLW; the conformal conservation law, Morawetz estimates, the vector fields method, and the conformal compactification methods are particularly useful. Because of finite speed of propagation, one can often reduce the case of compactly supported data. We will not discuss these results here, except in the energy-critical case which we treat in see Section 5.1, and refer the reader to [**Stra**], [**GV6**], [**BZS**], [**GV5**], [**Nak3**], [**Hid**].

EXERCISE 3.49. Complete the proof of Proposition 3.28. (Full details can also be found in [**Caz2**].)

EXERCISE 3.50. Establish the continuity component of Proposition 3.30. (One may need to divide the time axis into intervals on which certain spacetime norms are small, in order that the perturbative argument can apply to give local continuity. Then concatenate the results to obtain global continuity.)

EXERCISE 3.51. Suppose one replaces the H^1_x-subcritical cubic NLS in Proposition 3.31 with the H^1_x-critical quintic NLS (so $d = 3, p = 5, \mu = +1$). Show that one can still prove this Proposition if one fixes $q = 10$ (this is the unique value of q which is invariant under the scaling symmetry of the equation).

EXERCISE 3.52. Suppose one is working with a global $H^1_x(\mathbf{R}^3)$-wellposed solution u to either the cubic or quintic three-dimensional NLS (with either sign of nonlinearity). Suppose it is known that the potential energy $\frac{1}{p+1}\int_{\mathbf{R}^3}|u(t,x)|^{p+1}\,dx$ goes to zero as $t\to\infty$. Conclude that the solution scatters in H^1_x to an asymptotic state $e^{it\Delta/2}u_+$. (This is yet another affirmation of Principle 3.24.)

EXERCISE 3.53 (Blowup criterion for H^1_x-critical NLS). Suppose that $u \in C^0_{t,\mathrm{loc}}H^1_x([0,T_*)\times\mathbf{R}^3)$ is a strong H^1_x solution to quintic NLS (so $d=3$ and $p=5$) which cannot be continued beyond a finite time T_* as a strong solution. Show that the $L^{10}_{t,x}([0,T_*)\times\mathbf{R}^3)$ norm of u is infinite. (Hint: argue by contradiction and obtain an $\dot{S}^1([0,T_*)\times\mathbf{R}^3)$ bound on u. Conclude that for times t close to T, both the linear and nonlinear evolution of $u(t)$ will be small in $L^{10}_{t,x}([t,T_*)\times\mathbf{R}^3)$, and hence for $L^{10}_{t,x}([t,T_*+\varepsilon)\times\mathbf{R}^3)$ for some $\varepsilon>0$, contradicting the hypothesis that T_* is the maximal time of existence.)

EXERCISE 3.54. Consider the two-dimensional defocusing cubic NLS (thus $d=2, p=3, \mu=+1$). Show that if a global L^2_x-wellposed solution u is known to have finite $L^4_{t,x}(\mathbf{R}\times\mathbf{R}^2)$ norm, then $e^{-it\Delta/2}u(t)$ converges in L^2_x to an asymptotic state $u_+\in L^2_x$ as $t\to+\infty$.

3.7. Stability theory

True life is not lived where great external changes take place - where people move about, clash, fight, and slay one another - it is lived only where these tiny, tiny, infinitesimally small changes occur. (Leo Tolstoy, "Why Do People Intoxicate Themselves?")

The differential equations that one studies in mathematics, such as NLS and NLW, often arise from physics as simplified models for more complicated systems. In reality, the actual equations that govern a physical system will not evolve by these model equations exactly, but will contain some additional terms. For sake of discussion let us fix the model equation to be the NLS (3.1). Instead of solving NLS, the true system may be governed by a field $\tilde{u}$ which obeys a perturbed NLS

$$i\partial_t\tilde{u}+\frac{1}{2}\Delta\tilde{u}=\mu|\tilde{u}|^{p-1}\tilde{u}+e;\quad \tilde{u}(t_0)=u_0+e_0 \tag{3.52}$$

where the forcing term $e=e(t,x)$ and the initial datum error $e_0=e_0(x)$ are small, and possibly depending on $\tilde{u}$ and on some external forces[38]. It is thus of interest to develop a *stability theory* for equations such as NLS, which would guarantee that the solution to a perturbed NLS does not deviate too far from the solution to the actual NLS if e and e_0 are small some suitable norms. Note that this would generalise the property of *continuous dependence of the data*, which is already given by the wellposedness theory and corresponds to the special case $e=0$. It also generalises the uniqueness theory, which can be viewed as the case $e=e_0=0$. A strong stability theory lends confidence as to the robustness of the results obtained for the model equation. Conversely, if a PDE is known to be very unstable then this would

[38]The field of *stochastic partial differential equations* studies such equations with the assumption that e is some stochastic field, e.g. Gaussian white noise. These random fluctuations often serve to regularise the behaviour of the equation and keep it well-behaved even when the deterministic equation is not known to be wellposed. However, we shall focus on the deterministic theory in which e is fixed, which is the case needed for applications such as construction of solutions via approximate solutions.

cast doubt on the ability of that PDE to accurately simulate (either numerically or theoretically) a real-life system, except perhaps in some stochastic sense.

A stability theory is also useful to have in the analysis of PDE. It opens up a very useful strategy for constructing solutions u to an equation such as NLS, by first constructing an *approximate* solution $\tilde{u}$, for which $i\partial_t \tilde{u} + \frac{1}{2}\Delta\tilde{u}$ is very close to $\mu|\tilde{u}|^{p-1}\tilde{u}$ and $\tilde{u}(t_0) \approx u_0$. In other words, an approximate solution to NLS is nothing more than an exact solution to the perturbed NLS (3.52) for some small e and e_0. Stability theory can then let us pass from the approximate solution $\tilde{u}$ to a nearby exact solution u to the unperturbed NLS. This approach is quite powerful, because it is much easier to construct approximate solutions than exact solutions, for instance by asymptotic expansions[39], or by omitting certain terms from an equation that one believes to be negligible and then solving the reduced equation; see Section 3.8 below. To give another example, while the superposition of two solutions to a nonlinear equation will not in general yield another solution to that equation, such a superposition is often an *approximate* solution to the equation if the two component solutions are sufficiently "separated", either in space or frequency. This strategy can be used for instance to construct multisoliton solutions, and is also the main reason why the "induction on energy" strategy that we shall introduce in Chapter 5 is so powerful.

Fortunately, any equation with a good wellposedness theory is also likely to have a good stability theory, by modifying the arguments used to prove wellposedness suitably; indeed the stability theory is in many ways the culmination of the perturbation theory methods. The main trick (which we have already seen with the uniqueness and continuity theory) is to look at the *difference equation* between the approximate solution $\tilde{u}$ and the exact solution[40], and then solve this difference equation using the same types of techniques and estimates used for the wellposedness theory. Specifically, if we set $\tilde{u} = u + v$, then v solves the difference equation

$$i\partial_t v + \frac{1}{2}\Delta v = \mu(|u+v|^{p-1}(u+v) - |u|^{p-1}u) + e; \quad v(t_0) = e_0. \tag{3.53}$$

Thus the initial datum of v is small. As for the nonlinearity, we can use Taylor expansion to expand

$$\mu(|u+v|^{p-1}(u+v) - |u|^{p-1}u) + e = O(|u|^{p-1}|v|) + O(|u|^{p-2}|v|^2) + \ldots + O(|v|^p) + O(|e|).$$

assuming for simplicity that we are in the algebraic case where p is an odd integer. In practice, if e and e_0 are both small, then we expect v to be small also, and the dominant terms in the nonlinearity will be the terms $O(|u|^{p-1}|v|)$ which are linear in v. These terms can be dealt with for short times by iterative arguments based on the Duhamel formula, as well as estimates such as Strichartz estimates; for

[39]In particular, the theory of *nonlinear geometric optics* proceeds in this fashion, constructing solutions to an equation by first creating an ansatz consisting of an asymptotic series with certain amplitude and phase parameters. One then solves for these amplitudes and phases in order to make the partial sums of this series an approximate solution to the original equation, and then uses some stability theory to pass to an exact solution. These methods are very useful in constructing large classes of interesting solutions to many PDE, though they tend to require the initial data to be of a special form and are unsuited for the Cauchy problem with generic H^s_x initial data. Due to limitations of space we will not be able to discuss this important technique in this text.

[40]This assumes that the exact solution u exists for at least as long as the approximate solution u. In practice one can establish this by a continuity argument or by a suitable iteration of the wellposedness theory.

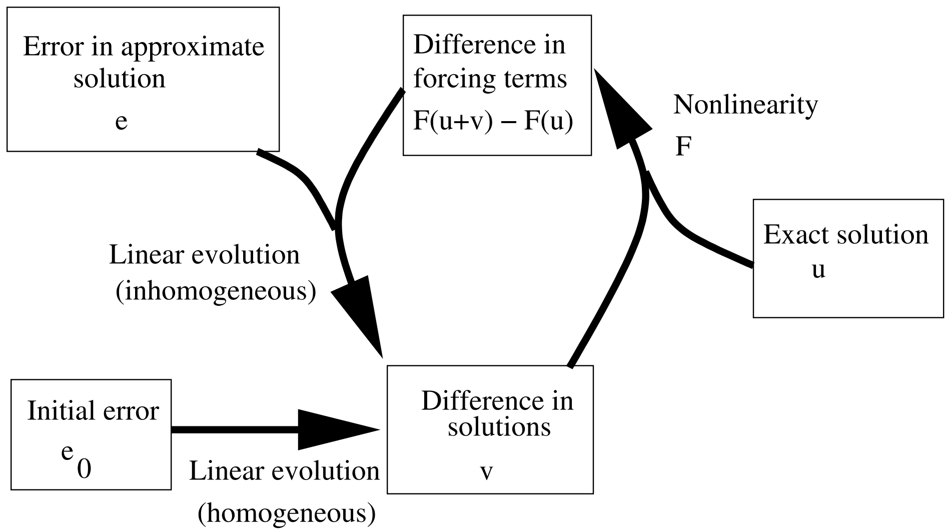

FIGURE 10. The difference scheme for an approximate perturbation $\tilde{u} = u + v$ to an exact solution u. This scheme can also be reversed to convert an approximate solution $\tilde{u}$ to an exact solution $u = \tilde{u} - v$ (by replacing $F(u+v) - F(u)$ with $F(\tilde{u}) - F(\tilde{u} - v)$).

longer times, one can use energy methods[41], combined with tools such as Gronwall's inequality, to try to keep control of the solution.

To illustrate the method, we shall consider asymptotics of one-dimensional defocusing NLS in the "short range" case $p > 3$. (The "critical range" case $p = 3$ and the "long range" case $p < 3$ are significantly more interesting, but also more difficult technically.) Applying the pseudoconformal transformation as in (3.16), we obtain the equation

$$i\partial_t v + \frac{1}{2}\partial_{xx} v = \frac{1}{t^{(5-p)/2}}|v|^{p-1}v \tag{3.54}$$

for $3 < p < 5$, which is obtained from the one-dimensional defocusing NLS via the pseudoconformal transformation (see (3.16)). To construct solutions near $t = 0$, we first omit the dispersive term $\frac{1}{2}\partial_{xx} v$ (using the intuition that this term will be dominated by the singular nonlinearity $\frac{1}{t^{(5-p)/2}}|v|^2 v$ for very small times t) and solve the simpler equation

$$i\partial_t \tilde{v} = \frac{1}{t^{(5-p)/2}}|\tilde{v}|^2\tilde{v}. \tag{3.55}$$

This equation just the ODE (3.17), and can be solved explicitly as

$$\tilde{v}(t,x) = \varepsilon e^{-i\frac{2}{p-3}\varepsilon^2|\psi(x)|^2 t^{(p-3)/2}}\psi(x) \tag{3.56}$$

[41]In some cases, when the exact solution u is an explicit form such as a soliton, one can use more advanced spectral analysis of the linearised equation $i\partial_t v + \frac{1}{2}\Delta v = O(|u|^{p-1}|v|)$ to obtain long-time control of the solution; this is an important tool in the theory of stability of solitons and multisolitons. However, such spectral methods are currently unavailable for more general classes u of solution.

for any complex-valued function $\psi(x)$, and $0 < \varepsilon < 1$ is a small parameter we have introduced to allow $\tilde{v}$ to be small (compare with (3.18)). Thus, if ε is small, we expect the original PDE (3.54) to have solutions which are approximately of the form (3.55). This can be established as follows.

PROPOSITION 3.34. *Let $\psi \in \mathcal{S}_x(\mathbf{R})$ and $0 < \varepsilon \ll 1$. If ε is sufficiently small depending on ψ, then we have a solution v to* (3.54) *on the slab* $(0,1) \times \mathbf{R}$ *obeying the bounds*

$$\|v(t) - \tilde{v}(t)\|_{H^1_x} \lesssim_\psi \varepsilon t$$

for all $0 < t < 1$, where $\tilde{v}$ was defined in (3.56).

PROOF. To construct v, we use the ansatz $v = \tilde{v} + w$. Subtracting (3.55) from (3.54), we see that w needs to solve the equation

$$i\partial_t w + \frac{1}{2}\partial_{xx} w = \frac{1}{t^{(5-p)/2}}(F(\tilde{v} + w) - F(\tilde{v})) - \partial_{xx}\tilde{v}$$

where $F(z) := |z|^{p-1}z$. We set initial datum $w(0) = 0$, and write the equation in integral form as $w = \Phi(w)$, where Φ is the nonlinear operator

$$\Phi(w) = \int_0^t \frac{1}{(t')^{(5-p)/2}}(F(\tilde{v}(t') + w(t')) - F(\tilde{v}(t'))) - \partial_{xx}\tilde{v}(t')\, dt'.$$

One can use energy estimates to verify that Φ is a contraction on the set $\{tu : \|u\|_{C^0_t H^1_x([0,1]\times\mathbf{R})} \lesssim_\psi \varepsilon\}$, if ε is sufficiently small depending on ψ; we leave this as an exercise. The claim now follows from the contraction mapping principle. $\square$

Informally, the above proposition gives the approximation

$$v(t,x) = \varepsilon e^{-i\frac{2}{p-3}\varepsilon^2|\psi(x)|^2 t^{(p-3)/2}}\psi(x) + l.o.t.$$

for $0 < t < 1$, where the lower order terms go to zero in a suitable sense as $t \to 0$. Inverting the pseudoconformal transformation, one obtains

$$u(t,x) = \varepsilon\frac{1}{(it)^{d/2}}\exp(\frac{i|x|^2}{2t} + \frac{2}{p-3}\varepsilon^2|\psi(x/t)|^2 t^{(p-3)/2})\overline{\psi(x/t)} + l.o.t. \tag{3.57}$$

for $1 < t < \infty$, where the lower order terms go to zero in a suitable sense as $t \to +\infty$. A similar argument applied to the linear Schrödinger equation (or using the fundamental solution) allows one to construct a solution u_{lin} to the linear equation with the asymptotics

$$u_{\mathrm{lin}}(t,x) = \varepsilon\frac{1}{(it)^{d/2}}\exp(\frac{i|x|^2}{2t})\overline{\psi(x/t)} + l.o.t.$$

Because we are in the short-range case $p > 3$, we can thus conclude that $u(t) - u_{\mathrm{lin}}(t)$ converges to zero in certain norms (for instance, it converges in $H^s_x(\mathbf{R})$ for any s). This suggests that the short-range case, one has scattering, at least for certain types of initial data. In the critical-range case $p = 3$ or the long-range case $p > 3$, it turns out that one can still construct solutions to NLS of the form (3.57); the arguments are similar though the singular nature of (3.54) now presents some delicate issues (cf. Exercise 1.19); see [**Oza**], [**GO**], [**HN**], [**CCT**] for some resolutions of this issue. These solutions fail to scatter to a solution to the linear Schrödinger equation in any H^s_x norm; thus long-range and critical-range equations do not exhibit scattering to the linear solution (this was first observed in [**Gla**]). However one can still hope to establish a modified scattering result, in which the approximating solution is not a linear solution, but rather a phase-shifted linear solution; see the above references.

Next, we illustrate how Gronwall type inequalities can be used to obtain stability for longer times than a simple iteration method (such as that given above) would give. The time interval on which one has non-trivial control is only extended by a logarithmic factor, but this is sometimes sufficient for applications. It would be of great interest to derive stability estimates on even longer intervals, perhaps by adapting the theory of Nekhoroshev stability from ODE, but this seems to be a difficult task (see [**BK**]).

PROPOSITION 3.35. *Let $\psi \in \mathcal{S}_x(\mathbf{R})$ and let $0 < \varepsilon \ll 1$ be a small number. Then there exists a time $T \sim_\psi \log^{1/3} \frac{1}{\varepsilon}$ and a strong H^1_x solution $u \in C^0_t H^1_x([0,T] \times \mathbf{R})$ to the small dispersion NLS*

$$i\partial_t u + \frac{\varepsilon^2}{2}\partial_{xx} u = |u|^2 u; \quad u(0) = \psi \tag{3.58}$$

such that $\|u - \tilde u\|_{C^0_t H^1_x([0,T]\times\mathbf{R})} \lesssim_\psi \varepsilon$, where

$$\tilde u(t,x) := e^{-i|\psi(x)|^2 t}\psi(x)$$

is the explicit solution to the ODE

$$i\partial_t \tilde u = |\tilde u|^2 \tilde u; \quad \tilde u(0) = \psi.$$

PROOF. From (a rescaled version of) Proposition 3.23 we know that a strong H^1_x solution u to (3.58) exists globally in time[42] Writing $u = \tilde u + w$, we see that w solves the equation

$$i\partial_t w + \varepsilon^2 \partial_{xx} w = (|(\tilde u + w)|^2(\tilde u + w) - |\tilde u|^2 \tilde u) - \frac{\varepsilon^2}{2}\partial_{xx}\tilde u; \quad w(0) = 0$$

which we write in Duhamel form as

$$w(t) = \int_0^t e^{i\varepsilon^2(t-t')\partial_{xx}}[(|(\tilde u + w)|^2(\tilde u + w) - |\tilde u|^2\tilde u) - \frac{\varepsilon^2}{2}\partial_{xx}\tilde u](t')\ dt'.$$

We take H^1_x norms of both sides, and use the fact that $e^{i\varepsilon^2(t-t')\partial_{xx}}$ is bounded in H^1_x, to obtain

$$\|w(t)\|_{H^1_x} \le \int_0^t \||(\tilde u + w)|^2(\tilde u + w)(t') - |\tilde u|^2\tilde u(t')\|_{H^1_x} + \frac{\varepsilon^2}{2}\|\partial_{xx}\tilde u(t')\|_{H^1_x}\ dt'.$$

A direct computation shows that

$$\|\tilde u(t')\|_{H^1_x} \lesssim_\psi \langle t'\rangle; \quad \|\partial_{xx}\tilde u(t')\|_{H^1_x} \lesssim_\psi \langle t'\rangle^3$$

while a computation using the algebra property of H^1_x (see Lemma A.8) gives

$$\||(\tilde u + w)|^2(\tilde u + w)(t') - |\tilde u|^2\tilde u(t')\|_{H^1_x} \lesssim \|w(t')\|_{H^1_x}(\|\tilde u(t')\|_{H^1_x} + \|w(t')\|_{H^1_x})^2.$$

Putting this all together, we obtain the bound

$$\|w(t)\|_{H^1_x} \lesssim_\psi \varepsilon^2\langle t\rangle^4 + \int_0^t \langle t\rangle^2 \|w(t')\|_{H^1_x} + \|w(t')\|^3_{H^1_x}\ dt'.$$

[42]This global wellposedness is convenient for the argument, but not absolutely necessary; the energy bounds we obtain in the proof, combined with the *local* H^1_x wellposedness theory, are sufficient (via a standard continuity argument) to construct the solution u on the given time interval $[0,T]$.

