# LOW-REGULARITY GLOBAL SOLUTIONS TO NONLINEAR DISPERSIVE EQUATIONS

TERENCE TAO

ABSTRACT. In these notes we discuss some recent developments in the study of global solutions to dispersive equations, especially for low regularity data.

## 1. INTRODUCTION

These notes are concerned with a certain class of PDE which generally fall under the title of non-linear dispersive equations. A typical such equation is the *cubic non-linear Schrodinger equation* (NLS)

$$iu_t + \Delta u = |u|^2 u,$$

where $u(x,t)$ is a complex-valued function of space $x \in \mathbf{R}^n$ and time $t \in \mathbf{R}$; this comes up in many physical situations such as modelling dilute Bose-Einstein condensates. Another example is the *Korteweg-de Vries* (KdV) equation

$$u_t + u_{xxx} = uu_x$$

where $u(x,t)$ is now a real-valued function of one space dimension $x \in \mathbf{R}$ and one time dimension $t \in \mathbf{R}$; this equation most famously models shallow waves in a canal. A third example is the *wave map equation* for the sphere

$$\Box u = -u_{tt} + \Delta u = -u(|\nabla u|^2 - |u_t|^2)$$

where $u(x,t)$ is now a function of $\mathbf{R}^n \times \mathbf{R}$ taking values in a sphere $S^{m-1} \subset \mathbf{R}^m$. This equation is the natural Minkowskian analogue of the harmonic map equation, and is also related to simplified models of Einsteinian gravity.

Although these equations are quite different, they share many features in common:

• They are evolution equations, involving a time parameter $t$. Because of this, the natural PDE problem to solve is the Cauchy problem, in which the initial position $u(x,0)$ is specified. For equations such as

1991 *Mathematics Subject Classification.* 42B15, 35L05.

the wave map equation which are second order in time, the initial velocity $u_t(x,0)$ must also be specified.

• They are all Hamiltonian equations. Recall that a Hamiltonian ODE flow on a phase space $\mathbf{R}^{2N}$ is any equation of the form

$$\dot{u} = J\nabla H(u)$$

where $H : \mathbf{R}^{2N} \to \mathbf{R}$ is the Hamiltonian function, and $J$ is the matrix of the symplectic form (so that the quadratic form $\langle Jx, y\rangle$ is anti-symmetric and non-degenerate, e.g. $J = \begin{pmatrix} 0 & I_N \\ -I_N & 0 \end{pmatrix}$). A Hamiltonian PDE is defined similarly except that the phase space is now an infinite dimensional Hilbert space such as $L^2(\mathbf{R}^n)$.

For NLS, the phase space is $L^2(\mathbf{R}^n; \mathcal{C})$ with symplectic form

$$\langle Ju, v\rangle := \mathrm{Im} \int u\overline{v}\, dx$$

and Hamiltonian

$$H(u) := \int \frac{1}{2}|\nabla u|^2 + \frac{1}{4}|u|^4.$$

For KdV, the phase space is (formally, at least) the homogeneous Sobolev space $\dot{H}^{-1/2}(\mathbf{R}; \mathbf{R})$ with symplectic form

$$\langle Ju, v\rangle := \int u(\frac{d}{dx})^{-1} v\, dx$$

and Hamiltonian

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

For wave maps, the phase space is $\dot{H}^{1/2}(\mathbf{R}^n; S^{m-1}) \times \dot{H}^{-1/2}(\mathbf{R}^n; S^{m-1})$ (with the phase space location at time $t$ given by $(u(t), u_t(t))$) with symplectic form[1]

$$\langle J(u, u_t), (v, v_t)\rangle := \int uv_t - vu_t dx$$

and Hamiltonian

$$H(u, u_t) := \int \frac{1}{2}|\nabla u|^2 - \frac{1}{2}|u_t|^2.$$

[1]This is not quite accurate because phase space is not linear here, but if one restricts the symplectic form to the non-linear phase space then these statements become valid.

Hamiltonian equations have several good features. To begin with, the Hamiltonian $H(u)$ is (formally, at least) an invariant of the flow:

$$\partial_t H(u(t)) = 0.$$

In most physical situations the Hamiltonian represents energy. More generally, any quantity which Poisson commutes with the Hamiltonian will be an invariant of the flow; for the NLS and KdV equations the $L^2$ norm $\int |u|^2$ is a typical example. Some equations, such as KdV (or NLS in one dimension) are fortunate to have a maximal number of independent functions which Poisson commute with the Hamiltonian; for finite dimensional phase space $\mathbf{R}^{2N}$ the maximal number is $N$, but for infinite dimensions the maximal number of such functions is of course infinite. Such equations are known as *completely integrable equations* and enjoy infinitely many conserved quantities. For instance, KdV has the conserved quantities

$$\int u$$
$$\int u^2$$
$$\int u_x^2 - \frac{2}{3}u^3$$
$$\int u_{xx}^2 - \frac{5}{3}uu_x^2 + \frac{5}{9}u^4$$

etc. However, complete integrability is a very special phenomenon, mostly restricted to one space dimension, and should not be expected in general.

Hamiltonian equations are also time reversible (unlike dissipative equations such as the heat equation).

• These equations are invariant under translations in both time and space. This suggests that the Fourier transform will be a useful tool; this is already apparent in our choice of using the Sobolev spaces $H^s$. Actually there are two Fourier transforms which are useful: the *spatial Fourier transform*

$$\hat{u}(\xi, t) := \int e^{-ix\cdot\xi} u(x,t)\ dx$$

and the *spacetime* Fourier transform

$$\tilde{u}(\xi, \tau) := \int\int e^{-i(x\cdot\xi + t\tau)} u(x,t)\ dxdt.$$

The spatial Fourier transform decomposes a function $u(x)$ into various plane waves $e^{ix\cdot\xi}$ of various frequencies $\xi$. It is quite profitable to think

of the evolution of the PDE in terms of these various frequency modes. For instance, the KdV equation

$$u_t + u_{xxx} = uu_x$$

becomes (on the Fourier side)

$$\partial_t \hat{u}(\xi,t) = i\xi^3 \hat{u}(\xi,t) + \int_{\xi_1+\xi_2=\xi} i\xi_2 \hat{u}(\xi_1,t)\hat{u}(\xi_2,t).$$

The first term causes $\hat{u}(\xi,t)$ to rotate in phase with angular velocity $\xi^3$. The second term is the non-linear frequency interaction term, causing Fourier modes at $\xi_1$ and $\xi_2$ to influence the evolution of the Fourier mode at $\xi$.