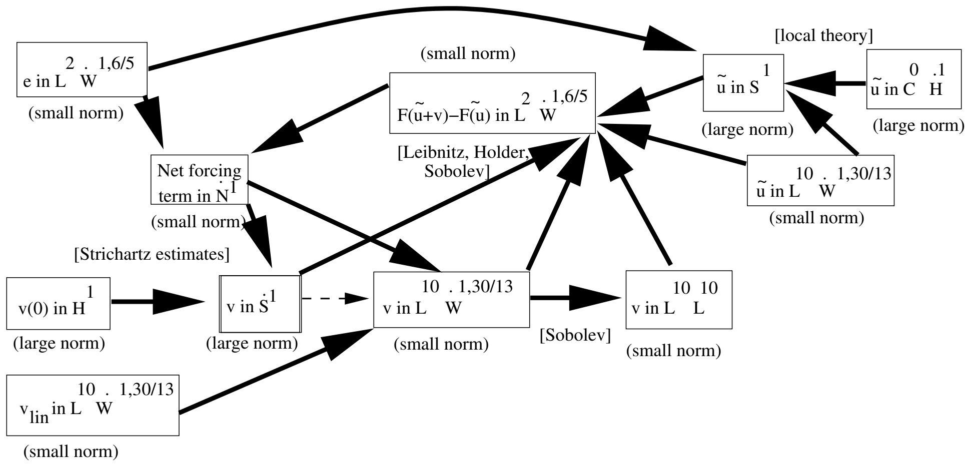

FIGURE 11. The scheme for estimating the difference $v = u - \tilde{u}$ in Lemma 3.36; it is thus a rather complex variation of the usual Strichartz iteration loop for this equation (see Figure 5).

If ε is sufficiently small depending on t, a continuity argument then gives

$$\|w(t)\|_{H^1_x} \lesssim_\psi \varepsilon^2 \langle t\rangle^4 \exp(C\langle t\rangle^3)$$

for all $0 < t \ll \log^{1/3} \frac{1}{\varepsilon}$ (cf. what one would obtain by Gronwall's inequality by dropping the nonlinear term $\|w(t')\|^3_{H^1_x}$), and the claim follows. $\square$

In the next section we will use this proposition to obtain some illposedness results for NLS.

Our final example of a stability theory result comes from the defocusing energy-critical three-dimensional NLS ($d = 3, p = 5, \mu = +1$). We shall show that H^1_x-wellposed solutions to this equation are stable as long as the $L^{10}_{t,x}$ norm stays bounded. We first state a preliminary result in which we assume that a certain spacetime norm on the solution is small.

LEMMA 3.36 (Short-time perturbations). [**CKSTT11**] *Let $d = 3, p = 5, \mu = +1$. Let I be a compact interval, and let $\tilde{u}$ be a field on $I \times \mathbf{R}^3$ which is a near-solution to NLS in the sense that*

$$(i\partial_t + \frac{1}{2}\Delta)\tilde{u} = |\tilde{u}|^4\tilde{u} + e \tag{3.59}$$

for some field e. Suppose that we also have the energy bound

$$\|\tilde{u}\|_{L^\infty_t \dot{H}^1_x(I\times\mathbf{R}^3)} \le E$$

for some $E > 0$. Let $t_0 \in I$, and let $u(t_0)$ be close to $\tilde{u}(t_0)$ in the sense that

$$\|u(t_0) - \tilde{u}(t_0)\|_{\dot{H}^1_x} \le E' \tag{3.60}$$

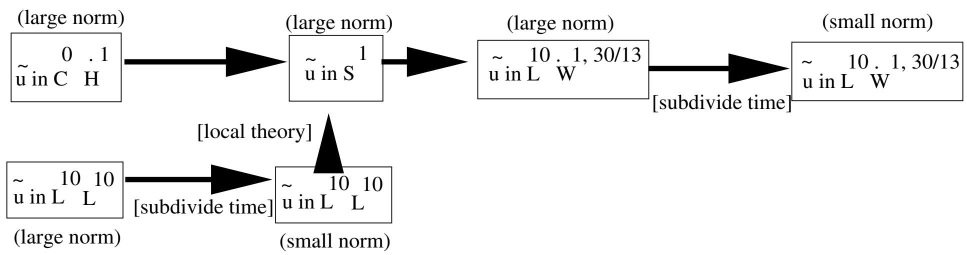

FIGURE 12. The additional time decompositions necessary in order to deduce Lemma 3.37 from Lemma 3.36.

for some $E' > 0$. *Assume also that we have the smallness conditions*

$$\|\tilde{u}\|_{L_t^{10}\dot{W}_x^{1,30/13}(I\times\mathbf{R}^3)} \leq \epsilon_0 \tag{3.61}$$

$$\|e^{i(t-t_0)\Delta/2}(u(t_0) - \tilde{u}(t_0))\|_{L_t^{10}\dot{W}_x^{1,30/13}(I\times\mathbf{R}^3)} \leq \varepsilon \tag{3.62}$$

$$\|e\|_{L_t^2\dot{W}_x^{1,6/5}} \leq \varepsilon \tag{3.63}$$

for some $0 < \varepsilon < \epsilon_0$, *where* ϵ_0 *is some constant* $\epsilon_0 = \epsilon_0(E, E') > 0$.

We conclude that there exists a solution u *to* (3.1) *on* $I \times \mathbf{R}^3$ *with the specified initial datum* $u(t_0)$ *at* t_0 *obeying the bounds*

$$\|u - \tilde{u}\|_{\dot{S}^1(I\times\mathbf{R}^3)} \lesssim E' \tag{3.64}$$

$$\|u\|_{\dot{S}^1(I\times\mathbf{R}^3)} \lesssim E' + E \tag{3.65}$$

$$\|u - \tilde{u}\|_{L^{10}_{t,x}(I\times\mathbf{R}^3)} \lesssim \|u - \tilde{u}\|_{L_t^{10}\dot{W}_x^{1,30/13}(I\times\mathbf{R}^3)} \lesssim \varepsilon \tag{3.66}$$

$$\|(i\partial_t + \Delta)(u - \tilde{u})\|_{L_t^2\dot{W}_x^{1,6/5}(I\times\mathbf{R}^3)} \lesssim \varepsilon. \tag{3.67}$$

Note that $u(t_0) - \tilde{u}(t_0)$ is allowed to have large energy, albeit at the cost of forcing ε to be smaller, and worsening the bounds in (3.64). From Strichartz estimates and (3.60) we see that the hypothesis (3.62) is redundant if one is willing to take $E' = O(\varepsilon)$.

We leave the proof of Lemma 3.36 to the exercises. One can amplify this lemma to deal with the more general situation of near-solutions with finite but arbitrarily large $L^{10}_{t,x}$ norms.

LEMMA 3.37 (Long-time perturbations). [**CKSTT11**] *Let* $d = 3, p = 5, \mu = +1$. *Let* I *be a compact interval, and let* $\tilde{u}$ *be a field on* $I \times \mathbf{R}^3$ *which obeys the bounds*

$$\|\tilde{u}\|_{L^{10}_{t,x}(I\times\mathbf{R}^3)} \leq M \tag{3.68}$$

and

$$\|\tilde{u}\|_{L_t^\infty\dot{H}_x^1(I\times\mathbf{R}^3)} \leq E \tag{3.69}$$

for some $M, E > 0$. *Suppose also that* $\tilde{u}$ *is a near-solution to NLS in the sense that it solves* (3.59) *for some* e. *Let* $t_0 \in I$, *and let* $u(t_0)$ *be close to* $\tilde{u}(t_0)$ *in the sense that*

$$\|u(t_0) - \tilde{u}(t_0)\|_{\dot{H}^1_x} \leq E'$$

for some $E' > 0$. Assume also that we have the smallness conditions,

$$\|e^{i(t-t_0)\Delta/2}(u(t_0) - \tilde{u}(t_0))\|_{L_t^{10}\dot{W}_x^{1,30/13}(I\times\mathbf{R}^3)} \leq \varepsilon \tag{3.70}$$
$$\|e\|_{L_t^2\dot{W}_x^{1,6/5}(I\times\mathbf{R}^3)} \leq \varepsilon$$

for some $0 < \varepsilon < \varepsilon_1$, where ε_1 is a small constant $\varepsilon_1 = \varepsilon_1(E, E', M) > 0$. We conclude there exists a solution u to (3.1) on $I \times \mathbf{R}^3$ with the specified initial datum $u(t_0)$ at t_0, and furthermore

$$\|u - \tilde{u}\|_{\dot{S}^1(I\times\mathbf{R}^3)} \lesssim_{M,E,E'} 1$$
$$\|u\|_{\dot{S}^1(I\times\mathbf{R}^3)} \lesssim_{M,E,E'} 1$$
$$\|u - \tilde{u}\|_{L_{t,x}^{10}(I\times\mathbf{R}^3)} \lesssim \|u - \tilde{u}\|_{L_t^{10}\dot{W}_x^{1,30/13}(I\times\mathbf{R}^3)} \lesssim_{M,E,E'} \varepsilon.$$

Again, we leave the details to the exercises. This stability lemma is quite powerful; it shows that approximate solutions can be adjusted to become exact solutions even when the energy of both initial data and their difference are large, as long as the approximate solution is bounded (but not necessarily small) in $L_{t,x}^{10}$ norm, and the error e is very small. It will play an important role in the large energy theory of this equation in Chapter 5.

In the preceding examples of stability theory, we approximated an exact solution u by an explicit approximate solution $\tilde{u}$. In some cases, most notably in the stability theory of solitons and multisolitons, it is better to approximate u by a *partially explicit* approximate solution, which involves some free parameters that one has some freedom to choose in order to make the analysis of the error terms as easy as possible. For instance, if considering perturbations u of a soliton solution such as $e^{it\tau}Q(x)$, the ansatz $u = e^{it\tau}Q(x)+w$ turns out to not be very effective (the bounds on w will grow exponentially in time if one applies perturbation theory naively). Instead, a better procedure is to perform an ansatz $u = e^{it\tau+\theta(t)}Q(x - x(t)) + w$, where $\theta : \mathbf{R} \to \mathbf{R}$ and $x : \mathbf{R} \to \mathbf{R}^d$ are parameters that one can choose. Typically, one chooses these parameters in order to obtain some moment conditions on w (for instance, one could try to force w to be orthogonal to functions such as iQ or ∇Q), which can improve the behaviour of the equation for w (by eliminating some degeneracies in the linearised operator associated to Q). This reflects the fact that perturbations to a soliton can cause that soliton to move in a significant manner along the directions given by the symmetries of the equation, namely phase rotation and spatial translation; these are major channels of propagation for the equation as motion in these directions does not conflict with any of the conservation laws. (In the case of the L_x^2-critical equation, motion in the scaling direction is also possible as it does not contradict conservation of mass.) We will not have space to devote attention to these tools, which are fundamental in the stability theory of solitons, but see [**Wei2**] and many subsequent papers (e.g. [**MR**], [**MR2**], [**MR3**] and the references therein).

EXERCISE 3.55 (Justification of energy conservation). Let $d = 3$ and $1 < p < 5$, $\mu = +1$, $t_0 = 0$, and $u_0 \in H_x^1(\mathbf{R}^d)$. For each $\varepsilon > 0$, show that there exists a global H_x^1-wellposed solution $u^{(\varepsilon)}$ solution to the regularised NLS

$$i\partial_t u^{(\varepsilon)} + \frac{1}{2}\Delta u^{(\varepsilon)} = (|u^{(\varepsilon)}|^2 + \varepsilon^2)^{(p-1)/2}u^{(\varepsilon)}; \quad u^{(\varepsilon)}(0) = u_0$$

with a conserved energy

$$E^{(\varepsilon)}[u^{(\varepsilon)}(t)] := \int_{\mathbf{R}^d} \frac{1}{2}|\nabla u^{(\varepsilon)}|^2 + \frac{2}{p+1}(|u^{(\varepsilon)}|^2 + \varepsilon^2)^{(p+1)/2}\ dx.$$

Then show that for any compact time interval I containing 0, $u^{(\varepsilon)}$ converges in $S^1(I \times \mathbf{R}^3)$ to a strong H^1_x solution $u \in S^1(I \times \mathbf{R}^3)$ to (3.1) with the conserved energy

$$E[u(t)] := \int_{\mathbf{R}^d} \frac{1}{2}|\nabla u^{(\varepsilon)}|^2 + \frac{2}{p+1}|u|^{p+1}\ dx.$$

This is one way in which to justify the conservation of energy for fractional-power NLS.

EXERCISE 3.56 (Weak solutions). Let $d = 3$ and $1 < p < 6$, $\mu = +1$, $t_0 = 0$, and $u_0 \in H^1_x(\mathbf{R}^d)$. Show that for any $\lambda > 0$ there exists a global H^1_x-wellposed solution $u^{(\lambda)}$ to the tempered NLS

$$i\partial_t u^{(\lambda)} + \frac{1}{2}\Delta u^{(\lambda)} = \max(|u^{(\lambda)}|^{p-1}, \lambda|u^{(\lambda)}|^4)u^{(\lambda)}; \quad u^{(\lambda)}(0) = u_0$$

with a conserved mass $\int_{\mathbf{R}^d} |u^{(\lambda)}|^2\ dx$ and conserved energy

$$E^{(\lambda)}[u^{(\lambda)}(t)] := \int_{\mathbf{R}^d} \frac{1}{2}|\nabla u^{(\lambda)}|^2 + V_\lambda(|u^{(\lambda)}|)\ dx,$$

where $V_\lambda(y) := \int_0^y \max(w^p, \lambda w^5)\ dw$. Using weak compactness, show that there exists a sequence $\lambda_n \to \infty$ such that the solutions $u^{(\lambda_n)}$ converges weakly in $L^\infty_t H^1_x(\mathbf{R} \times \mathbf{R}^3)$ to a global *weak* H^1_x solution $u \in L^\infty_t H^1_x(\mathbf{R} \times \mathbf{R}^3)$ to the NLS (3.1). Thus for certain supercritical equations it is still possible to construct global weak solutions. Existence or uniqueness of global strong H^1_x solutions for these equations is a major unsolved problem (sharing many difficulties with the notorious global regularity problem for Navier-Stokes). Even energy conservation for the global weak solution is not known (the above construction, combined with Fatou's lemma, only shows that the energy at time t is less than or equal to the energy at time 0). The analogous construction for global weak solutions for NLW dates back to [**Seg2**].

EXERCISE 3.57. Complete the proof of Proposition 3.34. (Hint: use energy estimates as in the proof of Proposition 3.8).

EXERCISE 3.58. [**CKSTT11**] Prove Lemma 3.36. (Hint: first establish $L^{10}_{t,x}$ and $\dot{S}^1$ control on $\tilde{u}$, then write $v := \tilde{u} - u$ and $S := \|(i\partial_t + \Delta)v\|_{L^2_t \dot{W}^{1,6/5}_x(I\times\mathbf{R}^3)}$, and use the Leibnitz rule, Hölder's inequality, Strichartz, and Sobolev to bound S in terms of itself and ε. Then use a continuity method argument to obtain an unconditional bound on S. See also Figure 11.)

EXERCISE 3.59. [**CKSTT11**] Prove Lemma 3.37. (Hint: first establish $\dot{S}^1$ control on $\tilde{u}$. Then divide up I into intervals where the $L^{10}_t L^{30/13}_x$ norm of ∇u is small, and apply Lemma 3.36 inductively on these intervals. See also Figure 12.)

EXERCISE 3.60. By refining the analysis used in the proof, replace the $\log^{1/3} \frac{1}{\varepsilon}$ in Proposition 3.35 with $\log \frac{1}{\varepsilon}$.

EXERCISE 3.61. [**CKSTT13**] Let $u \in C^0_{t,\mathrm{loc}} \mathcal{S}_x(\mathbf{R} \times \mathbf{T}^2)$ be a classical solution to the cubic defocusing NLS $i\partial_t u + \frac{1}{2}\Delta u = |u|^2 u$. Using the Fourier ansatz $u(t,x) = \sum_{k \in (2\pi\mathbf{Z})^2} e^{i(k\cdot x + \frac{1}{2}|k|^2 t)} a_k(t)$, deduce the infinite system of ODE

$$\partial_t a(t) = \mathcal{N}_t(a(t), a(t), a(t)) \tag{3.71}$$

where $a = (a_k)_{k \in (2\pi\mathbf{Z})^2}$ and $\mathcal{N}_t$ is the trilinear form

$$\mathcal{N}_t(a,b,c)_k := \sum_{k_1,k_2,k_3 \in (2\pi\mathbf{Z})^2 : k_1 - k_2 + k_3 = k} a_{k_1} \overline{b_{k_2}} c_{k_3} e^{\frac{i}{2}(|k_1|^2 - |k_2|^2 + |k_3|^2 - |k_4|^2)t}.$$

(Compare with (1.56)). Let $K \gg 1$ be a large number, let $0 < \sigma < 1$, and let $T \leq c(\sigma) K^2 \log K$ for some small $c(\sigma) > 0$ depending only on σ. Suppose we have a system $b(t) = (b_k(t))_{k \in (2\pi\mathbf{Z})^2}$ of functions with $b \in C^1_t l^1_k([0,T] \times (2\pi\mathbf{Z})^2)$ with $b(0) = a(0)$ which obeys the approximate equation

$$\partial_t b(t) = \mathcal{N}_t(b(t), b(t), b(t)) + e(t)$$

to (3.71), where $e(t)$ and $b(t)$ obey the l^1 bounds

$$\|b\|_{C^1_t l^1_k([0,T] \times (2\pi\mathbf{Z})^2)} \lesssim K^{-1}; \quad \sup_{0 \leq t \leq T} \int_0^t \| \int_0^t e(t')\, dt' \|_{l^1_k} \lesssim K^{-1-\sigma}.$$

Then if $c(\sigma)$ is sufficiently small depending on σ, we have the estimate $\|a - b\|_{C^1_t l^1_k([0,T] \times (2\pi\mathbf{Z})^2)} \lesssim K^{-1-\sigma/2}$. This lemma allows one to use near-solutions to NLS in Fourier space to approximate actual solutions to NLS, and is a key ingredient in establishing a certain weak turbulence result for this equation. See [**CKSTT13**].

3.8. Illposedness results

All happy families resemble one another; each unhappy family is unhappy in its own way. (Leo Tolstoy, "Anna Karénina")

In the past few sections we have developed a wellposedness theory for several types of NLS and NLW equations, for various regularities H^s_x (or $H^s_x \times H^{s-1}_x$). Despite the wide variety of equations and regularities considered, the wellposedness theory for these equations are remarkably similar to each other, especially for subcritical regularities. In such cases the time of existence depends only on the norm of the data, and the solution map not only exists and is unique, but enjoys very strong continuity properties; indeed, the solution map (from H^s_x to $C^0_t H^s_x$) is typically uniformly continuous, Lipschitz, infinitely differentiable, and even real analytic (see for instance Exercise 3.25).

However, there are certain equations and certain regularities for which the Cauchy problem does not agree with this picture, either locally or globally in time, in which case we say that that particular Cauchy problem is *illposed*. Unlike the situation with wellposedness, the type of illposedness exhibited can vary substantially on the equation and on the regularity. At one extreme, there are very dramatic examples of illposedness, such as *blowup* - various norms going to infinity in finite time - beyond which no reasonably strong notion of solution can be salvaged. At the other extreme there are very mild examples of illposedness, where it may still be that the solution map exists and could even be continuous, but that the solution map is known to be unstable (e.g. non-uniformly-continuous or non-Lipschitz), non-differentiable, or at least non-analytic. Intermediate between these extremes[43] are examples of *norm explosion* - when data of arbitrarily small norm can lead to solutions of arbitrarily large norm in arbitrarily small time. This is not quite as dramatic as blowup, because a solution may still exist for each given initial datum, but it certainly does prevent any continuous dependence of the solution map on the initial data.

For each of the types of illposedness discussed above, there are examples of equations and regularities that exhibit that illposedness. In contrast with the wellposedness theory, which is largely based around the single technique of Duhamel iteration, illposedness can be achieved by a surprisingly large number of unrelated methods. We will not be able to discuss all of them here, but we give a representative sample. For a recent survey of techniques and results, see [**Tzv**].