As we shall see, it is also profitable to group the Fourier modes into dyadic bands $|\xi| \sim N$, where $N$ is a power of two. (In harmonic analysis this technique falls under the domain of *Littlewood-Paley theory*, although usually we do not need the full power of this theory for PDE applications).

• While these equations are non-linear, they are only quite mild perturbations of linear equations. More precisely, they are *semi-linear*, meaning that they are all of the form

$$Lu = F(u)$$

where $L$ is a linear evolution operator and $F(u)$ is a purely non-linear expression (i.e. only containing quadratic or higher terms) which is lower order than $L$. In the above model equations the linear part $L$ is very simple (either the Schrodinger operator $i\partial_t + \Delta$, the Airy operator $\partial_t + \partial_{xxx}$, or the d'Alambertian (or wave operator) $\Box = -\partial_t^2 + \Delta$). This is in contrast to modern *linear* PDE theory, in which one does not add a non-linearity $F(u)$ but instead considers much more general linear operators $L$. For the most part, research in these two areas have gone in quite different directions, but perhaps they might be more strongly unified in the future.

Because of this semilinear structure, we expect the linear component $L$ to be somehow "dominant", especially if $u$ is small. (When $u$ is small, $Lu$ is also small, but $F(u)$ is even smaller, being purely non-linear). If $u$ is large, we still expect $L$ to dominate for short times (because the non-linearity, being a forcing term, requires some time to significantly influence the linear evolution of the initial data).

• The linear term $L$ is *dispersive*, which means that different frequency components of the linear evolution $Lu = 0$ move in different

directions. For instance, for the free Schrödinger equation $iu_t + \Delta u = 0$, a wave packet initial data of

$$u(x,0) = e^{i\xi\cdot x} a(x)$$

with $\xi$ a high frequency and $a$ a bump function, will roughly evolve like

$$u(x,t) \approx e^{i\xi\cdot x} e^{-i|\xi|^2 t} a(x - 2\xi t)$$

(at least if $\Delta a$ is negligible). Thus, apart from some phase oscillation, a wave packet with frequency $\xi$ will travel at velocity $2\xi$, and retain its frequency. In particular, wave packets with distinct frequencies will ultimately disperse from each other. The situation is similar for the Airy equation $u_t + u_{xxx} = 0$ except that the velocity is $-3\xi^2$ rather than $2\xi$. For the wave equation $\Box u$ the velocity is $\xi/|\xi|$; this is not quite as dispersive as the other two equations because the map $\xi \to \xi/|\xi|$ is not injective, but still has a significant amount of dispersion.

These phenomena can also be seen on the Fourier level. If $u(x,t)$ is a global solution to the free Schrödinger equation $iu_t + \Delta u = 0$, then its spacetime Fourier transform $\tilde{u}$ is (formally at least) supported on the paraboloid $\tau = |\xi|^2$. Similarly solutions to the free Airy equation are supported on the cubic $\tau = \xi^3$, while solutions to the free wave equation are supported on the cone $|\tau| = |\xi|$. The above wave packet analysis can then be viewed as a stationary phase computation of the Fourier transform of these characteristic surfaces. (The Fourier transform is useful here because these equations are translation-invariant).

For the non-linear equation $Lu = F(u)$ the above statements do not hold exactly, but we still expect them to hold approximately, especially if $F$ is somewhat weak, the analysis is over short times only, and $u$ is not too wild. In particular, if the initial data has a frequency component in the region $|\xi| \sim N$, we expect this component to stay roughly in the frequency band $|\xi| \sim N$ over time, although it may also create interactions at lower frequencies $|\xi| \ll N$ or higher frequencies $|\xi| \gg N$.

- All the above equations have a certain scale invariance of the form

$$u(x,t) \mapsto \lambda^{-a} u(\frac{x}{\lambda}, \frac{t}{\lambda^d})$$

for some constants $a, d$ depending on the equation: $(a,d)$ equals $(1,2)$ for NLS, $(2,3)$ for KdV, and $(0,1)$ for wave maps. This is in contrast to the linear counterparts $Lu = 0$, in which the scale invariance and homogeneity are decoupled:

$$u(x,t) \mapsto u(\frac{x}{\lambda}, \frac{t}{\lambda^d})$$
$$u(x,t) \mapsto \mu u(x,t)$$

Alternatively, one can interpret time has having units of $length^d$ and $u$ has having units of $length^{-a}$. Thus a time derivative has the same scaling as $d$ spatial derivatives, while $u$ itself has the same scaling as $a$ spatial derivatives. (This provides a very convenient checksum in computations, as all identities and estimates should be dimensionally consistent).

In particular, we see that Sobolev norms $\|u\|_{\dot{H}^s}$ themselves have a dimension, namely $length^{-a+n/2-s}$. The *critical regularity* $s_c$ is the value of $s$ for which $\|u\|_{\dot{H}^{s_c}}$ is dimensionless, i.e. $-a + n/2 - s_c = 0$. This quantity is then invariant under scaling. Regularities higher than $s_c$, $s > s_c$ are known (somewhat confusingly - it's because derivatives have negative dimension) as *sub-critical*; the $\dot{H}^s$ norms can then be made small by blowing up the solution in time and space. Regularities lower than $s_c$ are *super-critical*: one cannot shrink these norms by "zooming in" in time and space.

The scaling regularity $s_c$ is $n/2-1$ for NLS, $-3/2$ for KdV, and $n/2$ for wave maps.

Notice that the critical regularity $s_c$ has no obvious relation with the *phase space regularity* $s_p$ (0 for NLS, $-1/2$ for KdV, and $1/2$ for wave maps), or the *Hamiltonian regularity* $s_e$ (1 for NLS, KdV, and wave maps)[2]. A particularly important distinction is whether the Hamiltonian regularity $s_e$ (which controls the conserved quantity) is sub-critical, critical, or super-critical; for instance in wave maps, this occurs when $n < 2$, $n = 2$, and $n > 2$ respectively. It is fair to say that for large data global existence problems, we know a lot when the Hamiltonian is sub-critical, a little when the Hamiltonian is critical, and virtually nothing (in the positive direction) when the Hamiltonian is supercritical. (The three-dimensional Navier-Stokes and Einstein equations unfortunately fall into this last category; for Navier-Stokes the only useful quantity we can bound is $L^2$ of the velocity, but scaling is $\dot{H}^{1/2}$. For Einstein, we can bound the energy ($\dot{H}^1$ norm) of the metric, but scaling is $\dot{H}^{3/2}$.).

In these notes we discuss the problem of *global well-posedness* of non-linear dispersive equations of the above type. The basic question, for a fixed regularity $H^s$, is this: if the initial data is in $H^s$ (or $H^s \times H^{s-1}$ in the case of wave equations), does there exist a unique global solution to the Cauchy problem which remains in $H^s$ for all time? (One may have to specify "unique", "exist", and even "solution" more carefully).

[2]There is also $s_d$, the least regularity required to make sense of the equation as a distribution (often $s_d = 0$), but this regularity is usually unimportant because it is either much lower than what one can hope for, or there are various "local smoothing" effects that allow one to interpret solutions below $s_d$.

Additional control - such as continuous dependence of the solution on the data, or long-time asymptotics or growth bounds on the solution - is also desirable.

The classical theory of PDE focuses on smooth solutions, when $s$ is extremely large. In this case it is fairly easy to obtain *local well-posedness* for the above equations, but to obtain global well-posedness it is sometimes necessary to go down to much lower regularities, such as $s_c$, $s_p$, or $s_e$. The low regularity theory is also useful in obtaining more precise control on solutions (including smooth solutions), and more generally in deepening the intuition and set of tools for these equations.

## 2. Local well-posedness

For most of this section we shall discuss NLS for sake of argument, but the points made here are quite general.

Before we understand the global problem we must first understand the local problem, of how to obtain existence and uniqueness even for a short time interval $[0, T]$. Generally we expect that the larger the data, the shorter the time $T$ for which we can obtain existence. A simple model to keep in mind is the scalar ODE

$$\dot{u} = u^2.$$

This has an explicit solution

$$u(t) = \frac{u(0)}{1 - tu(0)}$$

and so blows up at time $T = u(0)^{-1}$.

In general it is not possible to obtain explicit solutions (except in special cases such as completely integrable equations, although even there the formulae are not so tractable). Current (deterministic) PDE existence theory then offers two basic methods to construct solutions: viscosity or compactness-type methods, and iteration-type methods. There are some variant methods to these two basic ones (e.g. difference element schemes, Glimm's method for conservation laws, etc.); it is certainly worth exploring other methods and getting a better understanding as to their comparative strengths and weaknesses.

*Viscosity methods* are based on introducing a damping term to the equation, for instance replacing KdV by the viscous KdV equation[3]

$$u_t^{(\varepsilon)} + u_{xxx}^{(\varepsilon)} = \varepsilon u_{xx}^{(\varepsilon)} + u^{(\varepsilon)} u_x^{(\varepsilon)};$$

[3]This is not the only way, or even the best way, to add viscosity terms to the KdV equation. See for instance [1].

solutions to these equations can usually be constructed using parabolic existence theory. One then tries to show that $u^{(\varepsilon)}$ converges to a solution $u$ of the original equation when $\varepsilon \to 0$ (taking subsequences if necessary).

In Fourier space, what the viscosity method is essentially doing is truncating the frequency variable $\xi$ to a bounded region ($\xi = O(\sqrt{T/\varepsilon})$ in the above example), obtaining a frequency-truncated evolution. One then relaxes the truncation radius to infinity. This works well as long as there is little interaction between frequencies. It does not work well when high frequency components often interact to create low frequency contributions to the wave (since such interactions are not picked up by the truncated equation) - but then again, no method seems to work when one has a strong "high-to-low cascade". The other problem is the "low-to-high cascade": it is possible for a component of the data to escape off to high frequencies and eventually cause a singularity in the solution. This will not cause the truncated solutions to diverge when $\varepsilon \to 0$ (at least if one is willing to take subsequences), but the limiting solution will not capture this singularity properly, in that the energy moving off to infinitely high frequencies may just disappear in the limit. This is the problem of "ghost solutions" - solutions which do not obey conservation of energy properly[4] (although energy monotonicity is usually guaranteed just from Fatou's lemma); related problems are that one is often unable to prove uniqueness or regularity for viscosity solutions. On the other hand, the technique is extremely robust and does not require the equation to be close to a linear equation.

*Iteration methods* provide much more control on the solution, but seem to be limited to equations which are close to linear. For the types of equations discussed here, iteration methods appear to be superior to viscosity methods and we shall use them exclusively.

The idea (dating back to Picard) is to treat the non-linear term $F(u)$ of the Cauchy problem

$$Lu = F(u); \quad u(0) = u_0$$

[4]A typical example is the critical focussing NLS $iu_t + u_{xx} = -|u|^4 u$ in one dimension. There exists solutions which are self-similar and oscillatory and collapse to a point singularity in finite time (after infinitely many rotations of phase). However if one adds a slight damping term then the solution behaves like the singular solution until very close to the singularity, at which point the solution disperses again (but with a phase which is highly dependent on the damping parameter $\varepsilon$). These solutions only converge in the weak sense as $\varepsilon \to 0$ because of the oscillation in phase, and the weak limit is in fact zero beyond the singularity, so conservation of energy or $L^2$ mass fails beyond the singularity. See [6].

as a perturbation. One starts with the linear solution

$$Lu^{(0)} = 0; \quad u^{(0)}(0) = u_0$$

and then constructs successive approximations $u^{(1)}, u^{(2)}, \ldots$ by solving the inhomogeneous linear problems

$$Lu^{(n+1)} = F(u^{(n)}); \quad u^{(n+1)}(0) = u_0.$$

Hopefully the iterates $u^{(n)}$ then converge (in a suitable normed vector space) to a solution $u$ of the original Cauchy problem. (This approach is closely related to the method of power series - trying to expand $u$ as a power series in the initial data $u_0$. The $n^{th}$ iterate $u^{(n)}$ is something like the $n^{th}$ partial sum of that power series).