We first discuss methods for generating blowup, by which we mean classical (or strong) solutions which develop a significant singularity in finite time (e.g. the H^s_x norm goes to infinity in finite time). One way to construct these solutions is via construction of explicit (or nearly explicit) blowup solutions. We have already seen two examples of this - the blowup solution (3.15) for the pseudoconformal focusing NLS and the ODE-based blowup solution (3.6) for the focusing NLW. The latter solution has no decay in space and thus does not lie in any $H^s_x \times H^{s-1}_x$ spaces, however this can be rectified by a finite speed of propagation; see Exercise 3.9.

In some cases, one cannot construct a blowup solution explicitly, but can create an explicit *approximate* solution to the equation which blows up in finite time. One can then hope to use perturbation theory to convert this to an exact blowup solution. This argument can be made to work, but is extremely delicate, because perturbation theory requires a great deal of wellposedness and stability on the equation, which is in obvious conflict with our need to make both the exact and approximate solution to blow up in finite time. One often needs to carefully renormalise the solution (usually via rescaling), and obtain stability control in one set of norms while obtaining blowup in another. See for instance [**Mer**], [**BW**] for some instances of this approach.

In the case of the NLS, there is another, much more indirect, way to force blowup of a solution, namely the *virial argument* of Glassey [**Gla2**], based on the nonlinear counterpart to (2.38). For simplicity let us consider a classical solution $u \in C^\infty_{t,\mathrm{loc}}\mathcal{S}_x(\mathbf{R} \times \mathbf{R}^d)$ to an algebraic NLS. Consider the quantity

$$V(t) := \int_{\mathbf{R}^d} |x|^2 \mathrm{T}_{00}(t,x)\ dx = \int_{\mathbf{R}^d} |x|^2 |u(t,x)|^2\ dx.$$

[43]There are several other "symptoms" of illposedness which we will not have space to discuss here, including breakdown of uniqueness (either for weak or strong solutions); failure of mass or energy conservation; loss of regularity; or examples of *approximate* solutions to the equation which blowup in finite time. The reader is invited to try to list several such illposedness phenomena and rank them in approximate order of severity.

This quantity is clearly non-negative. Applying (2.35) and integration by parts repeatedly, we obtain the *virial identity*

$$
\begin{aligned}
\partial_{tt} V(t) &= 2 \int_{\mathbf{R}^d} \mathrm{T}_{jj}(t,x)\ dx \\
&= \int_{\mathbf{R}^d} 2|\nabla u|^2 + \frac{\mu d(p-1)}{p+1} |u|^{p+1}\ dx \\
&= 4E[u] + \frac{\mu d(p - p_{L^2_x})}{p+1} \int_{\mathbf{R}^d} |u(t,x)|^{p+1}\ dx
\end{aligned}
\tag{3.72}
$$

where $E[u]$ is the conserved energy and $p_{L^2_x} := 1 + \frac{4}{2}$ is the pseudoconformal power. If we are in the L^2_x-critical or L^2_x-supercritical focusing cases $p \geq p_{L^2_x}$, $\mu = -1$, we thus conclude the bound

$$\partial_{tt} V(t) \leq 4E[u].$$

If the energy happens to be negative (which is possible in the focusing case $\mu = -1$), this shows that V is a strictly concave function of t. Since V is also non-negative, we conclude that the solution can only exist classically for a finite amount of time (in either direction). This argument thus demonstrates blowup in finite time (and even gives an upper bound on the time of existence in terms of the datum and the energy). It can be extended to demonstrate blowup for any H^1_x initial data u_0 which has negative energy[44] and obeys the decay condition $\langle x \rangle u_0 \in L^2_x$. The decay condition can be removed, basically by working with spatially truncated versions of the virial identity; see for instance [**Naw**]. We remark that while negative energy is a sufficient condition for blowup, it is hardly a necessary condition; for instance, the solution (3.15) has zero energy, and the solutions constructed in [**BW**] have positive energy. The blowup phenomenon has been analyzed much further for the L^2_x-critical equation, in the vicinity of soliton solutions; see [**MR**], [**MR2**], [**MR3**].

For the focusing NLW, one can also exploit some positivity properties of the fundamental solution to establish successively stronger lower bounds on a solution which eventually leads to blowup. One particularly striking example of this is a result of John [**Joh**], who showed that for the three-dimensional focusing NLW with initial data $u(0,x) = \varepsilon u_0(x)$, $\partial_t u(0,x) = 0$ for positive Schwartz u_0 and sufficiently small ε, one has blowup in finite time for $p < \sqrt{2}$ and global existence for $p > \sqrt{2}$. These results have since been generalised extensively; see for instance [**GLS**], [**Hor**].

Once one has one solution blowing up, one can use the symmetries of the equation to generate further solutions blowing up. When the regularity s is low enough, one can use the symmetries to create classical initial data of arbitrarily small H^s_x norm which blow up in arbitrarily small time, which is a very strong demonstration of illposedness in that data class H^s_x; we give some examples in the exercises.

All the known examples of blowup from classical data are for focusing equations; for many defocusing equations (e.g. H^1_x-subcritical or H^1_x-critical defocusing NLS or NLW) we have global existence of classical solutions. The question of whether blowup occurs from classical data for H^1_x-supercritical defocusing NLS or NLW equations is a major open problem in the subject (analogous to the Navier-Stokes global regularity problem) and remains very far from resolution. While blowup

[44] To put it another way, whenever the nonlinear component of the energy exceeds the linear component, blowup occurs. Compare this with Principle 3.24.

is not known for these equations, we can in many cases establish weaker forms of illposedness, which are not as dramatic as blowup but do indicate that many of the techniques discussed in earlier sections to establish wellposedness (e.g. iteration methods) must necessarily fail. One of the mildest types of illposedness of this form is that of *analytic illposedness*, in which one demonstrates that the solution map (say from H^s_x to $C^0_t H^s_x$), if it exists at all, is not real analytic. In fact one typically shows the stronger statement of C^k *illposedness* for some $k \geq 1$, which asserts that the solution map, if it exists, is not k-times differentiable. This is basically accomplished by the method of Taylor expansions (i.e. power series methods). Let us illustrate this with the three-dimensional cubic defocusing NLS ($d = 3, p = 3, \mu = +1$) with initial datum $u(0) = \varepsilon u_0$ for some fixed Schwartz u_0, thus we are considering solutions $u^{(\varepsilon)}$ to the Cauchy problem

$$i\partial_t u^{(\varepsilon)} + \frac{1}{2}\Delta u^{(\varepsilon)} = |u^{(\varepsilon)}|^2 u^{(\varepsilon)}; \quad u(0) = \varepsilon u_0. \tag{3.73}$$

The global existence theory of this equation (Exercise 3.38) guarantees that the solutions $u^{(\varepsilon)}$ exist and are smooth for all time. A refinement of this theory also shows us that $u^{(\varepsilon)}$ also depend smoothly on ε, uniformly on any compact time interval. In particular, we can obtain a Taylor expansion

$$u^{(\varepsilon)}(t,x) = \varepsilon u_1(t,x) + \varepsilon^2 u_2(t,x) + \varepsilon^3 u_3(t,x) + O(\varepsilon^4)$$

for some smooth functions u_1, u_2, u_3 (there is no zeroth order term since $u^{(0)}$ is clearly zero), where the error is uniformly smooth in t, x on any compact time interval. We can expand both sides of (3.73) using this expansion and compare coefficients. One learns that the first coefficient u_1 is just the linear solution:

$$i\partial_t u_1 + \frac{1}{2}\Delta u_1 = 0; \quad u_1(0) = u_0$$

or in other words $u_1(t) = e^{it\Delta/2}u_0$. The second term u_2 solves the equation

$$i\partial_t u_2 + \frac{1}{2}\Delta u_2 = 0; \quad u_2(0) = 0$$

and is hence zero. The third term u_3 solves the equation

$$i\partial_t u_3 + \frac{1}{2}\Delta u_3 = |u_1|^2 u_1; \quad u_3(0) = 0$$

and is hence given by a Duhamel integral

$$u_3(t) = -i\int_0^t e^{i(t-t')\Delta/2}(|u_1|^2 u_1(t'))\, dt' = -i\int_0^t e^{i(t-t')\Delta/2}(|e^{it'\Delta/2}u_0|^2 e^{it'\Delta/2}u_0)\, dt'.$$

From the Taylor expansion of $u^{(\varepsilon)}$ we thus have obtained the formula

$$\frac{d^3}{d\varepsilon^3}u^{(\varepsilon)}(t)|_{\varepsilon=0} = -3!i\int_0^t e^{i(t-t')\Delta/2}(|e^{it'\Delta/2}u_0|^2 e^{it'\Delta/2}u_0)\, dt'.$$

This shows that if the map $u_0 \mapsto \int_0^t e^{i(t-t')\Delta/2}(|e^{it'\Delta/2}u_0|^2 e^{it'\Delta/2}u_0)\, dt'$ is not a bounded map from $H^s_x(\mathbf{R}^3)$ to $C^0_t H^s_x([0,T] \times \mathbf{R}^3)$, then the solution map $u_0 \mapsto u$ will not be a C^3_{loc} map from $H^s_x(\mathbf{R}^3)$ to $C^0_t H^s_x([0,T] \times \mathbf{R}^3)$, even for data arbitrarily close to zero in H^s_x norm. This lack of boundedness can often be established by direct computation; in this case, we can achieve this for supercritical regularities $s < s_c = \frac{1}{2}$ (in contrast to the critical case $s = s_c$ and subcritical cases $s > s_c$, in which one does have analytic wellposedness); see Exercise 3.65.

The method above is fairly general. Roughly speaking, it shows that if the Duhamel iteration scheme used to construct solutions leaves a certain space X after finitely many iterations, then the solution map can only have a finite amount of differentiability in that space. This is not too surprising since the iteration scheme is closely akin to a power series expansion of the solution in terms of the initial datum.

Another approach for establishing illposedness is by constructing families of exact solutions to the equation which are close together at time zero but far apart at other times. In some cases one can use explicit solutions such as solitons and plane wave solutions, possibly after various symmetries of the equation have been applied; in other cases one needs to construct solutions by the methods of nonlinear geometric optics, or more generally by constructing an approximate solution first and then using stability theory to perturb the approximate solution into the exact solution. A typical result obtained by this method would be that a certain solution operator cannot be uniformly continuous from H^s_x to $C^0_t H^s_x$ even when the size of the datum and time of existence are set to be small. We have already seen some examples of this in Exercise 3.5 and the discussion after (3.20), using explicit solutions. We now briefly sketch how to achieve a similar effect using the approximate solutions of the preceding section. For sake of concreteness let us just consider the one-dimensional defocusing cubic NLS ($d = 1, p = 3, \mu = +1$). Let ψ be a Schwartz function. From Proposition 3.35 we have constructed (for small $\varepsilon > 0$ and $1 \leq a \leq 2$) solutions $w_{\varepsilon,a}$ to the small-dispersion equation $i\partial_t w_\varepsilon + \frac{\varepsilon^2}{2}\partial_{xx} w_\varepsilon = |v|^2 w_\varepsilon$ on the slab $[0,1] \times \mathbf{R}$ which has the approximate form

$$w_{\varepsilon,a}(t,x) = ae^{-ia^2|\psi(x)|^2 t}\psi(x) + O_\psi(\varepsilon)$$

for $0 \leq t \leq 1$, where the error can be controlled in a suitable H^1_x sense. One can apply the rescaling $u_{\varepsilon,a}(t,x) := w_{\varepsilon,a}(t,\varepsilon x)$ to obtain a class of solutions u_ε to the original NLS with the approximate form

$$u_{\varepsilon,a}(t,x) = ae^{-ia^2|\psi(\varepsilon x)|^2 t}\psi(\varepsilon x) + O_\psi(\varepsilon)$$

where the error is now controlled in some rescaled H^1_x sense. One can exploit scale invariance (3.9) and Galilean invariance (3.10) to obtain a wider class of exact solutions $u_{\varepsilon,a,\lambda,v}$ to NLS for $\lambda > 0$ and $v \in \mathbf{R}$ of the form

$$u_{\varepsilon,a,\lambda,v}(t,x) = \lambda^{-2/(p-1)} ae^{i(x\cdot v + \frac{t|v|^2}{2} - a^2|\psi(\varepsilon(x-vt)/\lambda)|^2 t/\lambda^2)}\psi(\varepsilon(x-vt)/\lambda) + O_\psi(\varepsilon\lambda^{-2/(p-1)})$$

where the error has to be interpreted in a suitable norm. If $s < 0$ is a negative regularity, then by making v large, and setting $\lambda \sim_\varepsilon |v|^{-(p-1)s/2}$, these solutions can become bounded in H^s_x. By a suitable variation of the parameters $\varepsilon, a, \lambda, v$ one can then show that the solution operator to this equation cannot be uniformly continuous from H^s_x to $C^0_t H^s_x$, even for small times and small norm, by exploiting the phase decoherence effect arising from the $a^2|\psi(\varepsilon(x-vt)/\lambda)|^2 t/\lambda^2$ term; see [**CCT2**] for details. Generally speaking, it is not difficult to create (for any equation and regularity) *large* data solutions which exhibit these types of instabilities at *large* times; the various symmetries are then used to create small data solutions which are similarly unstable at small times. In order for this to work, one needs the regularity s to be supercritical with respect to at least one of the symmetries (scaling, Galilean, or Lorentz). See [**Kuk**], [**Leb**], [**BGT**], [**CCT**], [**CCT2**], [**CCT3**] for several examples of this technique and further discussion.

One final type of illposedness is the *high-to-low frequency cascade*, in which a solution starts off initially with Fourier transform supported primarily at high frequencies, but quickly creates a substantial presence at low frequencies. For small s (e.g. negative s), such solutions typically have small H^s_x norm at time zero but large H^s_x norm at later times; this *norm explosion* is a fairly strong form of illposedness as it shows that the solution operator, if it exists at all, has a severe singularity in H^s_x near the zero solution. These cascading solutions can often be constructed using the stability theory arising from a higher regularity $H^{s'}_x$. See [**CCT2**], [**BT**] for examples of this strategy.

EXERCISE 3.62. Let u be a classical solution to an NLW, and let

$$V(t) := \int_{\mathbf{R}^d} |x|^2 \mathrm{T}_{00}(t,x) - \frac{d-1}{2}|u|^2\ dx.$$

Establish the following analogue of the virial identity for this quantity, namely

$$\partial_{tt} V(t) = 2E[u] + \frac{\mu(d-1)(p - p_{H^{1/2}_x})}{p+1} \int_{\mathbf{R}^d} |u(t,x)|^{p+1}\ dx$$

where $p_{H^{1/2}_x} := 1 + \frac{4}{d-1}$ is the conformal power. (Note the shifting of the dimension d by one; compare this with Exercise 3.2.) This identity is not as useful as the NLS virial identity because the quantity V does not have a definite sign in general.

EXERCISE 3.63. Consider a focusing NLS with $p \geq p_{L^2_x} = 1 + \frac{4}{d}$, and let $s < s_c$. Show that there exists classical data of arbitrarily small H^s_x norm such that the solution to the NLS blows up in arbitrarily small time. (Hint: use the virial identity to create a classical solution with Schwartz initial data which blows up in finite time, and then use the scaling symmetry to rescale the blowup time to be arbitrarily small.) This illustrates the principle that one usually does not have a wellposedness theory at supercritical regularities for focusing equations.

EXERCISE 3.64. Consider a focusing NLW and let $s < s_c$. Show that there exists classical data of arbitrarily small $H^s_x \times H^{s-1}_x$ norm such that the solution to the NLW blows up in arbitrarily small time. (Hint: use Exercise 3.9. In the case when s_c is negative, you may find it convenient to enforce moment conditions on the initial data to ensure some vanishing of the Fourier coefficients near the origin.) A similar result holds for $s < s_l$ (using the Lorentz invariance instead of the scaling invariance) but is a little trickier; see [**Sog**].

EXERCISE 3.65. Let $T > 0$ be arbitrary. Use a scaling argument to show that the map $u_0 \mapsto \int_0^t e^{i(t-t')\Delta/2}(|e^{it'\Delta/2}u_0|^2 e^{it'\Delta/2}u_0)\ dt'$ is not a bounded map from H^s_x to $C^0_t H^s_x([0,T] \times \mathbf{R}^3)$ when $s < 1/2$. Conversely, use Strichartz estimates to show that this map is bounded for $s \geq 1/2$. (For sake of this exercise, you may use the heuristic (A.15) as if it were rigorous. Alternatively, one may use Littlewood-Paley decomposition.)

EXERCISE 3.66. [**Kat4**] Consider the rather artificial nonlinear wave equation $\Box u = -|u|^p$ for some H^1_x-subcritical power $1 \leq p < 1 + \frac{4}{d-2}$. Let u be a strong $H^1_x \times L^2_x$ solution to this equation whose initial position is supported in the ball $\{|x| \leq 1\}$ and whose initial velocity is zero (for simplicity). It is possible to establish the finite speed of propagation property for such solutions, in particular you may assume that this solution is supported on the ball $\{|x| \leq 1 + t\}$ for all later times

$t > 0$ for which the solution exists. Show that if the integral $\int_{\mathbf{R}^d} u(0,x)\, dx$ is sufficiently large and positive depending on d and p, then the solution u can only exist for a finite amount of time in the forward direction (i.e. u cannot be a strong solution on $[0,+\infty)$). If $p < 1 + \frac{2}{d}$, show that one only needs the integral $\int_{\mathbf{R}^d} u(0,x)\, dx$ to be strictly positive to achieve the same result (i.e. no largeness hypothesis is required). Hints: obtain an integral inequality for the quantity $m(t) := \int_{\mathbf{R}^d} u(t,x)\, dx$, using Hölder's inequality and finite speed of propagation. First show that $m(t)$ is convex and monotone increasing, and then obtain even better lower bounds on this quantity. You may find the comparison principle, Exercise 1.7, to be useful. For further variations on this theme, see [**KTao3**].

EXERCISE 3.67. [**LSog**], [**Sog**] Consider a focusing NLW, and let s be such that $0 < s < s_l := \frac{d+1}{4} - \frac{1}{p-1}$ (so s is supercritical with respect to the Lorentz invariance). Start with the explicit blowup solution (3.6) with $t_0 = 0$ and apply a Lorentz transform to it, to create a solution which blows up at the point $(0,0)$ but is smooth in the backwards light cone $\{(t,x) : |x| < -t\}$. Now work on the time slice $t = -1$ and localise the initial data to a neighbourhood of the ball $\{|x| < 1\}$ using finite speed of propagation, to create smooth initial data $(u(-1), \partial_t u(-1))$ whose $H^s_x(\mathbf{R}^d) \times H^{s-1}_x(\mathbf{R}^d)$ norm is arbitrarily small, but which develops a singularity at time 0. Rescaling this, we can construct data of arbitrarily small $H^s_x(\mathbf{R}^d) \times H^{s-1}_x(\mathbf{R}^d)$ norm with a solution that blows up in arbitrarily small time, which defeats any hope of a reasonable wellposedness theory at this regularity.

3.9. Almost conservation laws

The Law of conservation of energy tells us we can't get something for nothing, but we refuse to believe it. (Isaac Asimov)

We have seen how the laws of conservation of mass and energy can be used to obtain global wellposedness results at the L^2_x and H^1_x regularities respectively; generally speaking, they assert that these two norms stay bounded for all time. One may then ask what happens to the other H^s_x norms; after all, the *linear* Schrödinger and wave flows preserve the H^s_x norm and $H^s_x \times H^{s-1}_x$ norms respectively. In particular, once one has global wellposedness for one norm, say H^1_x, one can ask whether the same wellposedness occurs for other regularities also.