Let's see how this works with a simple ODE, such as

$$\dot{u} = u^2; u(0) = 5.$$

We use the Fundamental Theorem of Calculus to rewrite this differential equation as an integral equation:

$$u(t) = N(u(t))$$

where $N$ is the non-linear operator

$$N(u)(t) := 5 + \int_0^t u^2(t')\ dt'.$$

We want a fixed point of $N$. We can try iteration, starting with $u \equiv 0$ (or $u \equiv 5$) and applying $N$ repeatedly. To make the iterates converge we need a contraction, say on the ball of radius 10 in $L^\infty([0,T])$. Certainly $N$ maps this ball to itself if $T$ is sufficiently small. Now we need a statement like

$$\|N(u) - N(v)\|_{L^\infty([0,T])} \leq \frac{1}{2}\|u - v\|_{L^\infty([0,T])}$$

for all $u, v$ in this ball. But this is easily accomplished if $T$ is small enough:

$$\|N(u)-N(v)\|_{L^\infty([0,T])} \leq \int_0^T |u^2(t')-v^2(t')|\ dt' \leq T\|u+v\|_\infty\|u-v\|_\infty \leq 20T\|u-v\|_\infty.$$

Thus we can obtain a unique fixed point of $N$ in this ball for sufficiently small times. Note that the time obtained $T$ is inversely proportional to the initial data 5, which agrees with what the explicit solution gives. Also the same argument shows that solution depends continuously (and even analytically) on the initial data 5.

The same principle works for, say, the non-linear Schrödinger equation

$$iu_t + \Delta u = |u|^2 u; u(0) = u_0 \in H^s$$

if $s$ is large enough. We can write this equation as an integral equation

$$u(t) = N(u(t))$$

where

$$N(u(t)) := e^{it\Delta} u_0 + \int_0^t e^{i(t-t')\Delta}(|u(t')|^2 u(t'))\ dt'.$$

We can again try iteration, now in a large ball in $L^\infty([0,T]; H^s(\mathbf{R}^n))$ (the trick of using this particular space is generally referred to as the *classical energy method*; as we shall see, it works providing that $s$ is larger than $n/2$). Note that $e^{it\Delta}$ is bounded on $H^s$ (which is why we use $L^2$ based Sobolev spaces to measure regularity for dispersive equations!), so the linear part $e^{it\Delta}u_0$ of $N(u(t))$ is already in this ball. To obtain the contraction we have to make an estimate such as

$$\| \int_0^t e^{i(t-t')\Delta}(|u(t')|^2 u(t') - |v(t')|^2 v(t'))\ dt' \|_{L^\infty([0,T];H^s)} \leq \frac{1}{2} \|u-v\|_{L^\infty([0,T];H^s)}$$

for all $u, v$ in the ball. By Minkowski's inequality and the boundedness of $e^{it\Delta}$ we can bound the left-hand side by

$$\int_0^T \| |u(t')|^2 u(t') - |v(t')|^2 v(t') \|_{H^s}\ dt'.$$

If we had an estimate such as

$$\|fgh\|_{H^s} \leq C \|f\|_{H^s} \|g\|_{H^s} \|h\|_{H^s}$$

then we would have

$$\| |u(t')|^2 u(t') - |v(t')|^2 v(t') \|_{H^s} \leq C \|u(t') - v(t')\|_{H^s}$$

for $u$, $v$ in the ball, so we have

$$\int_0^T \| |u(t')|^2 u(t') - |v(t')|^2 v(t') \|_{H^s}\ dt' \leq CT \|u - v\|_{L_t^\infty H^s}.$$

Thus we can again obtain a contraction if $T$ is sufficiently small. Again, the time of existence given by this argument is like $\sim \|u_0\|_{H^s}^{-1}$.

It remains to prove the product estimate. The left-hand side is basically

$$\|D^s(fgh)\|_2,$$

where $D^s$ is a differentiation operator of order $s$. We now use the *fractional Leibnitz rule*

$$D^s(fgh) \lessapprox (D^s f)gh + f(D^s g)h + fg(D^s h)$$

and Hölder to estimate

$$\|D^s(fgh)\|_2 \lesssim \|D^s f\|_2\|g\|_\infty\|h\|_\infty + \|f\|_\infty\|D^s g\|_2\|h\|_\infty + \|f\|_2\|g\|_\infty\|D^s h\|_2.$$

If we have $s > n/2$, then $H^s$ controls $L^\infty$, and we are done.

The fractional Leibnitz rule can be made more rigorous by using Plancherel's theorem and the Fourier transform. An informal explanation is as follows. Suppose that $f$ has frequency $\xi$, $g$ has frequency $\eta$, and $h$ has frequency $\zeta$. Then differentiating $f$ is like multiplying by $i\xi$, etc. Also $fgh$ has frequency $\xi + \eta + \zeta$. Putting all this together, the Leibnitz rule then (informally) says that

$$|\xi + \eta + \zeta|^s \lesssim |\xi|^s + |\eta|^s + |\zeta|^s$$

which is certainly true. Basically, this rule states that the frequency of $fgh$ is not much larger than the largest frequency of $f$, $g$, and $h$ individually.