Generally speaking, global wellposedness in lower norms implies global wellposedness in higher norms[45], due to persistence of regularity; we have already seen several examples of this phenomenon in this chapter. However, while the lower Sobolev norms such as H^1_x may remain bounded uniformly in time, the bounds one obtains on higher norms such as H^2_x may grow faster than this; see for instance the exponential bounds in Exercise 3.28. This reflects the fact that the persistence of regularity arguments do not prohibit a "low-to-high frequency cascade" scenario, in which the energy starts off concentrated in low frequencies but moves increasingly to higher frequencies as time progresses; it is easy to envisage a scenario of this form where the H^1_x norm stays bounded, but higher norms such as H^2_x go to infinity. Numerical simulations have confirmed this type of *weak turbulence* behaviour for the periodic analogues of NLS and NLW, but for the non-periodic defocusing setting it appears that such phenomena, if they exist at all, do not occur with any

[45] Assuming of course that the nonlinearity itself is smooth enough to support solutions at this level of regularity.

great strength generically. (Indeed, the *solition resolution conjecture* mentioned earlier is probably not consistent with weakly turbulent behaviour.) It would be of interest to obtain more theoretical results regarding this issue.

Somewhat dual to this is the problem of starting with a global wellposedness result, say at H^1_x, and trying to lower the regularity needed for global existence, say to H^s_x for some $0 < s < 1$. This is reasonable in the H^1_x-subcritical case with s subcritical, since in such cases one already knows that the *local* wellposedness theory can extend below H^1_x. Indeed, the H^s_x local wellposedness theory asserts in this case that the only way an H^s_x-solution can cease to exist is if the H^s_x norm blows up in finite time. Thus the difficulty is to establish some upper bounds on the growth of the H^s_x norm in time; by limiting arguments one can restrict attention to the global H^1_x solutions, so long as the final bound on the H^s_x norm growth depends only on the H^s_x norm of the initial datum rather than on the energy. Here, the major difficulty is caused by the "high-to-low frequency cascade" scenario, in which one starts initially with a very large amount of energy at high frequencies (which may have small H^s_x norm), but a significant fraction of this energy somehow makes its way to low frequencies, thus causing the H^s_x norm to grow substantially.

To summarise, in order to establish good global existence results either for s above or below the energy regularity H^1_x one needs to control the flow of energy either from low frequencies to high frequencies or vice versa. In recent years, two methods have been developed to achieve such a control, namely the *Fourier truncation method* of Bourgain, and the subsequent *method of almost conserved*[46] *quantities* or *I-method* of Colliander, Keel, Staffilani, Takaoka, and Tao. The two methods are similar (indeed, the former inspired the latter) but not identical. They both proceed by selecting a large frequency cutoff N, and declaring frequencies less than N to be "low" and greater than N to be "high". If the solution has regularity H^s_x for some $s < 1$, then the low frequency components will have bounded energy (but with a bound depending on N), but the high frequency components will have unbounded or infinite energy. The strategy is then to somehow suppress the unbounded energy high frequency component in order that the energy conservation law can be usefully applied. The Fourier truncation method achieves this by viewing the original equation as a weakly coupled system of the high and low frequency components. Then one attempts to omit the nonlinear effects of the high frequencies, so that one believes the high frequencies to evolve approximately linearly, and the low frequencies to evolve approximately via the original equation. In particular one expects the low frequencies to (approximately) conserve their energy (as opposed to exporting or importing energy with the high frequencies). On short time intervals, one can justify this approximation using stability theory; the strategy is then to iterate this control on short time intervals to control on long-time intervals. Thus turns out to be possible by choosing N to be large, provided that the initial regularity H^s_x is sufficiently close to H^1_x, and provided that the nonlinearity has a certain "smoothing" property (roughly speaking, one wants the effect of the nonlinearity to be bounded in H^1_x even when the solution is only as regular as H^s_x).

[46]We shall use the term "almost conserved quantity" rather loosely; for us, it shall mean a quantity whose time derivative is unexpectedly "small" or "low order" in some sense. The monotone quantities appearing in monotonicity formulae could also be viewed as a type of almost conserved quantity.

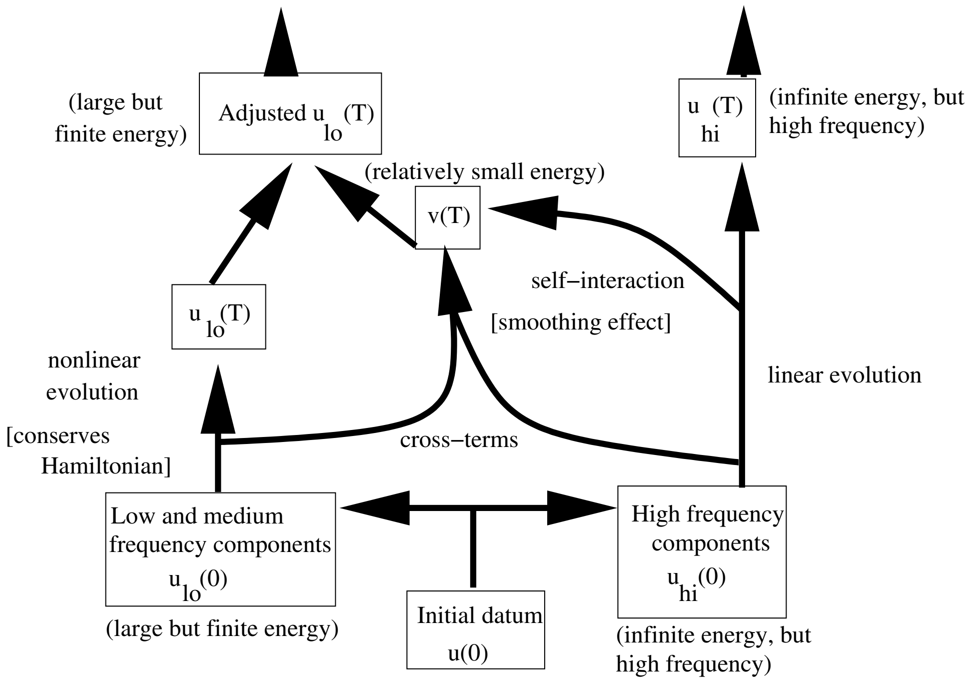

FIGURE 13. The first step in the Fourier truncation method. For a short time t, one evolves the high frequencies linearly and the low frequencies nonlinearly (thus preserving the Hamiltonian of the low frequencies). The error term v arises both from high-low frequency interactions and high-high frequency interactions; if the equation has enough smoothing properties, this error will be small in energy norm and can be safely absorbed into the low frequency component. One then iterates this scheme for as long as one has good control on all components.

The Fourier truncation method is surveyed in [**Bou9**] and will not be detailed here (but see Figure 13 and Table 2). The I-method proceeds slightly differently; rather than omit the high frequencies completely, it merely *damps* them using a Fourier multiplier I (hence the name "I-method"). This damping operator is essentially the mildest operator that makes the high frequencies bounded in energy; the low frequencies remain undamped by this operator. One then tries to control the energy $E[Iu(t)]$ of the damped solution Iu to the equation, which consists of the unadulterated low frequencies and the damped high frequencies. This quantity turns out to enjoy an *almost conservation law*, in that the quantity $E[Iu(t)]$ does not vary very quickly in t. (Note that if I is the identity then $E[Iu(t)]$ would be constant; thus this almost conservation nature reflects the "mild" nature of the operator I.) One can then use this almost conserved quantity to generate long-time control of the solution in much the same way that a genuine conservation law can be used to ensure global wellposedness. If all goes well, the time upon which one ultimately gets a useful control on the solution will be a *positive* power of N; letting N go to infinity will then yield the desired global wellposedness in H^s_x. This method can handle slightly more general nonlinearities and regularities than the Fourier truncation method, because no smoothing effect is required (this is due to a certain cancellation arising from a "commutator" term in the almost conservation law, which has no counterpart in the Fourier truncation approach), but provides slightly less information on the solution.

TABLE 2. An oversimplified comparison between the ways the Fourier restriction method and I-method treat different frequency interactions in a nonlinear equation. In both cases the low-low frequency interactions are considered large, but they do not alter the Hamiltonian, while the high-high frequency interactions are treated as error terms. The main difference lies in how the high-low interactions are treated, with the I-method taking advantage of commutator cancellations to show that these interactions approxiately conserve the damped Hamiltonian $E(Iu)$. Also, the Fourier restriction method takes advantage of smoothing effects and obtains better (energy-class) control of error terms.

Interaction	Fourier restriction method	I-method
Low-low	Conserves $E(u_{lo})$	Conserves $E(Iu)$
High-low	Small error in H_x^1	Approximately conserves $E(Iu)$
High-high	Small error in H_x^1	Small error in IH_x^1

Let us illustrate the method with the one-dimensional quintic defocusing NLS

$$i\partial_t u + \frac{1}{2}\partial_{xx} u = |u|^4 u, \tag{3.74}$$

which one can proceed by a relatively simple "energy method" implementation of the approach[47]. Indeed for this equation one can obtain global H_x^1-wellposed and classical solutions without any difficulty. (On the other hand, this equation is L_x^2-critical, and global wellposednes of L_x^2 solutions is unknown.) Now let u be a classical solution and $0 < s < 1$; we are interested in the behaviour of the H_x^s norm of $u(t)$ as $t \to \infty$. We already have conservation of the energy $E[u] := \int_{\mathbf{R}} \frac{1}{2}|\partial_x u|^2 + \frac{1}{3}|u|^6\, dx$ and mass $M[u] := \int_{\mathbf{R}} |u|^2\, dx$, but we will be reluctant to use $E[u]$ directly as it will not be controlled purely by the H_x^s norm. To create some almost conserved quantities at the regularity H_x^s, let us introduce a large frequency cutoff $N \gg 1$ and a spatial Fourier multiplier I defined by

$$\widehat{Iu}(\xi) := m_N(\xi) = m(\frac{\xi}{N}),$$

where $m_N(\xi) = m(\xi/N)$ and m is a smooth function which equals 1 for $|\xi| \leq 1$ and is equal to $|\xi|^{s-1}$ for $|\xi| \geq 2$. Thus I is the Identity operator on low frequencies $|\xi| \leq N$, and is essentially an Integration operator $N^{1-s}|\nabla|^{s-1}$ on high frequencies $|\xi| \geq$

[47] This is similar in spirit to the "energy cancellation" methods for establishing local existence for various nonlinear equations without performing an iteration scheme, and which can exploit certain structural cancellations arising from the nonlinearity; see for instance the high-regularity arguments in Section 4.1, Section 4.4, or Section 6.1. Most applications of the I-method, however, also require a modified local wellposedness statement which is obtained by standard iterative means, in order to exploit various local smoothing effects that can only be captured by spacetime norms.

N; this explains why this operator is denoted "I". We now show that the modified energy $E[Iu(t)] = \int_{\mathbf{R}} \frac{1}{2}|\partial_x Iu(t)|^2 + \frac{1}{3}|Iu(t)|^6\ dx$ obeys an almost conservation law.

PROPOSITION 3.38 (Almost conservation law). *Let $s > 1/2$. Suppose t is a time such that $E[Iu(t)] \lesssim 1$. Then $|\partial_t E[Iu(t)]| \lesssim_s N^{-1/2}$.*

The exponent $1/2$ here might not be best possible. An improvement of the exponent here will lead to a better global wellposedness result for a conclusion, as will be clear from the remainder of this argument.

PROOF. For a general classical field v, we have the identity

$$\partial_t E[v(t)] = -2\mathrm{Re} \int_{\mathbf{R}} \overline{\partial_t v}(i\partial_t v + \frac{1}{2}\partial_{xx} v - |v|^4 v)\ dx$$

which can be easily verified by integration by parts; note that this reproves the conservation of energy for (3.74). We now set $v := Iu$; by applying I to (3.74) we see that v solves the equation

$$i\partial_t v + \frac{1}{2}\partial_{xx} v = I(|u|^4 u)$$

and hence we have

$$\partial_t E[v(t)] = -2\mathrm{Re} \int_{\mathbf{R}} \overline{I\partial_t u}(I(|u|^4 u) - |Iu|^4 Iu)\ dx.$$

Thus it will suffice to establish the bound

$$|\int_{\mathbf{R}} \overline{I\partial_t u}(I(|u|^4 u) - |Iu|^4 Iu)\ dx| \lesssim N^{-1/2}.$$

Splitting $\partial_t u = \frac{1}{2}\partial_{xx} u - i|u|^4 u$, we can split this further into

$$|\int_{\mathbf{R}} \overline{I\partial_{xx} u}(I(|u|^4 u) - |Iu|^4 Iu)\ dx| \lesssim N^{-1/2} \tag{3.75}$$

and

$$|\int_{\mathbf{R}} \overline{I(|u|^4 u)}(I(|u|^4 u) - |Iu|^4 Iu)\ dx| \lesssim N^{-1/2}. \tag{3.76}$$

We shall just prove the top order estimate (3.75) and leave the lower order estimate (3.76) as an exercise. To avoid technicalities we shall "cheat" somewhat by assuming heuristics such as the fractional Leibnitz rule (A.14) as if they were rigorous; one can justify all the cheats performed here by Littlewood-Paley theory and other tools of Fourier analysis but we shall not do so here. We integrate one of the partial derivatives by parts, and observe from the hypothesis $E[Iu(t)] \lesssim 1$ that $\|I\partial_x u\|_{L^2_x} \lesssim 1$. Thus it suffices to show the commutator estimate

$$\|\partial_x[I(|u|^4 u) - |Iu|^4 Iu]\|_{L^2_x} \lesssim N^{-1/2}/$$

We split $u = u_{hi} + u_{lo}$, where $u_{hi} := P_{>N/100} u$ and $u_{lo} := P_{\leq N/100} u$. We can then expand $|u|^4 u$ and $|Iu|^4 Iu$ into a large number of terms involving five factors from u_{hi} and u_{lo}. There are three types of terms to consider. First consider the "low-low" terms that only involve u_{lo}:

$$\|\partial_x[I(|u_{lo}|^4 u_{lo}) - |Iu_{lo}|^4 Iu_{lo}]\|_{L^2_x}.$$

Because I is the Identity on low frequencies, we see that both $I(|u_{lo}|^4 u_{lo})$ and $|Iu_{lo}|^4 Iu_{lo}$ are equal to $|u_{lo}|^4 u_{lo}$ and so the net contribution of these terms is zero.

Next, consider any "high-high" term that involves two or more factors of u_{hi}. Here is a typical one (we do not attempt to exploit cancellation here):

$$\|\partial_x I(\overline{u_{hi}}^2 u_{lo}^3)\|_{L^2_x}.$$

The operator $\partial_x I$ is a pseudodifferential operator of positive order. Applying the fractional Leibnitz rule, we can distribute this operator and end up considering terms, of which the following is typical:

$$\|O(|\partial_x I u_{hi}||u_{hi}||u_{lo}|^3\|_{L^2_x}.$$

Now from the Gagliardo-Nirenberg inequality and the hypothesis $E[Iu(t)] \lesssim 1$ we have $\|Iu\|_{L^\infty_x} \lesssim 1$, and in particular $\|u_{lo}\|_{L^\infty_x} \lesssim 1$. Also we have already remarked that $\|\partial_x I u_{hi}\|_{L^2_x} \lesssim 1$, which by an easy Fourier analytic argument (exploiting the high frequency nature of u_{hi} and the hypothesis $s > 1/2$) implies that $\|u_{hi}\|_{L^\infty_x} \lesssim_s N^{-1/2}$; see Exercise 3.68. The desired bound then follows from Hölder's inequality.

Finally, we must consider "high-low" terms involving only one factor of u_{hi}. Here we must use[48] the cancellation present in (3.75). A typical term to consider is

$$\|\partial_x[I(|u_{lo}|^4 u_{hi}) - |u_{lo}|^4 I u_{hi}]\|_{L^2_x},$$

where we have used the fact that $Iu_{lo} = u_{lo}$. The expression in brackets is the commutator of I and $|u_{lo}|^4$, applied to u_{hi}. Let us write $w := |u_{lo}|^4$. The Fourier transform of $\partial_x(I(wu_{hi}) - wIu_{hi})$ at ξ can be computed to be

$$i \int_{\mathbf{R}} \xi[m(\frac{\xi}{N}) - m(\xi - \eta N)]\hat{w}(\eta)\hat{u_{hi}}(\xi - \eta)\, d\eta.$$

The integrand vanishes unless $|\eta| \lesssim N$ and $|\xi| \gtrsim N$. In such a case, an application of the mean-value theorem gives the bound $\xi(m(\frac{\xi}{N}) - m(\xi - \eta N)) = O(|\eta|m((\xi - \eta)/N))$. On the other hand, the expression

$$\int_{\mathbf{R}} |\eta| m((\xi - \eta)/N)\hat{w}(\eta)\hat{u_{hi}}(\xi - \eta)\, d\eta$$

is essentially the Fourier transform of $(|\nabla| w) I u_{hi}$. Thus we morally have

$$\partial_x[I(|u_{lo}|^4 u_{hi}) - |u_{lo}|^4 I u_{hi}] \text{“} \lesssim'' \ (|\nabla| w) I u_{hi}$$

in some Fourier sense. Assuming this to be rigorous, we are reduced to establishing that

$$\|(|\nabla| w) I u_{hi}\|_{L^2_x} \lesssim N^{-1/2}.$$

Now we already know that $\|Iu_{hi}\|_{L^\infty_x} \lesssim N^{-1/2}$. Also by distributing the derivative $|\nabla|$ using the fractional Leibnitz rule, we can (morally) replace $|\nabla| w$ by an expression such as $O(||\nabla| u_{lo}| \times |u_{lo}|^3)$. Since we already know that $\|u_{lo}\|_{L^\infty_x} \lesssim \|Iu\|_{L^\infty_x} \lesssim 1$ and $\||\nabla| u_{lo}\|_{L^2_x} \lesssim \|\nabla I u\|_{L^2_x} \lesssim 1$, the claim now follows from Hölder's inequality. $\square$

From the above proposition and the continuity method, we conclude that if $E[Iu(0)] = O(1)$, then in fact $E[Iu(t)] = O(1)$ for all $|t| \ll_s N^{1/2}$. Thus the quantity $E[Iu(t)]$ is stable for long periods of time. One can now apply scaling arguments and some Fourier analysis to conclude

[48] An alternative would be to try to average in time and exploit bilinear refinements to Strichartz' inequality here; this is related to the extra smoothing effect alluded to earlier. However, the approach given in the text demonstrates that one can use the commutator cancellation in the I-method as a substitute for such smoothing effects.

PROPOSITION 3.39. *If* $\|u(0)\|_{H^s_x} \lesssim 1$, *then* $\|u(t)\|_{H^s_x} \lesssim_s N^{1-s}$ *for all* $|t| \ll_s N^{\frac{1}{2}-2(1-s)}$.

We leave the derivation of this proposition to the exercises. If $s > 3/4$, the exponent of N in the bound on $|t|$ is positive, and so by letting $N \to \infty$ we can conclude a growth bound on the H^s_x norm. In fact we obtain the polynomial bound

$$\|u(t)\|_{H^s_x} \lesssim_s \langle t \rangle^{(1-s)/(\frac{1}{2}-2(1-s))}.$$

This, combined with the local H^s_x wellposedness theory, easily gives global wellposedness for this equation in H^s_x for all $3/4 < s < 1$. (Wellposedness for $s \geq 1$ already follows from energy conservation and persistence of regularity.)

The above strategy is rather flexible and can be adapted to a variety of subcritical equations; see for instance [**CKSTT**], [**CKSTT3**], [**CKSTT4**], [**Mat**], [**Pec**], [**Pec2**], [**Pec3**], [**Car**]. It also combines well with scattering theory (see [**CKSTT7**], [**CKSTT10**]; also see [**Bou6**] for an application of the Fourier restriction method to the scattering problem), to the growth of higher Sobolev norms (see [**CDKS**], [**Sta**], [**CKSTT8**], [**Bou10**]) and to the stability theory of solitons (see [**CKSTT8**], [**CKSTT9**]). In many cases it is not practical to obtain a *pointwise* bound on the time derivative $\partial_t E[Iu(t)]$ as in Proposition 3.38, but all one really needs anyway is a bound on the *integral* $\int_{t_0}^{t_1} \partial_t E[Iu(t)]\, dt$ of this time derivative. This additional time averaging allows one to use additional spacetime norms such as Strichartz norms, which can lead to better estimates. In such cases, one needs an additional ingredient in the argument, namely a "modified local existence theorem" that asserts that whenever $E[Iu(t)]$ is bounded, then certain spacetime norms of Iu (such as Strichartz norms) are bounded on a time interval centred at t. This however can be achieved by a routine modification of the local existence theory; see Figure 14 for a summary of this scheme. One can also exploit other conservation laws (e.g. mass conservation) to try to improve the powers of N which appear in the above argument. However, the most powerful methods for improving the exponents here has proceeded by modifying either the Hamiltonian $E[u]$ or the almost conserved quantity $E[Iu(t)]$ with additional correction terms to damp out some "nonresonant" fluctuations; see Section 4.2. For instance, the quintic NLS discussed above is in fact known to be globally wellposed in H^s_x for all $s > 4/9$ using this technique; see [**Tzi**].

EXERCISE 3.68. Prove that $\|u_{hi}\|_{L^\infty_x} \lesssim_s \|\partial_x I u_{hi}\|_{L^2_x}$ whenever u_{hi} is a Schwartz function supported on frequencies $> N/100$. (Hint: use frequency decomposition and either Bernstein's inequality (A.6) or Sobolev embedding.)

EXERCISE 3.69. Prove (3.76). (Here one will have to make some use of the potential energy component of $E[Iu]$, which gives a useful bound on $\|Iu\|_{L^6_x}$. This can be combined with the bound one already has on $\|Iu\|_{L^\infty_x}$, after decomposing into high and low frequencies.)

EXERCISE 3.70. Prove Proposition 3.39. (Hint: choose a $\lambda \geq 1$ such that the rescaled solution $u_\lambda(t,x) := \frac{1}{\lambda^{1/2}} u(\frac{t}{\lambda^2}, \frac{x}{\lambda})$ obeys $E[Iu_\lambda(0)] \lesssim 1$ (you may find taking the Fourier transform to be helpful). Then apply the almost conservation of $E[Iu_\lambda]$ for a long period of time, and then undo the scaling. One can use mass conservation to control the lower order component of the H^s_x norm.)

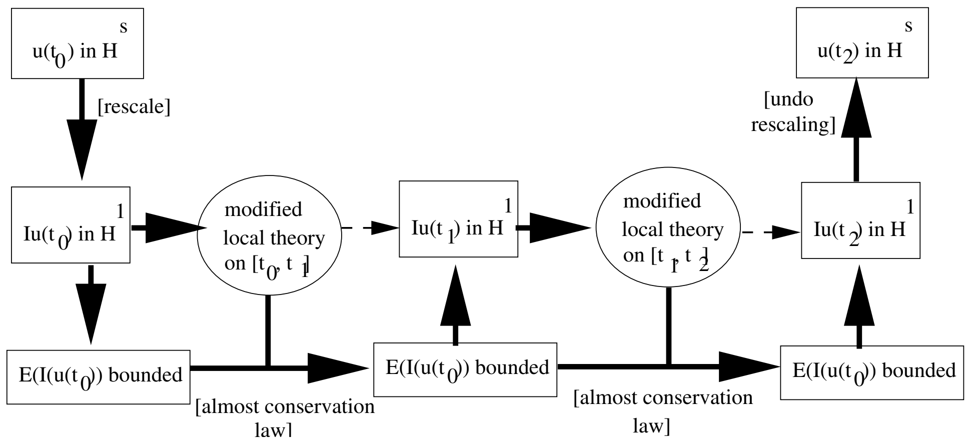

FIGURE 14. The general scheme of the I-method; compare with Figure 7. Of course, one would usually iterate the method for more than the two time steps indicated here. Apart from the rescaling and the presence of the I operator, one new feature is that the (modified) local theory plays a *quantitative* role rather than merely a qualitative one, as this theory is necessary to control the error terms in the almost conservation law. However, it is important that the local theory does not impact the main term in that law, otherwise the H^1_x norm of $Iu(t)$ could increase exponentially with each time step.

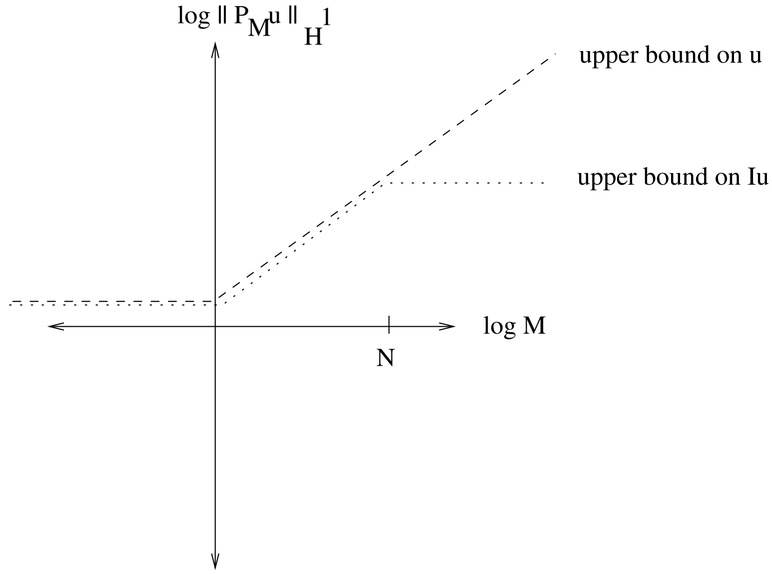

FIGURE 15. A log-log plot of the energy of the Littlewood-Paley pieces $P_M u(0)$ of $u(0)$ as a function of M, when $u(0)$ is assumed to only lie in H^s_x for some $s < 1$. Note the infinite energy at high frequencies. The operator I smooths out the energy at high frequencies, giving $Iu(0)$ a large but finite energy. A rescaling is then needed to make the energy bounded by 1.

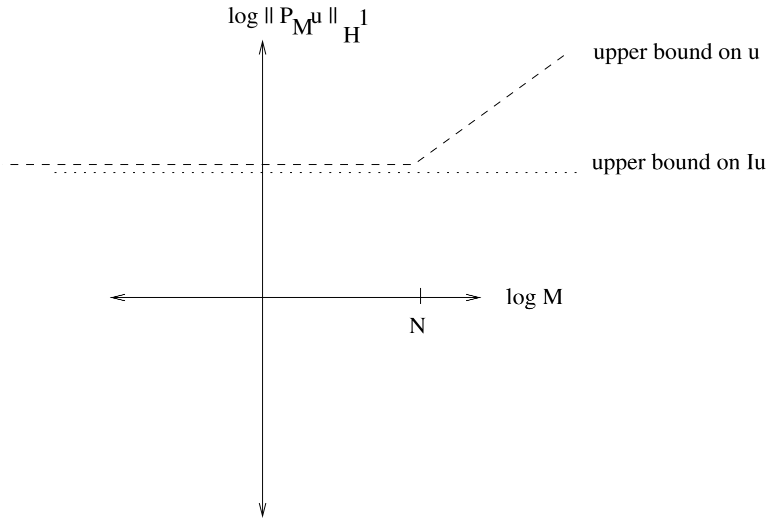

FIGURE 16. A log-log plot of the energy of the Littlewood-Paley pieces $P_M u(t)$ of $u(t)$ as a function of M, for some late time t (a power of N). The almost conservation law keeps the energy of Iu bounded but large (because of the scaling), but we do not exclude the possibility that energy has been moved around in the frequency ranges $M \lesssim N$. Thus at each time t, the high frequencies still evolve in a somewhat linear (non-interacting) fashion, but the low and medium frequencies may share their energy with each other. As time progresses, more and more frequencies could mix their energy, potentially leading to a polynomial growth in the H^s_x norm.

CHAPTER A

Appendix: tools from harmonic analysis

Every action of our lives touches on some chord that will vibrate in eternity. (Sean O'Casey)

The nonlinear evolution equations studied here can be profitably analyzed by viewing these equations as describing the oscillation and interaction between low, medium, and high frequencies. To make this type of analysis rigorous, we of course need the notation and tools of harmonic analysis, and in particular the Fourier transform and Littlewood-Paley theory; the purpose of this appendix is to review that material. This is only an outline of the material; for a more thorough introduction to these tools from a PDE-oriented perspective, see [**Tay**], [**Tay2**].

It is convenient to work in the Schwartz class $\mathcal{S}_x(\mathbf{R}^d)$. One particularly important operation on Schwartz functions (and hence on their dual) is the (spatial) *Fourier transform* $f \mapsto \hat{f}$, defined for $f \in \mathcal{S}_x(\mathbf{R}^d)$ by the formula[1]

$$\hat{f}(\xi) := \int_{\mathbf{R}^d} f(x) e^{-ix\cdot\xi}\ dx.$$

As is well known, the Fourier transform $f \mapsto \hat{f}$ is a Frechet space automorphism on the Schwartz space $\mathcal{S}_x(\mathbf{R}^d)$, with the inversion formula

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbf{R}^d} \hat{f}(\xi) e^{ix\cdot\xi}\ d\xi.$$

Thus every Schwartz function can be formally viewed as a superposition of plane waves $e^{ix\cdot\xi}$. We also have the fundamental *Plancherel identity*

$$\int_{\mathbf{R}^d} |f(x)|^2\ dx = \frac{1}{(2\pi)^d} \int_{\mathbf{R}^d} |\hat{f}(\xi)|^2\ d\xi$$

as well as the closely related *Parseval identity*[2]

$$\int_{\mathbf{R}^d} f(x)\overline{g(x)}\ dx = \frac{1}{(2\pi)^d} \int_{\mathbf{R}^d} \hat{f}(\xi)\overline{\hat{g}(\xi)}\ d\xi. \tag{A.1}$$

The Fourier transform enjoys many symmetry properties, several of which we list in Table 1. Of particular importance to PDE is the relation lies in the fact that it

[1]It is customary to omit the factor of 2π from the Fourier exponent in PDE, in order to simplify the Fourier multipliers associated to any given PDE; of course, this factor then surfaces in the Fourier inverion formula. In any event, the factors of 2π make only a negligible impact on the theory, so much so that some authors choose to abuse notation slightly, and simply omit all factors of 2π in their arguments.

[2]Some authors reverse the attribution of these two identities, which are easily shown to be equivalent. Strictly speaking, Parseval's original identity was for Fourier series, whereas Plancherel's theorem concerned Fourier integrals.

diagonalises any constant coefficient operator $P(\nabla)$:

$$\widehat{P(\nabla)f}(\xi) = P(i\xi)\hat{f}(\xi).$$

Thus differential operators amplify high frequencies and attenuate low frequencies; integration operators of course do the reverse. Note that if $P(\nabla)$ is skew-adjoint, then $P(i\xi)$ is automatically skew-adjoint; this can be shown directly, and also follows from (A.1). Indeed in this case we have $P(\nabla) = ih(\nabla/i)$ for some real-valued polynomial $h : \mathbf{R}^d \to \mathbf{R}$.

TABLE 1. Some operations on functions $f(x)$, and their Fourier transform. Here $x_0, \xi_0 \in \mathbf{R}$, $f, g \in \mathcal{S}_x(\mathbf{R}^d)$, $\lambda \in \mathbf{R}\backslash\{0\}$, $P : \mathbf{R}^d \to \mathbf{C}$ is a polynomial, and $f * g(x) := \int_{\mathbf{R}^d} f(y)g(x-y)\,dy$.

$f(x)$	$\hat{f}(\xi)$
$f(x-x_0)$	$e^{-ix_0\cdot\xi}\hat{f}(\xi)$
$e^{ix\cdot\xi_0}f(x)$	$\hat{f}(\xi-\xi_0)$
$\overline{f(x)}$	$\overline{\hat{f}(-\xi)}$
$f(x/\lambda)$	$\|\lambda\|^d\hat{f}(\lambda\xi)$
$f * g(x)$	$\hat{f}(\xi)\hat{g}(\xi)$
$f(x)g(x)$	$\frac{1}{(2\pi)^d}\hat{f} * \hat{g}(\xi)$
$P(\nabla)f$	$P(i\xi)f$

The Fourier transform can be extended to Lebesgue spaces such as $L^2_x(\mathbf{R}^d)$ using Plancherel's theorem (where it essentially becomes an isometry), and also to the space of tempered distributions $\mathcal{S}_x(\mathbf{R}^d)^*$.

An important concept for us shall be that of a *Fourier multiplier*. If we are given a locally integrable function $m : \mathbf{R}^d \to \mathbf{C}$ of at most polynomial growth, we can define the associated multiplier $m(\nabla/i) : \mathcal{S}_x(\mathbf{R}^d) \to \mathcal{S}_x(\mathbf{R}^d)^*$ via the Fourier transform by the formula

$$\widehat{m(\nabla/i)f}(\xi) := m(\xi)\hat{f}(\xi)$$

or equivalently

$$m(\nabla/i)f(x) := \frac{1}{(2\pi)^d}\int_{\mathbf{R}^d} m(\xi)\hat{f}(\xi)e^{2\pi i x\cdot\xi}\,d\xi.$$

This notation is consistent with that of constant-coefficient differential operators $h(\nabla/i)$. We also have (formally at least) the *multiplier calculus*

$$\begin{aligned} m(\nabla/i)^* &= \overline{m}(\nabla/i); \\ m_1(\nabla/i) + m_2(\nabla/i) &= (m_1+m_2)(\nabla/i); \\ m_1(\nabla/i)m_2(\nabla/i) &= (m_1m_2)(\nabla/i). \end{aligned}$$

In particular, Fourier multipliers all (formally) commute with each other. The function $m(\xi)$ is known as the *symbol* of the operator $m(\nabla/i)$. Important examples of Fourier multipliers include the constant coefficient differential operators $h(\nabla/i)$, the propagators $e^{ith(\nabla/i)}$ discussed in Section 2.1, and the fractional differentiation and integration operators $|\nabla|^\alpha$ and $\langle\nabla\rangle^\alpha$ defined[3] for all $\alpha \in \mathbf{R}$, with symbols $|\xi|^\alpha$ and $\langle\xi\rangle^\alpha$ respectively. This in turn leads to the *Sobolev spaces* $W^{s,p}_x(\mathbf{R}^d)$ and the *homogeneous Sobolev spaces* $\dot W^{s,p}_x(\mathbf{R}^d)$, defined for $s \in \mathbf{R}$ and $1 < p < \infty$ as the closure of the Schwartz functions under their respective norms

$$\|f\|_{W^{s,p}_x(\mathbf{R}^d)} := \|\langle\nabla\rangle^s f\|_{L^p_x(\mathbf{R}^d)}$$

and

$$\|f\|_{\dot W^{s,p}_x(\mathbf{R}^d)} := \||\nabla|^s f\|_{L^p_x(\mathbf{R}^d)}.$$

Thus these spaces generalise the Lebesgue spaces, which correspond to the cases $s = 0$. In the special case $p = 2$, we write H^s_x and $\dot H^s_x$ for $W^{s,2}_x$ and $\dot W^{s,2}_x$ respectively. From Plancherel's theorem we observe that

$$\|f\|_{H^s_x(\mathbf{R}^d)} = \frac{1}{(2\pi)^{d/2}}\|\langle\xi\rangle^s \hat f\|_{L^2_\xi(\mathbf{R}^d)}$$

and similarly

$$\|f\|_{\dot H^s_x(\mathbf{R}^d)} = \frac{1}{(2\pi)^{d/2}}\||\xi|^s \hat f\|_{L^2_\xi(\mathbf{R}^d)}.$$

Using Calderón-Zygmund theory (see e.g. [**Stei**]), one can show the identities

$$\begin{aligned}
\|f\|_{W^{s,p}_x(\mathbf{R}^d)} &\sim_{s,p,d} \|f\|_{W^{s-1,p}_x(\mathbf{R}^d)} + \|\nabla f\|_{W^{s-1,p}_x(\mathbf{R}^d)};\\
\|f\|_{\dot W^{s,p}_x(\mathbf{R}^d)} &\sim_{s,p,d} \|\nabla f\|_{\dot W^{s-1,p}_x(\mathbf{R}^d)}
\end{aligned}$$

for any $1 < p < \infty$ and $s \in \mathbf{R}$. Iterating the above inequalities, we obtain that these Sobolev norms are equivalent (up to constants) to their classical counterparts, thus

$$\|f\|_{W^{k,p}_x(\mathbf{R}^d)} \sim_{k,p,d} \sum_{j=0}^k \|\nabla^j f\|_{L^p_x(\mathbf{R}^d)}$$

and

$$\|f\|_{\dot W^{k,p}_x(\mathbf{R}^d)} \sim_{k,p,d} \|\nabla^k f\|_{L^p_x(\mathbf{R}^d)}.$$

We will not define Sobolev spaces at $p = 1$ or $p = \infty$ to avoid the technicalities associated with endpoint Calderón-Zygmund theory.

Another important class of Fourier multipliers are the *Littlewood-Paley multipliers*. Let us fix a real-valued radially symmetric bump function $\varphi(\xi)$ adapted to the ball $\{\xi \in \mathbf{R}^d : |\xi| \le 2\}$ which equals 1 on the ball $\{\xi \in \mathbf{R}^d : |\xi| \le 1\}$; the exact choice of bump function turns out in practice to not be important[4]. Define a *dyadic number* to be any number $N \in 2^{\mathbf{Z}}$ of the form $N = 2^j$ where $j \in \mathbf{Z}$ is an integer; any sum over the dummy variable N or M is understood to be over

[3]For $\alpha \le -d$, the operator $|\nabla|^\alpha$ is only defined for Schwartz functions which obey enough moment conditions that their Fourier transform vanishes to high order at the origin. As we shall never use integration operators of such low order, we shall ignore this technicality.

[4]In the classical Littlewood-Paley theory (see e.g. [**Stei**]), one uses the harmonic extension or heat extension, which would correspond to the (non-compactly-supported) choices $\varphi(\xi) := e^{-|\xi|}$ or $\varphi(\xi) := e^{-|\xi|^2}$ respectively. However in the modern theory it has turned out to be more convenient to use compactly supported bump functions (but see Section 6.4).

dyadic numbers unless otherwise specified. For each dyadic number N, we define the Fourier multipliers

$$\begin{aligned}
\widehat{P_{\leq N} f}(\xi) &:= \varphi(\xi/N)\hat{f}(\xi) \\
\widehat{P_{>N} f}(\xi) &:= (1 - \varphi(\xi/N))\hat{f}(\xi) \\
\widehat{P_N f}(\xi) &:= (\varphi(\xi/N) - \varphi(2\xi/N))\hat{f}(\xi).
\end{aligned}$$

We similarly define $P_{<N}$ and $P_{\geq N}$. Thus P_N, $P_{\leq N}$, $P_{>N}$ are smoothed out projections to the regions $|\xi| \sim N$, $|\xi| \leq 2N$, $|\xi| > N$ respectively. Note in particular the telescoping identities

$$P_{\leq N} f = \sum_{M \leq N} P_M f; \quad P_{>N} f = \sum_{M > N} P_M f; \quad f = \sum_M P_M f$$

for all Schwartz f, where M ranges over dyadic numbers. We also define

$$P_{M < \cdot \leq N} := P_{\leq N} - P_{\leq M} = \sum_{M < N' \leq N} P_{N'}$$

whenever $M \leq N$ are dyadic numbers. Similarly define $P_{M \leq \cdot \leq N}$, etc.

Littlewood-Paley projections are extremely handy in the rigorous study of PDE, because they separate (in a quantitative manner) the rough (high-frequency, oscillating, low regularity) components of a solution from the smooth (low-frequency, slowly varying, high regularity) components, thus clarifying the nature of various components of the equation, such as derivatives and various nonlinear interactions of the solution with itself. The following heuristics are quite useful (see Figure 1).