The above argument gave local existence if the initial data was in $H^s$ for $s > n/2$, and even keeps the solution bounded in $H^s$. It is tempting to then try to iterate this argument to give global existence (using $u(T)$ as new initial data for another time step) but the trouble is that the $H^s$ norm at each step can be something like twice the $H^s$ norm of the previous step, so that the time of existence gained at each step is like half the previous one, and one can still obtain blowup in finite time (as in the ODE $\dot{u} = u^2$). However if we can keep the $H^s$ norm from blowing up then we should be able to obtain global existence.[5]

To convert local existence to global existence, we can then seek ways to control the $H^s$ norms for which we have a local result. In the case of NLS, the conservation of the Hamiltonian

$$H(u) := \int \frac{1}{2}|\nabla u|^2 + \frac{1}{4}|u|^4$$

(as well as the $L^2$ norm) allows us to say that if the solution is in $H^1 \cap L^4$ initially, then it stays bounded in $H^1$ for all time. In one dimension

[5]In fact, the above argument gives something more, namely *persistence of regularity*: if the $H^s$ norm is bounded for some $s > n/2$ and a higher norm $H^{s'}$ is finite, then the solution stays in the high regularity space $H^{s'}$ for as long as the low regularity norm $H^s$ is bounded. The idea is first to control the low regularity norm $H^s$, and then to begin iterating in the higher norm $H^{s'}$. Even if the $H^{s'}$ norm is large, the fact that the $H^s$ norm is bounded will still allow for a contraction for a time $T$ which depends only on the low-regularity norm $H^s$. Ultimately, this comes down to the fact that the Leibnitz rule only places the big derivative $D^{s'}$ on one of the three terms in the non-linearity. There is a general principle which roughly states that if one has good control on low regularity norms, one automatically gets good control on higher norms too.

this is already enough to give global existence for this equation since $1 > n/2$ in this case (note $H^1$ controls $L^4$ in dimensions $n \leq 3$).

This still leaves open the question of what happens in higher dimensions. It is clear that we might be able to answer this question better if we could push the local well-posedness theory down below $s > n/2$, and hopefully down to the energy regularity $s = 1$ or below. More generally we are interested in making the regularity needed for local existence as low as possible.

We can try an abstract iteration scheme; replacing the space $L_t^\infty H_x^s$ used before by some general Banach space $X$ of spacetime functions $u(x,t)$. To obtain a contraction we thus need estimates such as

$$\|e^{it\Delta} u_0\|_X \lesssim \|u_0\|_{H^s}$$

(i.e. free $H^s$ solutions must lie in $X$), and

$$\| \int_0^t e^{i(t-t')\Delta} |u(t')|^2 u(t')\, dt' \|_X \lesssim T^\alpha \|u\|_X^3 \tag{1}$$

for some $\alpha \geq 0$ (in order for $N$ to map $X$ to $X$; a similar estimate will also make $N$ a contraction on $X$ if $T$ is sufficiently small).

The first estimate forces $X$ to have the same scaling as $H^s$, which is $length^{-(s-s_c)}$. Now a dimensional analysis of (1)[6] gives

$$length^{-(s-s_c)} \leq length^{2\alpha} length^{-3(s-s_c)}$$

so the best value of $\alpha$ we can obtain is

$$\alpha \leq s - s_c.$$

(note: for the energy method $s > n/2$, $s_c = n/2 - 1$, and $\alpha = 1$).

In particular, we see that this type of iteration argument cannot possibly work for supercritical regularities $s < s_c$. (It is absurd for the non-linear estimate (1) to have a negative value of $\alpha$, since the left-hand side will increase in $T$). So iteration methods have a natural barrier at the critical regularity $s_c$.

A little work shows that the time of existence given by the above arguments are something like $T \sim \|u\|_{H^s}^{-1/\alpha}$. So the lower we can push $s$ toward $s_c$, the closer $\alpha$ gets toward zero, and the longer we can push the existence of small data. In particular if we can obtain a space $X$ to iterate in which has the critical regularity, then we have $\alpha = 0$ and we can get $T = +\infty$ if $\|u\|_{H^{s_c}}$ is sufficiently small. Thus it is possible to obtain global well-posedness for small data without requiring a conservation law if we can iterate at the critical regularity.

[6]This is a little inaccurate because the $H^s$ norm is inhomogeneous, but the conclusion is still essentially correct.

It turns out that for the NLS one can in fact push the local theory all the way down to the critical regularity $s_c = n/2 - 1$, except in one dimension when we hit another barrier at $s = 0$ (below $s = 0$ the non-linearity doesn't make much sense, and more pertinently there is an obstruction coming from Gallilean invariance rather than scale invariance (note that the space $H^s$ is only Gallilean invariant when $s = 0$)). Roughly the idea is as follows. An inspection of the energy method proof shows that we didn't really need to control $L_t^\infty H^s$, we wanted $L_t^\infty L_x^\infty$. And in fact we didn't even need that, we actually would be happy with $L_t^1 L_x^\infty$. Now normally to obtain $L_x^\infty$ control on a function, Sobolev embedding requires more than $n/2$ derivatives in $L^2$. But the averaging in time given by the $L_t^1$ term allows one to do better, because of the *dispersive* nature of the equation. The point is that Sobolev embedding is only efficient when the solution is concentrated on a small ball, but the dispersion prevents a solution from staying like a ball for extended periods of time.

This leads to the theory of *Strichartz estimates*, which seek to find good $L_t^1 L_x^\infty$ (or more generally $L_t^q L_x^r$) estimates on a solution (to either a free equation or an inhomogeneous equation) given that the initial data is, say, in $H^s$, and there is some reasonable control on the forcing term. Further details would take us too far afield, but let me just say that one of the main tools is the *dispersive inequality*, which for Schrödinger equations is

$$\|e^{it\Delta} f\|_\infty \lesssim t^{-n/2} \|f\|_1.$$

This is an immediate consequence of the kernel bounds for $e^{it\Delta}$ and is crucial to getting $L_x^\infty$ (or $L_x^r$, if one interpolates with $L^2$ estimates) bounds on solutions.

Things get more complicated when there are derivatives in the non-linearity (as with KdV and wave maps), and one has to make heavy reliance on the space-time Fourier transform and the characteristic surface of the free equation. Again, this would take us too far afield. Suffice to say that by means of multilinear estimates using the space-time Fourier transform, one can get local well-posedness for KdV down to $s \geq -3/4$ on the real line [17], [8], or down to $s \geq -1/2$ for periodic data [18]. For wave maps, one can obtain local well-posedness for $s > s_c = n/2$, with some recent progress at the critical regularity $s = s_c = n/2$ (see below).

## 3. Global existence

We now try to extend the above local existence methods to give global existence. Viscosity methods can give global solutions without

much difficulty (in fact it is no harder than to generate local solutions, usually) but it is possible that these are "ghost" solutions, continuing long after the solution makes any physical sense (because of energy moving off to infinitely high frequencies, and thus leaving the viscosity limit). One symptom of this is that these solutions usually only have energy monotonicity rather than energy conservation, and are not time reversible.