PRINCIPLE A.1 (Uncertainty principle). *Let N be a dyadic number, and let f be a function on $\mathbf{R}^d$.*

- *(Low frequencies) If f has Fourier transform supported on frequencies of magnitude $|\xi| \lesssim N$ (e.g. if $f = P_{\leq N} g$ for some g), then f should be approximately constant on spatial balls of radius c/N for small c, and $\nabla^s f$ should be "dominated" by $N^s f$ for any $s \geq 0$. (Thus localisation at frequency scales N forces a spatial uncertainty of $1/N$; this is a manifestation of the* Heisenberg uncertainty principle $|\delta x \cdot \delta \xi| \gtrsim 1$*.)*
- *(High frequencies) If f has Fourier transform supported on frequencies of magnitude $|\xi| \gtrsim N$ (e.g. if $f = P_{\geq N} g$ for some g), then f should have approximate mean zero[5] on balls of radius C/N for large C, and $\nabla^{-s} f$ should be "dominated" by $N^{-s} f$ for any $s \geq 0$. (Thus exclusion of frequencies at scales N and below forces spatial oscillation at scale $1/N$.)*
- *(Medium frequencies) If f has Fourier transform supported on frequencies of magnitude $|\xi| \sim N$ (e.g. if $f = P_N g$ for some g), then both of the above heuristics should apply, and $\nabla^s f$ should be "comparable" to $N^s f$ for both positive and negative s.*

We now present some concrete estimates that make the above intuition rigorous. One easily verifies that $P_{\leq N}$ is a convolution operator, in fact

$$P_{\leq N} f(x) = \frac{1}{(2\pi)^d} \int_{\mathbf{R}^d} \hat{\varphi}(y) f(x + \frac{y}{N})\, dy.$$

[5]In fact, we expect higher moments to vanish as well, so that f should be approximately orthogonal to any bounded degree polynomials on these balls.

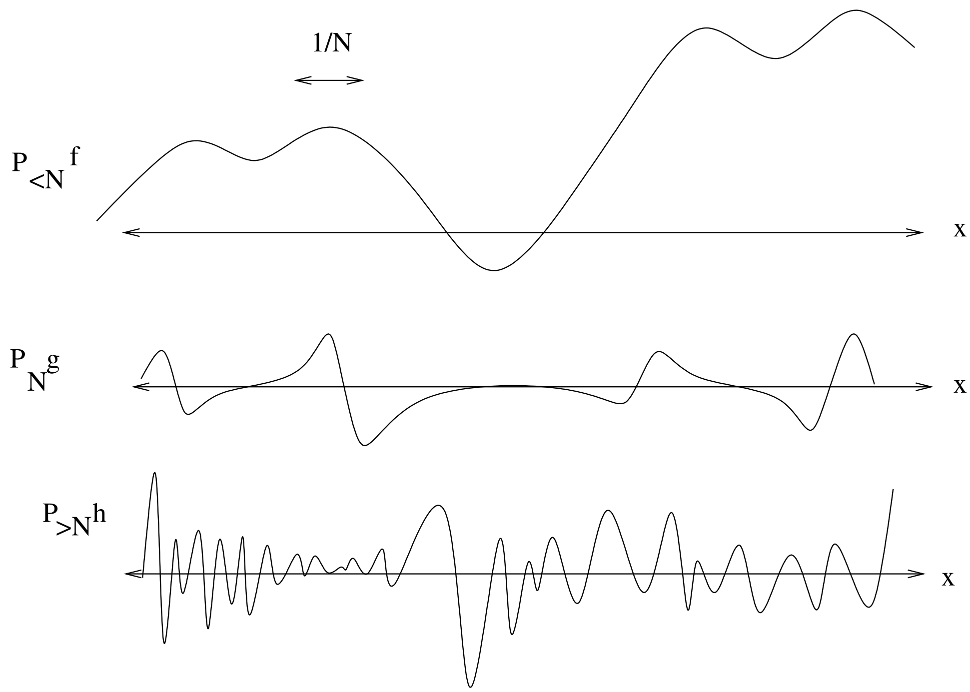

FIGURE 1. The uncertainty principle. The low-frequency function $P_{<N}f$ has frequencies less than N and is thus essentially constant at spatial scales $\ll 1/N$. The high-frequency function $P_{>N}h$ has frequencies greater than N and thus oscillates (with mean essentially zero) at spatial scales $\gg 1/N$. The medium-frequency function $P_N g$ behaves in both fashions simultaneously.

Since $\hat{\varphi}$ is rapidly decreasing and has unit mass, one thus can think of $P_{\leq N}$ as an averaging operator that "blurs" f by a spatial scale of $O(1/N)$, and localises f in frequency to the ball of radius $O(N)$, which is consistent with Principle A.1. From this identity one can easily verify (using Young's inequality, and the commutativity of all Fourier multipliers) that the above Littlewood-Paley operators are bounded (uniformly in N or M) on every Lebesgue space $L^p_x(\mathbf{R}^d)$ with $1 \leq p \leq \infty$, as well as every Sobolev space $W^{s,p}_x(\mathbf{R}^d)$, $\dot{W}^{s,p}_x(\mathbf{R}^d)$ for $s \in \mathbf{R}$ and $1 < p < \infty$. Furthermore, they obey the following easily verified (see Exercise A.1) and extremely useful *Bernstein inequalities* for $\mathbf{R}^d$ with $s \geq 0$ and $1 \leq p \leq q \leq \infty$:

$$\|P_{\geq N} f\|_{L^p_x(\mathbf{R}^d)} \lesssim_{p,s,d} N^{-s} \||\nabla|^s P_{\geq N} f\|_{L^p_x(\mathbf{R}^d)} \tag{A.2}$$

$$\|P_{\leq N} |\nabla|^s f\|_{L^p_x(\mathbf{R}^d)} \lesssim_{p,s,d} N^s \|P_{\leq N} f\|_{L^p_x(\mathbf{R}^d)} \tag{A.3}$$

$$\|P_N |\nabla|^{\pm s} f\|_{L^p_x(\mathbf{R}^d)} \sim_{p,s,d} N^{\pm s} \|P_N f\|_{L^p_x(\mathbf{R}^d)} \tag{A.4}$$

$$\|P_{\leq N} f\|_{L^q_x(\mathbf{R}^d)} \lesssim_{p,q,d} N^{\frac{d}{p}-\frac{d}{q}} \|P_{\leq N} f\|_{L^p_x(\mathbf{R}^d)} \tag{A.5}$$

$$\|P_N f\|_{L^q_x(\mathbf{R}^d)} \lesssim_{p,q,d} N^{\frac{d}{p}-\frac{d}{q}} \|P_N f\|_{L^p_x(\mathbf{R}^d)}. \tag{A.6}$$

Thus when the frequency is localised, one can upgrade low Lebesgue integrability to high Lebesgue integrability, at the cost of some powers of N; when the frequency

N is very low, this cost is in fact a gain, and it becomes quite desirable to use Bernstein's inequality whenever the opportunity arises. These estimates can be verified by computation of the distributional kernel of P_N and $P_{\leq N}$, and their derivatives, followed by Young's inequality. A deeper estimate, requiring some Calderón-Zygmund theory, is the *Littlewood-Paley inequality*

$$\|f\|_{L^p_x(\mathbf{R}^d)} \sim_{p,d} \|(\sum_N |P_N f|^2)^{1/2}\|_{L^p_x(\mathbf{R}^d)}; \tag{A.7}$$

see for instance [**Stei2**]. In a similar spirit, from Plancherel's theorem we have the estimate

$$\|f\|_{\dot H^s_x(\mathbf{R}^d)} \sim_{s,d} (\sum_N N^{2s} \|P_N f\|_{L^2_x(\mathbf{R}^d)}^2)^{1/2} \tag{A.8}$$

and

$$\|f\|_{H^s_x(\mathbf{R}^d)} \sim_{s,d} \|P_{\leq 1} f\|_{L^2_x(\mathbf{R}^d)} + (\sum_{N>1} N^{2s} \|P_N f\|_{L^2_x(\mathbf{R}^d)}^2)^{1/2} \tag{A.9}$$

As a sample application of these estimates, let us present

LEMMA A.2 (Hardy's inequality). *If $0 \leq s < d/2$ then*

$$\||x|^{-s} f\|_{L^2_x(\mathbf{R}^d)} \lesssim_{s,d} \|f\|_{\dot H^s_x(\mathbf{R}^d)}.$$

PROOF. The case $s = 0$ is trivial, so suppose $0 < s < d/2$. Using (A.8) it suffices to show that

$$\int_{\mathbf{R}^d} \frac{|f(x)|^2}{|x|^{2s}}\ dx \lesssim_{s,d} \sum_N N^{2s} \|P_N f\|_{L^2_x(\mathbf{R}^d)}^2.$$

We subdivide the left-hand side into dyadic shells and estimate

$$\int_{\mathbf{R}^d} \frac{|f(x)|^2}{|x|^{2s}}\ dx \lesssim_{s,d} \sum_R R^{-2s} \int_{|x|\leq R} |f(x)|^2\ dx$$

where R ranges over dyadic numbers. Using Littlewood-Paley decomposition and the triangle inequality, we have

$$(\int_{|x|\leq R} |f(x)|^2\ dx)^{1/2} \leq \sum_N (\int_{|x|\leq R} |P_N f(x)|^2\ dx)^{1/2}.$$

On the one hand we have the trivial estimate

$$(\int_{|x|\leq R} |P_N f(x)|^2\ dx)^{1/2} \leq \|P_N f\|_{L^2_x(\mathbf{R}^d)}$$

while on the other hand by Bernstein (A.6) and Hölder we have

$$(\int_{|x|\leq R} |P_N f(x)|^2\ dx)^{1/2} \lesssim_d R^{d/2} \|P_N f\|_{L^\infty_x(\mathbf{R}^d)} \lesssim_d (NR)^{d/2} \|P_N f\|_{L^2_x(\mathbf{R}^d)}.$$

Combining all these estimates together, we reduce to establishing that

$$\sum_R R^{-2s} (\sum_N \min(1, (NR)^{d/2}) \|P_N f\|_{L^2_x(\mathbf{R}^d)})^2 \lesssim_{s,d} \sum_N N^{2s} \|P_N f\|_{L^2_x(\mathbf{R}^d)}^2.$$

Writing $c_N := N^s \|P_N f\|_{L^2_x(\mathbf{R}^d)}$, this becomes

$$\|\sum_N \min((NR)^{-s}, (NR)^{d/2-s}) c_N\|_{l^2_R(2^{\mathbf{Z}})} \lesssim_{s,d} \|c_N\|_{l^2_N(2^{\mathbf{Z}})}$$

where $2^{\mathbf{Z}}$ is the space of dyadic numbers. But since $0 < s < d/2$, the kernel $\min(M^{-s}, M^{d/2-s})$ is absolutely convergent over dyadic numbers. The claim now follows from Young's inequality (or Minkowski's inequality, or Schur's test). □

In a similar spirit we have

PROPOSITION A.3 (Gagliardo-Nirenberg inequality). *Let $1 < p < q \leq \infty$ and $s > 0$ be such that*

$$\frac{1}{q} = \frac{1}{p} - \frac{\theta s}{d}$$

for some $0 < \theta < 1$. Then for any $u \in W_x^{s,p}(\mathbf{R}^d)$ we have

$$\|u\|_{L_x^q(\mathbf{R}^d)} \lesssim_{d,p,q,s} \|u\|_{L_x^p(\mathbf{R}^d)}^{1-\theta} \|u\|_{\dot{W}_x^{s,p}(\mathbf{R}^d)}^{\theta}.$$

In the special case $q = \infty$, we conclude in particular (by the usual approximation by Schwartz function argument) that u is in fact continuous (so it lies in $C_x^0(\mathbf{R}^d)$).

PROOF. We may of course assume that u is non-zero. The inequality is invariant under homogeneity $u(x) \mapsto \lambda u(x)$ and scaling $u(x) \mapsto u(x/\lambda)$ for any $\lambda > 0$. Using these invariances we may normalise $\|u\|_{L_x^p(\mathbf{R}^d)} = \|u\|_{\dot{W}_x^{s,p}(\mathbf{R}^d)} = 1$.

The next step is the Littlewood-Paley decomposition $u = \sum_N P_N u$, where N ranges over dyadic numbers. From the triangle inequality followed by Bernstein's inequality we have

$$\begin{aligned}
\|u\|_{L_x^q(\mathbf{R}^d)} &\leq \sum_N \|P_N u\|_{L_x^q(\mathbf{R}^d)} \\
&\lesssim_{d,p,q} \sum_N N^{\frac{d}{p} - \frac{d}{q}} \|P_N u\|_{L_x^p(\mathbf{R}^d)} \\
&= \sum_N N^{\theta s} \|P_N u\|_{L_x^p(\mathbf{R}^d)}
\end{aligned}$$

On the other hand, from (A.4) and the boundedness of P_N have

$$\|P_N u\|_{L_x^p(\mathbf{R}^d)} \lesssim_{d,p} \|u\|_{L_x^p(\mathbf{R}^d)} = 1; \quad \|P_N u\|_{L_x^p(\mathbf{R}^d)} \lesssim_{d,p,s} N^{-s} \||\nabla|^s u\|_{L_x^p(\mathbf{R}^d)} = N^{-s}.$$

Inserting this into the previous estimate we obtain

$$\|u\|_{L_x^q(\mathbf{R}^d)} \lesssim_{d,p,s} \sum_N N^{\theta s} \min(1, N^{-s}) \lesssim_{\theta,s} 1$$

and the claim follows (note that θ is determined by d, p, q, s). □

Closely related to the above two inequalities is the *Hardy-Littlewood-Sobolev theorem of fractional integration*, which asserts that

$$\|f * \frac{1}{|x|^\alpha}\|_{L_x^q(\mathbf{R}^d)} \lesssim_{p,q,d} \|f\|_{L_x^p(\mathbf{R}^d)} \tag{A.10}$$

whenever $1 < p < q < \infty$ and $0 < \alpha < d$ obey the scaling condition $\frac{1}{p} = \frac{1}{q} + \frac{d-\alpha}{d}$. This implies the *homogeneous Sobolev embedding*

$$\|f\|_{L_x^q(\mathbf{R}^d)} \lesssim_{p,q,d} \|f\|_{\dot{W}_x^{s,p}(\mathbf{R}^d)} \tag{A.11}$$

whenever $1 < p < q < \infty$ and $s > 0$ obey the scaling condition $\frac{1}{p} = \frac{1}{q} + \frac{s}{d}$, which in turn implies the *inhomogeneous Sobolev embedding*

$$\|f\|_{L_x^q(\mathbf{R}^d)} \lesssim_{p,q,s,d} \|f\|_{W_x^{s,p}(\mathbf{R}^d)} \tag{A.12}$$

whenever $1 < p < q < \infty$ and $s > 0$ is such that $\frac{1}{p} \leq \frac{1}{q} + \frac{s}{d}$. We leave the proofs as exercises. Note that the non-endpoint case $\frac{1}{p} < \frac{1}{q} + \frac{s}{d}$ of (A.12) already follows from Proposition A.3, and we also have an extension to the $q = \infty$ case, namely

$$\|f\|_{C^0_x(\mathbf{R}^d)} = \|f\|_{L^\infty_x(\mathbf{R}^d)} \lesssim_{p,s,d} \|f\|_{W^{s,p}_x(\mathbf{R}^d)}$$

whenever $1 < p < \infty$ and $s > 0$ is such that $\frac{1}{p} < \frac{s}{d}$. In particular we have

$$\|f\|_{C^0_x(\mathbf{R}^d)} = \|f\|_{L^\infty_x(\mathbf{R}^d)} \lesssim_{s,d} \|f\|_{H^s_x(\mathbf{R}^d)} \tag{A.13}$$

when $s > n/2$.

The Sobolev embedding theorem (A.11) is sharp in the following sense: if f is a rescaled bump function, say $f = N^\alpha \psi(Nx)$ for some $\psi \in \mathcal{S}_x(\mathbf{R}^d)$ and some $N > 0$ and $\alpha \in \mathbf{R}$, then one can verify that $\|f\|_{L^q_x(\mathbf{R}^d)} \sim_{\psi,q,d} N^{-d/q} N^\alpha$ and $\|f\|_{\dot W^{s,p}_x(\mathbf{R}^d)} \sim_{\psi,s,p,d} N^s N^{-d/p} N^\alpha$, and so from the scaling condition $\frac{1}{p} = \frac{1}{q} + \frac{s}{d}$ we see that both sides of (A.11) are comparable. A useful fact is that these bump functions are in some sense the *only* way in which both sides of Sobolev embedding estimate can be close to comparable. Indeed, we have

PROPOSITION A.4 (Inverse Sobolev theorem). *Let $1 < p < q < \infty$, $s > 0$, and $0 < \eta \leq 1$.*

- *If $\frac{1}{p} = \frac{1}{q} + \frac{s}{d}$ and f is such that $\|f\|_{\dot W^{s,p}_x(\mathbf{R}^d)} \lesssim 1$ and $\|f\|_{L^q_x(\mathbf{R}^d)} \gtrsim \eta$, then there exists a dyadic number N and a position $x_0 \in \mathbf{R}^d$ such that $|P_N f(x_0)| \sim_{p,q,d,\eta} N^{d/q}$, and furthermore*

$$\left(\int_{|x-x_0| \leq C/N} |P_N f(x)|^r \, dx\right)^{1/r} \sim_{p,q,d,\eta} N^{\frac{d}{q} - \frac{d}{r}}$$

  for all $1 \leq r \leq \infty$ and some large constant $C = C(p,q,d,\eta) > 0$.
- *If $\frac{1}{p} < \frac{1}{q} + \frac{s}{d}$ and f is such that $\|f\|_{W^{s,p}_x(\mathbf{R}^d)} \lesssim 1$ and $\|f\|_{L^q_x(\mathbf{R}^d)} \gtrsim \eta$, then there exists a dyadic number $N \sim_{p,q,s,d,\eta} 1$ and a position $x_0 \in \mathbf{R}^d$ such that $|P_N f(x_0)| \sim_{p,q,s,d,\eta} 1$, and furthermore*

$$\left(\int_{|x-x_0| \leq C} |P_N f(x)|^r \, dx\right)^{1/r} \sim_{p,q,s,d,\eta} 1$$

  for all $1 \leq r \leq \infty$ and some large constant $C = C(p,q,s,d,\eta) > 0$.

More informally, in order for (A.11) to be close to sharp, f must contain a large normalised bump function at some position x_0 and some frequency N (and wavelength $1/N$); in order for (A.12) to be sharp, we have a similar conclusion but with the additional information that the frequency N is comparable to 1. To put it another way, in order to come within a constant to saturating the Sobolev embedding theorem, the function must concentrate a significant portion of its $W^{s,p}_x$ "energy" in a ball. (See also Lemma B.4, which essentially asserts that if one comes within an *epsilon* of the best constant in a Sobolev embedding type theorem, then one must concentrate *nearly all* of one's energy in a ball.) The implicit constants here can be made more explicit, for instance the dependence on η is polynomial, but we will not need such quantitative bounds here. See [**BG**] for an application of these types of theorems to nonlinear wave equations.

PROOF. We will just prove the (easier) second half of the theorem here, and leave the first to Exercise A.7. We have

$$\eta \lesssim \|f\|_{L^q_x(\mathbf{R}^d)} \lesssim \sum_N \|P_N f\|_{L^q_x(\mathbf{R}^d)}.$$

Now from (A.6), (A.4), and the hypothesis $\|f\|_{W^{s,p}_x(\mathbf{R}^d)} \lesssim 1$ we have

$$\|P_N f\|_{L^q_x(\mathbf{R}^d)} \lesssim_{p,q,d,s} N^{\frac{d}{p}-\frac{d}{q}} \|P_N f\|_{L^p_x(\mathbf{R}^d)} \lesssim_{p,q,d,s} N^{\frac{d}{p}-\frac{d}{q}} \min(1, N^{-s}).$$

The hypotheses on p, q, s ensure that $\sum_N N^{\frac{d}{p}-\frac{d}{q}} \min(1, N^{-s})$ is geometrically decreasing as $N \to 0$ or $N \to \infty$ and is thus convergent. We conclude that there exists $N \sim_{p,q,s,d,\eta} 1$ such that $\|P_N f\|_{L^q_x(\mathbf{R}^d)} \gtrsim_{p,q,s,d,\eta} 1$. Since $\|P_N f\|_{L^p_x(\mathbf{R}^d)} \lesssim_{p,q,s,d,\eta} 1$, we conclude from Hölder's inequality that $\|P_N f\|_{L^\infty_x(\mathbf{R}^d)} \gtrsim_{p,q,s,d,\eta} 1$. Thus there exists $x_0 \in \mathbf{R}^d$ such that $|P_N f(x_0)| \gtrsim_{p,q,s,d,\eta} 1$. Writing $P_N f = P_{N/4<\cdot<4N} P_N f$, we can express $P_N f(x_0)$ as the inner product of $P_N f$ with a rapidly decreasing approximation to the identity centred at x_0. Since we also have $\|P_N f\|_{L^p_x(\mathbf{R}^d)} \lesssim_{d,p,s} 1$, an easy application of Hölder's inequality then gives

$$\int_{|x-x_0|\leq C} |P_N f(x)|\, dx \gtrsim_{p,q,s,d,\eta} 1$$

for some large $C = C(p, q, s, d, \eta)$. On the other hand, from Bernstein's inequality we have $\|P_N f\|_{L^\infty_x(\mathbf{R}^d)} \lesssim_{d,p,s} 1$. The claim then follows from Hölder's inequality again. □

We have seen how Littlewood-Paley technology is useful for understanding linear operations such as fractional integration. It is also invaluable in understanding *nonlinear* operations, such as multiplication $(f, g) \mapsto fg$ or composition $u \mapsto F(u)$ for various explicit functions F. Both of these operations arise constantly in nonlinear PDE, and there are two very useful heuristics that can be used to understand them:

PRINCIPLE A.5 (Fractional Leibnitz rule). *Let f, g be functions on $\mathbf{R}^d$, and let D^α be some sort of differential or pseudodifferential operator of positive order $\alpha > 0$.*

- *(High-low interactions) If f has significantly higher frequency than g (e.g. if $f = P_N F$ and $g = P_{<N/8} G$ for some F, G), or is "rougher" than g (e.g. $f = \nabla u$ and $g = u$ for some u) then fg will have comparable frequency to f, and we expect $D^\alpha(fg) \approx (D^\alpha f)g$. In a similar spirit we expect $P_N(fg) \approx (P_N f)g$.*
- *(Low-high interactions) If g has significantly higher frequency or is rougher than f, then we expect fg to have comparable frequency to g. We also expect $D^\alpha(fg) \approx f(D^\alpha g)$, and $P_N(fg) \approx f(P_N g)$.*
- *(High-high interactions) If f and g have comparable frequency (e.g. $f = P_N F$ and $g = P_N G$ for some F, G) then fg should have frequency comparable or lower than f, and we expect $D^\alpha(fg) \lessapprox (D^\alpha f)g \approx f(D^\alpha g)$.*
- *(Full Leibnitz rule) With no frequency assumptions on f and g, we expect*

$$D^\alpha(fg) \approx f(D^\alpha g) + (D^\alpha f)g. \tag{A.14}$$

PRINCIPLE A.6 (Fractional chain rule). *Let u be a function on $\mathbf{R}^d$, and let $F : \mathbf{R} \to \mathbf{R}$ be a "reasonably smooth" function (e.g. $F(u) = |u|^{p-1}u$). Then we* *have the* fractional chain rule

$$D^\alpha(F(u)) \approx F'(u)D^\alpha u \tag{A.15}$$

for any differential operator D^α *of positive order* $\alpha > 0$*, as well as the Littlewood-Paley variants*

$$P_{<N}(F(u)) \approx F(P_{<N}u)$$

and

$$P_N(F(u)) \approx F'(P_{<N}u)P_N u. \tag{A.16}$$

If F *is complex instead of real, we have to replace* $F'(u)D^\alpha u$ *by* $F_z(u)D^\alpha u + F_{\overline{z}}\overline{D^\alpha u}$*, and similarly for* (A.16).

Observe that when D^α is a differential operator of order k, then the heuristics (A.14), (A.15) are accurate to top order in k (i.e. ignoring any terms which only differentiate f, g, u $k-1$ or fewer times). Indeed, the above two principles are instances of a more general principle:

PRINCIPLE A.7 (Top order terms dominate). *When distributing derivatives, the dominant terms are usually*[6] *the terms in which all the derivatives fall on a single factor; if the factors have unequal degrees of smoothness, the dominant term will be the one in which all the derivatives fall on the roughest (or highest frequency) factor.*

A complete and rigorous treatment of these heuristics (sometimes called *paradifferential calculus*) is beyond the scope of this text, and we refer the reader to [**Tay2**]. We will however give some representative instances of these heuristics in action.

LEMMA A.8 (Product lemma). *If* $s \geq 0$*, then we have the estimate*

$$\|fg\|_{H^s_x(\mathbf{R}^d)} \lesssim_{s,d} \|f\|_{H^s_x(\mathbf{R}^d)}\|g\|_{L^\infty_x(\mathbf{R}^d)} + \|f\|_{L^\infty_x(\mathbf{R}^d)}\|g\|_{H^s_x(\mathbf{R}^d)} \tag{A.17}$$

for all $f, g \in H^s_x(\mathbf{R}^d) \cap L^\infty_x(\mathbf{R}^d)$*. In particular, if* $s > d/2$*, we see from the Sobolev embedding* (A.13) *that we have the algebra property*

$$\|fg\|_{H^s_x(\mathbf{R}^d)} \lesssim_{s,d} \|f\|_{H^s_x(\mathbf{R}^d)}\|g\|_{H^s_x(\mathbf{R}^d)}. \tag{A.18}$$

Observe that (A.17) heuristically follows from (A.14), since that latter heuristic suggests that

$$\langle\nabla\rangle^s(fg) \approx (\langle\nabla\rangle^s f)g + f(\langle\nabla\rangle^s g)$$

and the claim then (non-rigorously) follows by taking L^2_x norms of both sides and then using the triangle and Hölder inequalities.

PROOF. The basic strategy with these multilinear estimates is to decompose using the Littlewood-Paley decomposition, eliminate any terms that are obviously zero (because of impossible frequency interactions), estimate each remaining component using the Bernstein and Hölder inequalities, and then sum. One should always try to apply Bernstein on the lowest frequency possible, as this gives the most efficient estimates. In some cases one needs to apply Cauchy-Schwarz to conclude the summation.

[6] In some cases, there is a special cancellation which allows one to treat the dominant terms directly. In such cases one often then has to look at the next term in the "Taylor expansion" in which all but one derivative falls on one term, and the remaining derivative falls on another. This phenomenon underlies a number of *commutator estimates*, such as those discussed in Section 3.9.

The claim is trivial for $s = 0$, so assume $s > 0$. From (A.9) we have

$$\|fg\|_{H^s_x(\mathbf{R}^d)} \lesssim_{s,d} \|P_{\leq 1}(fg)\|_{L^2_x(\mathbf{R}^d)} + (\sum_{N>1} N^{2s}\|P_N(fg)\|^2_{L^2_x(\mathbf{R}^d)})^{1/2}. \tag{A.19}$$

We shall just bound the latter term, and leave the former term to the exercises. We split[7]

$$\|P_N(fg)\|_{L^2_x(\mathbf{R}^d)} \lesssim \|P_N((P_{<N/8}f)g)\|_{L^2_x(\mathbf{R}^d)} + + \sum_{M>N/8} \|P_N((P_M f)g\|_{L^2_x(\mathbf{R}^d)}.$$

For the first term, observe from Fourier analysis that we may freely replace g by $P_{N/8<\cdot<8N}g$, and so by Hölder's inequality

$$\begin{aligned}\|P_N((P_{<N/8}f)g)\|_{L^2_x(\mathbf{R}^d)} &\lesssim_d \|(P_{<N/8}f)P_{N/8<\cdot<8N}g\|_{L^2_x(\mathbf{R}^d)}\\ &\lesssim_d \|f\|_{L^\infty_x(\mathbf{R}^d)} \sum_{M\sim N} \|P_M g\|_{L^2_x(\mathbf{R}^d)}\end{aligned}$$

and so the total contribution of this term to (A.19) is $O_{s,d}(\|f\|_{L^\infty_x(\mathbf{R}^d)}\|g\|_{H^s_x(\mathbf{R}^d)})$. For the second term, we simply bound

$$\begin{aligned}\sum_{M>N/8} \|P_N((P_M f)g)\|_{L^2_x(\mathbf{R}^d)} &\lesssim_d \sum_{M>N/8} \|(P_M f)g\|_{L^2_x(\mathbf{R}^d)}\\ &\lesssim_d \|g\|_{L^\infty_x(\mathbf{R}^d)} \sum_{M\gtrsim N} M^{-s}\|P_M f\|_{L^2_x(\mathbf{R}^d)}\end{aligned}$$

and so by Cauchy-Schwarz

$$(N^s \sum_{M>N/8} \|P_N((P_M f)g)\|_{L^2_x(\mathbf{R}^d)})^2 \lesssim_d \|g\|^2_{L^\infty_x(\mathbf{R}^d)} \sum_{M\gtrsim N} N^s M^{-s}\|P_M f\|^2_{L^2_x(\mathbf{R}^d)}.$$

Summing this in N (and using the hypothesis $s > 0$) we see that the total contribution of this term is $O_{s,d}(\|f\|_{H^s_x(\mathbf{R}^d)}\|g\|_{L^\infty_x(\mathbf{R}^d)})$, and we are done. □

Lemma A.9 (Schauder estimate). *Let V be a finite-dimensional normed vector space, let $f \in H^s_x(\mathbf{R}^d \to V) \cap L^\infty_x(\mathbf{R}^d \to V)$ for some $s \geq 0$. Let k be the first integer greater than s, and let $F \in C^k_{\mathrm{loc}}(V \to V)$ be such that $F(0) = 0$. Then $F(f) \in H^s_x(\mathbf{R}^d \to V)$ as well, with a bound of the form*

$$\|F(f)\|_{H^s_x(\mathbf{R}^d)} \lesssim_{F,\|f\|_{L^\infty_x(\mathbf{R}^d)},V,s,d} \|f\|_{H^s_x(\mathbf{R}^d)}.$$

Note that when F is real analytic, one can deduce this from Lemma A.8; but the argument below is rather robust and extends to rougher types of function F. For instance, when $s \leq 1$ the argument in fact only requires Lipschitz control on F. The reader should heuristically verify that Lemma A.9 follows immediately from Principle A.6 in much the same way that Lemma A.8 follows from Principle A.5. The reader may also wish to verify the $s = 1$ case of this estimate by hand (with F Lipschitz) in order to get some sense of why this type of estimate should hold.

[7]This is a basic example of a *paraproduct decomposition*, in which a genuine product such as fg is split as the sum of *paraproducts* (combinations of products of Littlewood-Paley pieces). Paraproducts are usually easier to estimate analytically, especially if derivatives are involved, because they specifically identify which of the factors is high frequency and which is low frequency, allowing one to use the flexible estimates (A.2)-(A.6) in a manner adapted to the paraproduct at hand, instead of relying only on "one-size-fits-all" tools such as Sobolev embedding.

PROOF. The strategy to prove nonlinear is related, though not quite the same as, that used to prove multilinear estimates. Basically, one should try to split $F(f)$ using Taylor expansion into a rough error, which one estimates crudely, and a smooth main term, which one estimates using information about its derivatives. Again, one uses tools such as Hölder, Bernstein, and Cauchy-Schwarz to estimate the terms that appear.

Let us write $A := \|f\|_{L^\infty_x(\mathbf{R}^d)}$. Since F is C^k_{loc} and $F(0) = 0$, we see that $|F(f)| \lesssim_{F,A,V} |f|$. This already establishes the claim when $s = 0$. Applying (A.9), it thus suffices to show that

$$(\sum_{N>1} N^{2s} \|P_N F(f)\|^2_{L^2_x(\mathbf{R}^d)})^{1/2} \lesssim_{F,A,V,s} \|f\|_{H^s_x(\mathbf{R}^d)}.$$

for all $s > 0$.

We first throw away a "rough" portion of $F(f)$ in $P_N F(f)$. Fix N, s, and split $f = P_{<N} f + P_{\geq N} f$. Note that f and $P_{<N} f$ are both bounded by $O_{V,d}(A)$. Now F is C^k_{loc}, hence Lipschitz on the ball of radius $O_{V,d}(A)$, hence we have

$$F(f) = F(P_{<N} f) + O_{F,A,V,d}(|P_{\geq N} f|),$$

and thus

$$\|P_N F(f)\|_{L^2_x(\mathbf{R}^d)} \lesssim_{F,A,V,d} \|P_N F(P_{<N} f)\|_{L^2_x(\mathbf{R}^d)} + \|P_{\geq N} f\|_{L^2_x(\mathbf{R}^d)}.$$

To control the latter term, observe from the triangle inequality and Cauchy-Schwarz that

$$N^{2s} \|P_{\geq N} f\|^2_{L^2_x(\mathbf{R}^d)} \lesssim_s \sum_{N' \geq N} (N')^s N^s \|P_{N'} f\|^2_{L^2_x(\mathbf{R}^d)}$$

and summing this in N and using (A.9) we see that this term is acceptable. Thus it remains to show that

$$(\sum_{N>1} N^{2s} \|P_N F(P_{<N} f)\|^2_{L^2_x(\mathbf{R}^d)})^{1/2} \lesssim_{F,A,V,s} \|f\|_{H^s_x(\mathbf{R}^d)}.$$

We will exploit the smoothness of $P_{<N} f$ and F by using (A.4) to estimate

$$\|P_N F(P_{<N} f)\|_{L^2_x(\mathbf{R}^d)} \lesssim_{d,k} N^{-2k} \|\nabla^k F(P_{<N} f)\|_{L^2_x(\mathbf{R}^d)}. \tag{A.20}$$

Applying the chain rule repeatedly, and noting that all derivatives of F are bounded on the ball of radius $O_{V,d}(A)$, we obtain the pointwise estimate

$$|\nabla^k F(P_{<N} f)| \lesssim_{F,A,V,d,k} \sup_{k_1 + \ldots + k_r = k} |\nabla^{k_1}(P_{<N} f)| \ldots |\nabla^{k_r}(P_{<N} f)|$$

where r ranges over $1, \ldots, k$ and $k_1, \ldots, k_r$ range over non-negative integers that add up to k. We split this up further using Littlewood-Paley decomposition as

$$|\nabla^k F(P_{<N} f)| \lesssim_{F,A,V,d,k} \sup_{k_1 + \ldots + k_r = k} \sum_{1 \leq N_1, \ldots, N_r < N} |\nabla^{k_1}(\tilde{P}_{N_1} f)| \ldots |\nabla^{k_r}(\tilde{P}_{N_r} f)|$$

where we adopt the convention that $\tilde{P}_N := P_N$ when $N > 1$ and $\tilde{P}_1 := P_{\leq 1}$. By giving up a factor of $r! = O_k(1)$ we may take $N_1 \leq N_2 \ldots \leq N_r$. where $k_1, \ldots, k_r$ range over all positive integers that add up to k. Now from (A.4) we have

$$\|\nabla^{k_i}(\tilde{P}_{N_i} f)\|_{L^\infty_x(\mathbf{R}^d)} \lesssim_{d,k} N_i^{k_i} \|f\|_{L^\infty_x(\mathbf{R}^d)} \lesssim_{d,k,A} N^{k_i}$$

for $i = 1, \ldots, r-1$, and similarly

$$\|\nabla^{k_r}(\tilde{P}_{N_r} f)\|_{L^\infty_x(\mathbf{R}^d)} \lesssim_{d,k} N_r^{k_r} \|\tilde{P}_{N_r} f\|_{L^2_x(\mathbf{R}^d)}$$

and hence we have

$$\|\nabla^k F(P_{<N} f)\|_{L^2_x(\mathbf{R}^d)} \lesssim_{F,A,V,d,k} \sup_{k_1+\ldots+k_r=k} \sum_{1\leq N_1\leq\ldots\leq N_r<N} N_1^{k_1}\ldots N_r^{k_r}\|\tilde P_{N_r} f\|_{L^2_x(\mathbf{R}^d)}.$$

Performing the sum in N_1, then N_2, then finally N_{r-1}, and rewriting $N' := N_r$, we obtain

$$\|\nabla^k F(P_{<N} f)\|_{L^2_x(\mathbf{R}^d)} \lesssim_{F,A,V,d,k} \sum_{1\leq N'<N} (N')^k \|\tilde P_{N'} f\|_{L^2_x(\mathbf{R}^d)}.$$

By Cauchy-Schwarz and (A.20) we conclude

$$\|P_N F(P_{<N} f)\|_{L^2_x(\mathbf{R}^d)} \lesssim_{F,A,V,d,k} \sum_{1\leq N'<N} (N')^k N^{-k} \|\tilde P_{N'} f\|^2_{L^2_x(\mathbf{R}^d)}.$$

Summing this in N and using (A.9) we see that this term is acceptable (note that k depends only on s, so the dependence on k is not a concern). □

In computations involving momentum, one often encounters expressions such as $\int_{\mathbf{R}^d} u\nabla v\ dx$ or $\int_{\mathbf{R}^d} u(x)(\frac{x}{|x|}\cdot\nabla v)$. The following lemma is useful for controlling these quantities.

Lemma A.10 (Momentum estimate). *Let $u, v \in \mathcal{S}(\mathbf{R}^d)$ for some $d \geq 3$, and let K be a kernel on $\mathbf{R}^d$ which is smooth away from the origin, and obeys the estimates*

$$|K(x)| \lesssim_d 1; \quad |\nabla K(x)| \lesssim_d |x|^{-1}$$

away from the origin. (For instance, we could have $K(x) \equiv 1$, or $K(x) \equiv \frac{x_j}{|x|}$ for some $j = 1, \ldots, d$.) Then we have

$$|\int_{\mathbf{R}^d} u(x)K(x)\nabla v(x)\ dx| \lesssim_d \|u\|_{\dot H^s_x(\mathbf{R}^d)}\|v\|_{\dot H^{1-s}_x(\mathbf{R}^d)}$$

for all $0 \leq s \leq 1$ (in particular, the estimate is true for $s = 1/2$).

Intuitively speaking, the justification for this lemma is that we can integrate by parts "1/2 times" to move half of the derivative from v onto u, ignoring the mild symbol K in between. By standard limiting arguments we may now extend the bilinear form $(u, v) \mapsto \int_{\mathbf{R}^d} u(x)K(x)\nabla v(x)\ dx$ to all $u, v \in \dot H^{1/2}_x(\mathbf{R}^d)$, dropping the hypothesis that u, v is Schwartz.