Iteration methods can give global solutions if the critical norm $\dot{H}^{s_c}$ of the initial data is small (by direct iteration globally in time), or if a suitable $H^s$ norm is controlled for all time (by iterating locally in time and then advancing one timestep at a time). Another possibility if there is some decay in the solution, so that some norm (e.g. the $L^\infty$ norm) actually gets smaller with time, possibly counteracting growth in other norms such as $H^s$ norms. This however only seems to be possible when one assumes the initial data is small and rapidly decreasing. There is a large body of work pursuing this type of result (generally known as global existence for small data) but we will not pursue it here.

For now, we focus on the problem of controlling $H^s$ norms globally in time. Some $H^s$ norms are bounded directly from conservation laws. For instance, for NLS the $L^2$ norm is exactly conserved, while the $H^1$ norm is closely related to the Hamiltonian and stays bounded if it is bounded at time zero. For the KdV equation one can in fact show that for any integer $k$, the $H^k$ norm stays bounded if it is bounded at time zero.

The question then arises as to what happens to other norms, such as the $H^{1/2}$ norm. If the initial data is in $H^1$ then it is bounded in $H^1$ for all time and hence the $H^{1/2}$ norm is similarly bounded, but if the initial data is only in $H^{1/2}$ then the $H^1$ norm may be infinite, and so the conservation of the Hamiltonian appears to be useless. Conservation of the $L^2$ norm also appears to be unhelpful for this particular problem (but for the opposite reason).

To understand the relationship between the $H^s$ norms further it is convenient to introduce the *Littlewood-Paley decomposition*

$$u = \sum_{k \in \mathbf{Z}} P_k u$$

where

$$\widehat{P_k u}(\xi) = \eta(\xi/2^k)\hat{u}(\xi)$$

and $\eta(\xi)$ is a suitable bump function localized to the region $|\xi| \sim 1$. Thus $P_k u$ is the portion of $u$ having frequencies $\sim 2^k$.

A little application of Plancherel's theorem shows that

$$\|\nabla P_k u\|_2 \sim 2^k \|P_k u\|_2$$

and more generally that

$$\|P_k u\|_{H^s} = \|D^s P_k u\|_2 \sim (1+2^k)^s \|P_k u\|_2.$$

Thus for each frequency piece $P_k u$, the $H^s$ norms are all related by a suitable power of $2^k$ (or $(1+2^k)$). By orthogonality we have

$$\|u\|_{H^s} \sim (\sum_k ((1+2^k)^s \|P_k u\|_2)^2)^{1/2}.$$

It is now convenient to plot the distribution of energy as a function of $k$. There are several ways to do this; one nice way is to plot $\log \|P_k u\|_{\dot{H}^1}$ as a function of $k$.

If $u$ is bounded in $\dot{H}^1$, so $\|u\|_{\dot{H}^1} \lesssim 1$, then $\log \|P_k u\|_{\dot{H}^1} \leq O(1)$, so the energy distribution is essentially bounded above by the $x$-axis. The converse is true up to logarithms: if $u$ is truncated to frequencies $N^{-1} \leq |\xi| \leq N$ and one has $\log \|P_k u\|_{\dot{H}^1} \leq O(1)$ for all $k$, then $\|u\|_{\dot{H}^1} \lesssim \log N$.

Conservation of energy implies that the $\dot{H}^1$ norm stays bounded, which means that the energy distribution can move from one frequency mode $k$ to another as time evolves but the "sup norm" must stay bounded.

Now what does it mean for the initial data to lie in, say, $H^{1/2}$? This implies that $(1+2^k)^{1/2} \|P_k u(0)\|_2$ is bounded, so that

$$\log \|P_k u(0)\|_{\dot{H}^1} \leq \min(k, k/2) + O(1).$$

Thus we have very small energy at low frequencies, but potentially infinite energy at high frequencies. If at later times all this energy comes down to low frequencies we can have blow up (the "high to low" cascade).

Now if we have $L^2$ norm conservation, then we know that for all time $\|P_k u\|_2$ is bounded, so that

$$\log \|P_k u(t)\|_{\dot{H}^1} \leq k + O(1).$$

This goes some way towards preventing a high to low cascade, but still does not prevent the $H^{1/2}$ norm from becoming infinite in finite time.

So it seems that to stop the $H^{1/2}$ norm from blowing up we cannot just use $L^2$ and $H^1$ conservation, but must somehow also prevent energy from moving from high to low frequencies. Similarly for any $H^s$ norm with $0 < s < 1$.

The major breakthrough came from Bourgain [3] in 1997, who came up with the idea of introducing a large frequency parameter $N$ and dividing the solution $u$ into the low frequency portion $u_{low}$ (where

$|\xi| \lesssim N$) and the high frequency portion $u_{high}$ (where $|\xi| \lesssim N$). The low frequency term had finite (but large) energy (for instance, if $u$ was bounded in $H^s$ for some $s < 1$, then the $H^1$ norm of $u_{low}$ was $O(N^{1-s})$), while the high frequency term was still quite rough (and had infinite energy), but was small with respect to lower regularity norms (for instance, the $L^2$ norm was $N^{-s}$). The idea was then to treat the high frequency term as a perturbation of the low frequency term (since the low frequency term, if left by itself, would evolve globally for all time thanks to finite energy). The strategy was to show that the non-linear interaction terms arising from the high frequencies could be kept under control for long times (of the order of $T \sim N^\alpha$ for some $\alpha > 0$). The main tool in this was an *extra smoothing estimate*, which showed that if the high frequencies were merely in $H^s$ for some $s < 1$, then the non-linear interactions arising from the high frequencies were significantly smoother, in fact they were in the energy class $H^1$. Thus these terms could be absorbed harmlessly into the low frequency term without destroying the finite energy property. (One can think of this as kind of a compactness property of the non-linearity). Thus Bourgain's method effectively decouples the equation into a high-frequency component evolving by the free evolution, and a low-frequency component which evolves in a forced Hamiltonian system, but with the forcing term controllable in energy norm for a reasonably long period of time.

Applying this strategy to the 2D NLS, Bourgain [3] was able to obtain global well-posedness below $H^1$, and in fact down to $s > 3/5$. The reason for the limitation is that the lower $s$ gets, the larger the energy of $u_{low}$ gets, and this eventually overwhelms the gain in the local smoothing estimate (so the $\alpha$ eventually dips below zero).

The argument also shows that the $H^s$ norm grows at most polynomially in time, so is $O(T^{\beta(s)})$ for some $\beta(s) > 0$. As one might expect, this $\beta$ becomes infinite in Bourgain's argument as $s \to 3/5$.

A similar argument works for KdV, but is substantially more difficult because of the derivative in the non-linearity; one has to work much harder to obtain extra smoothing. Ignoring some technicalities, Colliander, Staffilani, and Takaoka [9] were able to obtain global existence down from $s = 0$ (which follows from $L^2$ conservation and local existence theory) to $s > -1/24$, which is some way short of the local theory, which goes down to $s > -3/4$. However for wave maps there is no extra smoothing (too many derivatives in the non-linearity), and Bourgain's method does not seem to work directly. (Similar results have been achieved for a wide array of model non-linear dispersive equations).

However, a new approach was developed later by Colliander, Keel, Staffilani, Takaoka, and myself - which we call the "$I$-method" - which achieves similar or better results without needing an extra smoothing estimate. To illustrate the idea, let us return to NLS and consider the Hamiltonian

$$H(u) = \int \frac{1}{2}|\nabla u|^2 + \frac{1}{4}|u|^4.$$

We know that the Hamiltonian is an invariant of the flow:

$$\frac{d}{dt}H(u(t)) = 0.$$

Since the Hamiltonian is roughly comparable to the $H^1$ norm, this allows us to control the $H^1$ norm globally in time. Let us quickly prove this conservation law:

$$\begin{aligned}\frac{d}{dt}H(u(t)) &= \mathrm{Re}\int \nabla u\overline{\nabla} u_t + |u|^2 u\overline{u}_t \\ &= \mathrm{Re}\langle -\Delta u + |u|^2 u, u_t\rangle \\ &= \mathrm{Re}\langle -iu_t - \Delta u + |u|^2 u, u_t\rangle.\end{aligned}$$

If $u$ solves NLS then the last line is clearly zero.

Could we use the Hamiltonian to control, say the $H^2$ norm? The obvious thing to do is to look at the expression $H(Du)$, where $D$ is a differentiation operator of order 1. By the above computation we have

$$\frac{d}{dt}H(Du(t)) = \mathrm{Re}\langle -iDu_t - D\Delta u + |Du|^2 Du, Du_t\rangle.$$

On the other hand, if we apply $D$ to the NLS equation we have

$$-iDu_t - D\Delta u + D(|u|^2 u) = 0.$$

Thus the only thing left is a "commutator" term:

$$\frac{d}{dt}H(Du(t)) = \mathrm{Re}\langle |Du|^2 Du - D(|u|^2 u), Du_t\rangle.$$

Using the NLS equation to expand out $Du_t$, then some integration by parts followed by Hölder and Sobolev, one can eventually get the estimate

$$|\frac{d}{dt}H(Du(t))| \lesssim H(Du(t))^2 + l.o.t.$$

which allows one to control $H(Du(t))$ for short times only.

A similar thing happens if we try to control the evolution of the $H^s$ norm for any $s$ (either $s > 1$ or $s < 1$; the difference being that $D$ is replaced by an integration operator when $s < 1$). In all these

cases it appears that one only obtains control for short times. However, inspired by Bourgain's method, we again introduce a frequency parameter $N$.

Suppose we wish to control the $H^s$ norm for some $s < 1$. We introduce the operator $I$ defined by

$$\widehat{Iu}(\xi) = m(\xi)\hat{u}(\xi)$$

where $m(\xi)$ equals 1 for $|\xi| \leq N$, equals $(|\xi|/N)^{s-1}$ for $|\xi| \geq 2N$, and is a smooth interpolant in between. Thus $I$ is the *I*dentity for frequencies less than $N$, and an *I*ntegration operator for frequencies greater than $N$. Using the logarithmic energy distribution picture, $I$ does nothing when $k < \log(N)$ but brings down the logarithmic energy by $(1-s)(k - \log N)$ for $k > \log N$. From this we see that $\|Iu\|_{H^1}$ is somewhat comparable with $\|u\|_{H^s}$:

$$\|Iu\|_{H^1} \lesssim \|u\|_{H^s} \lesssim N^{1-s}\|Iu\|_{H^1}$$

Now look at how the Hamiltonian $H(Iu)$ evolves. By the same calculation as before we have

$$\frac{d}{dt}H(Iu(t)) = \mathrm{Re}\langle |Iu|^2 Iu - I(|u|^2 u), Iu_t\rangle.$$

Now we begin to see $I$ work its magic. If $I$ were the identity, then the above expression vanishes. In particular, if $u$ consisted only of low frequencies ($< N/3$) then there is no variation in the above expression. Thus if we split $u = u_{low} + u_{high}$ as before then the only terms which are non-zero are the terms with at least one high frequency, e.g.

$$\mathrm{Re}\langle |Iu_{low}|^2 Iu_{high} - I(|u_{low}|^2 u_{high}), Iu_t\rangle.$$

The point is now to exploit the fact that the high frequency is small, at least when measured in rough norms.

The terms which are quadratic or better in the high frequencies (the "high-high" interactions) tend to be quite small, so let us concentrate on the terms which are linear in the high frequencies (the "low-high" interactions). These are the interactions which cause the most trouble in Bourgain's method, as a low-high interaction is again a high frequency and so they prevent the high frequency term from being approximated by a the linear evolution. Here however the I-method has an advantage, in that it can exploit some cancellation in the commutator expression

$$|Iu_{low}|^2 Iu_{high} - I(|u_{low}|^2 u_{high}).$$

Indeed, since $I$ is the identity in $u_{low}$, this is

$$[|u_{low}|^2, I]u_{high}.$$

If $u_{low}$ was 0 frequency (i.e. constant) then the commutator vanishes. Thus we expect the commutator to be quite small for low frequencies. To make this more precise, if the low frequency object $|u_{low}|^2$ had frequency $\eta$ and $u_{high}$ had frequency $\xi$ (so $|\xi| \gtrsim N \gtrsim |\eta|$), then the commutator behaves like $|u_{low}|^2(m(\xi) - m(\xi + \eta))$. We thus expect cancellation between the two $m$'s thanks to the smoothness of $m$ (in practice we obtain this using the mean value theorem).