Proof. A standard regularisation argument (replacing K by $K * \phi_\varepsilon$ for some approximation to the identity ϕ_ε, and then letting $\varepsilon \to 0$, taking advantage of the hypothesis that u, v are Schwartz) allows us to assume that K is smooth on all of $\mathbf{R}^d$ (including the origin), provided of course that our estimates are uniform in K. By real or complex interpolation it will suffice to establish the estimates

$$|\int_{\mathbf{R}^d} u(x)K(x)\nabla v(x)\ dx| \lesssim_d \|u\|_{L^2_x(\mathbf{R}^d)}\|v\|_{\dot H^1_x(\mathbf{R}^d)}$$

and

$$|\int_{\mathbf{R}^d} u(x)K(x)\nabla v(x)\ dx| \lesssim_d \|u\|_{\dot H^1_x(\mathbf{R}^d)}\|v\|_{L^2_x(\mathbf{R}^d)}.$$

The first estimate is immediate from Hölder's inequality (estimating $u, \nabla v$ in L^2_x and K in L^∞_x). For the second estimate, we integrate by parts (again taking advantage of the hypothesis that u, v are Schwartz) and use the triangle inequality to estimate

$$|\int_{\mathbf{R}^d} u(x)K(x)\nabla v(x)\,dx| \leq |\int_{\mathbf{R}^d} (\nabla u)(x)K(x)v(x)\,dx| + |\int_{\mathbf{R}^d} u(x)(\nabla K)(x)v(x)\,dx|.$$

The first term can be estimated by Hölder's inequality as before. The second term can be estimated by Cauchy-Schwarz (placing v in L^2_x) followed by Lemma A.2 (with $s = 1$ and $d \geq 3$), and we are done. □

Exercise A.1. Prove (A.2)-(A.6). (Hint: for each of the estimates, use Fourier analysis to write the expression in the left-hand norm as the convolution of the expression in the right-hand norm with some explicit kernel, and then use Young's inequality.) Discuss why these estimates are consistent with Principle A.1.

Exercise A.2. Deduce (A.13) directly from the Fourier inversion formula and Cauchy-Schwarz, and show that it fails at the endpoint $s = d/2$.

Exercise A.3 (Lorentz characterisation of L^p_x). [**KTao**] Let $f \in L^p_x(\mathbf{R}^d)$ for some $1 < p < \infty$. Show that one can decompose $f = \sum_k c_k \chi_k$, where k ranges over the integers, χ_k is a function bounded in magnitude by 1 and supported on a set of measure at most 2^k, and c_k are a sequence of non-negative reals such that $(\sum_k 2^k |c_k|^p)^{1/p} \sim_p \|f\|_{L^p_x(\mathbf{R}^d)}$. (Hint: let $f^*(x) := \inf\{\alpha : |\{|f| > \alpha\}| < x\}$ be the (left-continuous) nondecreasing rearrangement of $|f|$. Set c_k to equal $f^*(2^{k-1})$, and $c_k\chi_k$ be the portion of f where $f^*(2^k) < |f(x)| \leq f^*(2^{k-1})$.)

Exercise A.4 (Dual Lorentz characterisation of L^q_x). Let $f \in L^q_x(\mathbf{R}^d)$ for some $1 < q < \infty$. Show that

$$\|f\|_{L^q_x(\mathbf{R}^d)} \sim_q \sup_{E_k} (\sum_k 2^{k(1-q)} |\int_{E_k} f(x)\ dx|^q)^{1/q}$$

where for each k, E_k ranges over all bounded open sets of measure 2^k. (Hint: use the nondecreasing rearrangement again. Show that $\sup_{E_k} |\int_{E_k} f(x)\,dx| \sim_q \int_0^{2^k} f^*(t)\,dt$, and then decompose the interval $[0, 2^k]$ dyadically.)

Exercise A.5. Use Exercises A.3, A.4 to prove (A.10). (Hint: first establish the estimate

$$|\int_{E_{k'}} \chi_k * \frac{1}{|x|^\alpha}| \lesssim_{d,\alpha} \min(2^{\frac{d-\alpha}{d}k} 2^{k'}, 2^{\frac{d-\alpha}{d}k'} 2^k)$$

for all k, k', where $E_{k'}$ and χ_k are as in the preceding exercises.) Deduce (A.11) and (A.12) as a consequence.

Exercise A.6 (Lorentz refinement of Sobolev embedding). For $1 < p, q < \infty$, define the *Lorentz norm*

$$\|f\|_{L^{q,p}_x(\mathbf{R}^d)} := \sup_{E_k} (\sum_k 2^{k(\frac{p}{q} - p)} |\int_{E_k} f(x)\ dx|^p)^{1/p}$$

where E_k is as in Exercise A.4. By repeating the proof of Exercise A.5, refine the estimate (A.11) to

$$\|f\|_{L^q_x(\mathbf{R}^d)} \lesssim_{p,q,d} \|f\|_{L^{q,p}_x(\mathbf{R}^d)} \lesssim_{p,q,d} \|f\|_{\dot W^{s,p}_x(\mathbf{R}^d)}$$

under the same hypotheses on p, q, d, s.

Exercise A.7. Prove the first half of Proposition A.4. (Hint: first use Exercise A.6 to show that $|\int_{E_k} f(x)\, dx| \sim_{p,q,d,\eta} 2^{k(1-1/q)}$ for some $k \in \mathbf{Z}$ and some set E_k of measure 2^k. Then perform a Littlewood-Paley decomposition of f to conclude that $\|P_N f\|_{L^\infty_x(\mathbf{R}^d)} \gtrsim_{p,q,d,\eta} 2^{k(1-1/q)}$ for some $N \sim_{p,q,d,\eta} 2^{-k/d}$.)

Exercise A.8 (Relationship between Sobolev and isoperimetric inequalities). Prove the endpoint Sobolev estimate

$$|\{|f(x)| \geq \lambda\}| \lesssim_d \frac{\||\nabla f|\|^d_{L^1_x(\mathbf{R}^d)}}{\lambda^d}$$

for any $\lambda > 0$ and $f \in \mathcal{S}_x(\mathbf{R}^d)$. (Hint: estimate $|f(x)|$ pointwise by $|\nabla f(x)| * \frac{1}{|x|^{d-1}}$. Let $E := \{|f(x)| \geq \lambda\}$ and obtain a pointwise bound for $1_E * \frac{1}{|x|^{d-1}}$.) If $\Omega \subset \mathbf{R}^d$ is a bounded domain with smooth boundary, deduce the *isoperimetric inequality*

$$|\partial\Omega| \gtrsim_d |\Omega|^{(d-1)/d}$$

where $|\partial\Omega|$ is the surface area of Ω. (Hint: set f to be a smoothed out version of 1_Ω.) It is well known that among all domains with fixed volume, the ball has the smallest surface area; comment on how this is compatible with the heuristics supporting Proposition A.4.

Exercise A.9. Give a heuristic justification of Principle A.5 using the Fourier transform and the elementary estimate $\langle \xi + \eta \rangle^s \lesssim_s \langle \xi \rangle^s + \langle \eta \rangle^s$ for all $\xi, \eta \in \mathbf{R}^d$.

Exercise A.10. Complete the proof of Lemma A.8.

Exercise A.11. Generalise Lemma A.8 by replacing H^s_x with $W^{s,p}_x$ for some $1 < p < \infty$, and replacing the condition $s > d/2$ with $s > d/p$. (Hint: you will need the Littlewood-Paley estimate (A.7).)

Exercise A.12. Let the assumptions and notation be as in Lemma A.9, but suppose that F lies in C^{k+1}_{loc} rather than just in C^k_{loc}. Establish the Lipschitz estimate

$$\|F(f) - F(g)\|_{H^s_x(\mathbf{R}^d)} \lesssim_{F, \|f\|_{L^\infty_x(\mathbf{R}^d)}, \|g\|_{L^\infty_x(\mathbf{R}^d)}, V, s, d} \|f - g\|_{H^s_x(\mathbf{R}^d)}.$$

(Hint: One could repeat the proof of Lemma A.9, but a slicker proof is to use the fundamental theorem of calculus to write $F(f) - F(g) = \int_0^1 DF((1-\theta)f + \theta g) \cdot (f - g)\, d\theta$, where DF is the differential of F, and then apply Lemma A.9 and Lemma A.8.)

Exercise A.13 (Fractional chain rule). [**CWein**] Let $p > 1$, and let $F \in C^1_{\mathrm{loc}}(\mathbf{C} \to \mathbf{C})$ be a function of p^{th} power type, in the sense that $F(z) = O(|z|^p)$ and $\nabla F(z) = O_p(|z|^{p-1})$. Let $0 \leq s < 1$ and $1 < q < r < \infty$ obey the scaling condition $\frac{d}{q} = \frac{dp}{r} - (p-1)s$. Show that

$$\|F(f)\|_{W^{s,q}_x(\mathbf{R}^d)} \lesssim_{d,p,q,r,s} \|f\|^p_{W^{s,r}_x(\mathbf{R}^d)}$$

for all $f \in W^{s,r}_x(\mathbf{R}^d)$. (Note that this is rather easy to justify heuristically from Principle A.6.) If furthermore $p > 2$, and F is C^2_{loc} with $\nabla^2 F(z) = O_p(|z|^{p-2})$, establish the stronger estimate

$$\|F(f) - F(g)\|_{W^{s,q}_x(\mathbf{R}^d)} \lesssim_{d,p,q,r,s} (\|f\|_{W^{s,r}_x(\mathbf{R}^d)} + \|g\|_{W^{s,r}_x(\mathbf{R}^d)})^{p-1} \|f - g\|_{W^{s,r}_x(\mathbf{R}^d)}$$

for all $f, g \in W^{s,r}_x(\mathbf{R}^d)$.

EXERCISE A.14. If I is an interval in $\mathbf{R}$ and $2 \leq q, r < \infty$, establish the inequality

$$\|u\|_{L^q_t L^r_x(I \times \mathbf{R}^d)} \lesssim_{d,q,r} (\sum_N \|P_N u\|^2_{L^q_t L^r_x(I \times \mathbf{R}^d)})^{1/2}.$$

(Hint: use (A.7). To interchange the norm and square function, first consider the extreme cases $q, r = 2, \infty$ and then interpolate, for instance using the complex method.)

EXERCISE A.15. If I is a bounded interval in $\mathbf{R}$, and $u, \partial_t u \in L^2_t(I)$, use elementary arguments to obtain a localised Gagliardo-Nirenberg inequality

$$\|u\|_{L^\infty_t(I)} \lesssim \|u\|^{1/2}_{L^2_t(I)} \|\partial_t u\|^{1/2}_{L^2_t(I)}$$

and the Poincaré inequality

$$\|u - \frac{1}{I}\int_I u\|_{L^2_t(I)} \lesssim |I| \|\partial_t u\|_{L^2_t(I)}.$$

EXERCISE A.16 (Hardy inequality revisited). Let $u \in \mathcal{S}_x(\mathbf{R}^d)$. Use integration by parts to establish the identity

$$\int_{\mathbf{R}^d} |x|^\alpha |x \cdot \nabla u(x)|^2 \, dx = \frac{(d+\alpha)^2}{4} \int_{\mathbf{R}^d} |x|^\alpha |u(x)|^2 \, dx + \int_{\mathbf{R}^d} |x|^\alpha |x \cdot \nabla u(x) - \frac{d+\alpha}{2} u(x)|^2 \, dx$$

for any $\alpha > -d$, and use this to establish another proof of Hardy's inequality in the case $s = 1$.

EXERCISE A.17. Give another proof of Lemma A.10 which does not use interpolation, but relies instead on Littlewood-Paley decomposition. (Hint: you may need to decompose K smoothly into dyadic pieces also and use arguments similar to those used to prove Lemma A.2.) The techniques of interpolation and of dyadic decomposition are closely related; the latter tends to be messier but more flexible.

EXERCISE A.18 (Localisation of H^s_x functions). Let $u \in H^s_x(\mathbf{R}^d)$ for some $s \geq 0$, and let $\psi \in \mathcal{S}_x(\mathbf{R}^d)$. Show that for any $R \geq 1$ we have

$$\|u(x)\psi(\frac{x}{R})\|_{H^s_x(\mathbf{R}^d)} \lesssim_{s,d,\psi} \|u\|_{H^s_x(\mathbf{R}^d)}.$$

This very useful fact allows one to smoothly localise functions in H^s_x to large balls, *uniformly in the size of the ball.* (Hint: prove this for s a positive integer by induction first, and then use interpolation. You may find the Hardy or Sobolev inequalities to be useful. An alternate approach is to perform a Fourier decomposition of ψ and work entirely in frequency space.) Similarly, if $u \in \dot H^1_x(\mathbf{R}^d)$ and $d \geq 3$, establish the bound

$$\|u(x)\psi(\frac{x}{R})\|_{\dot H^1_x(\mathbf{R}^d)} \lesssim_{\psi,d} \|u\|_{\dot H^1_x(\mathbf{R}^d)}.$$

EXERCISE A.19 (Radial Sobolev inequality). Let $d \geq 3$, and let u be a Schwartz function on $\mathbf{R}^d$. Establish the inequality

$$\||x|^{\frac{d}{2}-1}|u|\|_{L^\infty_x(\mathbf{R}^d)} \lesssim_d \|u\|_{\dot H^1_x(\mathbf{R}^d)}$$

for all $x \in \mathbf{R}^d$, as well as the variant

$$\||x|^s|u|\|_{L^\infty_x(\mathbf{R}^d)} \lesssim_{d,s} \|u\|_{H^1_x(\mathbf{R}^d)}$$

for all $\frac{d}{2} - 1 \leq s \leq \frac{d-1}{2}$. (Hint: if $|x| = R$, truncate u smoothly to the region $|x| \sim R$ using Exercise A.18, use polar coordinates, and use the Gagliardo-Nirenberg inequality.)

EXERCISE A.20. If f is spherically symmetric, show that one can take $x_0 = O_{s,p,q,d,\eta}(1/N)$ in Proposition A.4; thus Sobolev embedding is only sharp near the origin (using the natural length scale associated to the frequency). (Hint: if x_0 is too far away from the origin, use the symmetry to find a large number of disjoint balls, each of which absorb a significant portion of energy.)

EXERCISE A.21 (Littlewood-Paley characterisation of Hölder regularity). Let $0 < \alpha < 1$ and $1 \leq p \leq \infty$. If $f \in \mathcal{S}_x(\mathbf{R}^d)$, we define the *Hölder norm* $\|f\|_{\Lambda^p_\alpha(\mathbf{R}^d)}$ by the formula

$$\|f\|_{\Lambda^p_\alpha(\mathbf{R}^d)} := \|f\|_{L^p_x(\mathbf{R}^d)} + \sup_{h \in \mathbf{R}^d : 0 < |h| \leq 1} \frac{\|f^h - f\|_{L^p_x(\mathbf{R}^d)}}{|h|^\alpha}$$

where $f^h(x) = f(x+h)$ is the translate of f by h. Show that

$$\|f\|_{\Lambda^p_\alpha(\mathbf{R}^d)} \sim_{p,\alpha,d} \|f\|_{L^p_x(\mathbf{R}^d)} + \sup_{N \geq 1} N^\alpha \|P_N f\|_{L^p_x(\mathbf{R}^d)}.$$

(Hint: to control the latter by the former, express $P_N f$ as an average of functions of the form $f^h - f$. To control the former by the latter, obtain two bounds for the L^p_x norm of $P_N f^h - P_N f$, using the triangle inequality in the high frequency case $N \gtrsim |h|^{-1}$ and the fundamental theorem of calculus in the low frequency case $N \lesssim |h|^{-1}$.) The latter expression is essentially an example of a *Besov norm*, which often functions as a substitute for the Sobolev norm which is a little more technically convenient in several PDE applications, particularly those in which one is concerned about controlling interactions between high and low frequencies. Conclude in particular that

$$\|f\|_{W^{\alpha-\varepsilon,p}_x(\mathbf{R}^d)} \lesssim_{p,\alpha,d,\varepsilon} \|f\|_{\Lambda^p_\alpha(\mathbf{R}^d)} \lesssim_{p,\alpha,d} \|f\|_{W^{\alpha,p}_x(\mathbf{R}^d)}$$

for any $\varepsilon > 0$; thus Hölder norms are "within an epsilon" of their Sobolev counterparts.

EXERCISE A.22 (Morrey-Sobolev inequality). If $0 < \alpha < 1$ and $d < p \leq \frac{d}{1-\alpha}$, show that

$$\|f\|_{\Lambda^\infty_\alpha(\mathbf{R}^d)} \lesssim_{p,\alpha,d} \|f\|_{W^{1,p}_x(\mathbf{R}^d)}$$

for all $f \in \mathcal{S}_x(\mathbf{R}^d)$, where Λ^∞_α was defined in the previous exercise. This reflects a general principle, that if there is some "surplus" regularity in the Sobolev embedding theorem that causes one to go past L^∞_x, this additional regularity will then manifest itself as Hölder continuity, and one can again recover endpoint estimates.

EXERCISE A.23 (Hodge decomposition). Let $\phi : H^s_x(\mathbf{R}^d \to \mathbf{R}^d)$ be a vector field. Show that one has a unique decomposition $\phi = \phi^{\mathrm{cf}} + \phi^{\mathrm{df}}$ into a curl-free vector field $\phi^{\mathrm{cf}} \in H^s_x(\mathbf{R}^d \to \mathbf{R}^d)$ and a divergence-free vector field $\phi^{\mathrm{df}} \in H^s_x(\mathbf{R}^d \to \mathbf{R}^d)$, thus $\mathrm{curl}\phi^{\mathrm{cf}} = \nabla \wedge \phi^{\mathrm{cf}} = 0$ and $\mathrm{div}\phi^{\mathrm{df}} = \nabla \cdot \phi^{\mathrm{df}} = 0$ in the sense of distributions. Verify the identities $\phi^{\mathrm{cf}} = \Delta^{-1}\nabla(\nabla \cdot \phi)$ and $\phi^{\mathrm{df}} = \Delta^{-1}\nabla\neg(\nabla \wedge \phi)$. If $s = 0$, show that ϕ^{cf} and ϕ^{df} are orthogonal. (You may either use the Fourier transform, or take divergences and curls of the decomposition $\phi = \phi^{\mathrm{cf}} + \phi^{\mathrm{df}}$ to solve for ϕ^{cf} and ϕ^{df}.)

EXERCISE A.24 (Div-curl lemma). Let $\phi, \psi : L^2_x(\mathbf{R}^d \to \mathbf{R}^d)$ be vector fields such that $\mathrm{div}\phi = 0$ and $\mathrm{curl}\psi = 0$. Show that $\int_{\mathbf{R}^d} \phi \cdot \psi = 0$, and also that $\phi \cdot \psi \in \dot H^{-d/2}_x(\mathbf{R}^d)$; this is a simple example of a *div-curl lemma*, that exploits a certain "high-high" frequency cancellation between divergence-free and curl-free vector fields, and forms a key component of the theory of *compensated compactness*; see for instance [**CLMS**]. (Note that Hölder's inequality would place $\phi \cdot \psi$ in L^1_x, which is not enough for Sobolev embedding to place into $\dot H^{-d/2}_x 3$. To prove the lemma, use Hodge theory to write ϕ and ψ as the curl and gradient of a $\dot H^1_x$ 2-form and scalar field respectively, then use Littlewood-Paley decomposition.)

EXERCISE A.25 (Sobolev trace lemma). Let $f \in \mathcal{S}_x(\mathbf{R}^d)$ for $d \geq 2$, and view $\mathbf{R}^{d-1} \equiv \mathbf{R}^{d-1} \times \{0\}$ as a subset of $\mathbf{R}^d$ in the usual manner. Show that

$$\|f\|_{\dot H^s_x(\mathbf{R}^{d-1})} \lesssim_{d,s} \|f\|_{\dot H^{s+1/2}_x(\mathbf{R}^d)}$$

and

$$\|f\|_{H^s_x(\mathbf{R}^{d-1})} \lesssim_{d,s} \|f\|_{H^{s+1/2}_x(\mathbf{R}^d)}$$

for all $s > 0$. (This can be done from the Fourier transform; it is also worthwhile to try to prove this from Littlewood-Paley theory, following the heuristics in Principle A.1.) Show that these estimates fail when $s = 0$, and also that the loss of $1/2$ a derivative cannot be reduced. One can of course generalise this lemma to other subsets of $\mathbf{R}^d$ of various codimension, and other Sobolev spaces, but we shall not do so here.

EXERCISE A.26 (Agmon division lemma). Let $f \in \mathcal{S}_x(\mathbf{R})$ be such that $f(0) = 0$. Show that $\|f(x)/x\|_{H^{s-1}_x(\mathbf{R})} \lesssim_s \|f\|_{H^s_x(\mathbf{R})}$ for all $s > 1/2$. (Hint: write $f(x)/x = \int_0^1 f'(tx)\, dt$, then take Fourier transforms.)