Using this method we were able to push GWP for 2D NLS down to $s > 4/7$, a slight improvement over Bourgain's results, and similarly for KdV down to $s > -3/10$ and for 1D wave maps down to $s > 3/4$. There is still a barrier to going all the way down to match the local theory because the smaller $s$ gets, the larger $H(Iu)$ gets and the more difficult it is to control $\frac{d}{dt}H(Iu(t))$.

In terms of the frequency distribution, the $I$-method asserts that for times up to some time $T = N^\alpha$, the high frequencies do not move much in energy (so that one can damp them with the $I$ without affecting conservation of energy much), while the low frequencies stay low but can slosh around between low modes. As time increases, more and more of the frequencies are free to move around, causing a possible polynomial growth in the energy.

The above strategy can be viewed as an attempt to construct an *almost conserved quantity*, namely $H(Iu)$. This quantity is not exactly conserved, but varies slowly in time. To improve the above results we had to refine the almost conserved quantity somewhat.

To illustrate the procedure let us consider the KdV equation again. Instead of the Hamiltonian, let us consider the simpler conserved quantity

$$L(u) = \int u^2.$$

Let us give a Fourier proof why this quantity is conserved. We can write this quantity in Fourier space as

$$L(u) = \int_{\xi_1+\xi_2=0} \hat{u}(\xi_1)\hat{u}(\xi_2).$$

But $\hat{u}$ obeys the equation

$$\hat{u}_t(\xi) = i\xi^3\hat{u}(\xi) + \int_{\xi_1+\xi_2=\xi} i\xi_2\hat{u}(\xi_1)\hat{u}(\xi_2).$$

If we differentiate $L(u)$ using this equation, we obtain

$$\frac{d}{dt}L(u) = \int_{\xi_1+\xi_2=0} (i\xi_1^3+i\xi_2^3)\hat{u}(\xi_1)\hat{u}(\xi_2) + \int_{\xi_1+\xi_2+\xi_3=0} (i\xi_2+i\xi_3)\hat{u}(\xi_1)\hat{u}(\xi_2)\hat{u}(\xi_3).$$

The first integral is zero because $\xi_1^3 + \xi_2^3$ vanishes on $\xi_1 + xi_2$. The second integral can be symmetrized, replacing $\xi_2 + \xi_3$ by $\frac{2}{3}(\xi_1 + \xi_2 + \xi_3)$ which also vanishes.

If this was the only conserved quantity we knew, and we wanted to construct an almost conserved quantity with the scaling of $H^1$, we would try

$$L(u_x) = \int u_x^2.$$

In Fourier space this is

$$L(u_x) = \int_{\xi_1+\xi_2=0} (i\xi_1)(i\xi_2)\hat{u}(\xi_1)\hat{u}(\xi_2).$$

Differentiating this again

$$\frac{d}{dt}L(u_x) = \int_{\xi_1+\xi_2=0} (i\xi_1^3 + i\xi_2^3)(i\xi_1)(i\xi_2)\hat{u}(\xi_1)\hat{u}(\xi_2)$$
$$+ \int_{\xi_1+\xi_2+\xi_3=0} (i\xi_2 i(\xi_1+\xi_2)i\xi_3 + i\xi_3 i\xi_1 i(\xi_2+\xi_3))\hat{u}(\xi_1)\hat{u}(\xi_2)\hat{u}(\xi_3).$$

The first integral vanishes as before, but the second integral does not, even after symmetrization. Instead, it eventually simplifies to $\frac{2i}{3}(\xi_1^3 + \xi_2^3 + \xi_3^3)$.

This looks bad, but we can fix the problem by adding a correction term to $L(u_x)$, namely

$$L(u_x) - \frac{2}{3}\int u^3.$$

In Fourier space the correction term is

$$-\frac{2}{3}\int_{\xi_1+\xi_2+\xi_3=0} \hat{u}(\xi_1)\hat{u}(\xi_2)\hat{u}(\xi_3).$$

If one differentiates this, one obtains a trilinear term

$$-\frac{2}{3}\int_{\xi_1+\xi_2+\xi_3=0} (i\xi_1^3 + ixi_2^3 + i\xi_3^3)\hat{u}(\xi_1)\hat{u}(\xi_2)\hat{u}(\xi_3)$$

which happily cancels the term coming from $L(u_x)$, together with a quadrilinear term

$$-\frac{2}{3}\int_{\xi_1+\xi_2+\xi_3+\xi_4=0} (i\xi_2 + i\xi_3 + i\xi_4)\hat{u}(\xi_1)\hat{u}(\xi_2)\hat{u}(\xi_3)$$

which vanishes after symmetrization. Thus the above expression is perfectly conserved (though this is not so surprising since it is just twice the Hamiltonian).

Now the same strategy can be applied to the expression $\int u_{xx}^2$, and this will eventually generate the second KdV conserved quantity, and so forth. But we can also apply it to the expression

$$L(Iu) = \int (Iu)^2$$

which is the KdV analogue of $H(Iu)$. In Fourier this is

$$L(u_x) = \int_{\xi_1+\xi_2=0} m(\xi_1)m(\xi_2)\hat{u}(\xi_1)\hat{u}(\xi_2).$$

If we differentiate this, we get a bilinear term (which again vanishes) and a trilinear term

$$\int_{\xi_1+\xi_2+\xi_3=0} (i\xi_2 m(\xi_1+\xi_2)m(\xi_3) + i\xi_3 m(\xi_1)m(\xi_2+\xi_3))\hat{u}(\xi_1)\hat{u}(\xi_2)\hat{u}(\xi_3),$$

which symmetrizes to

$$\frac{-2i}{3}\int_{\xi_1+\xi_2+\xi_3=0} (\xi_1 m^2(\xi_1) + \xi_2 m^2(\xi_2) + \xi_3 m^2(\xi_3))\hat{u}(\xi_1)\hat{u}(\xi_2)\hat{u}(\xi_3).$$

Thus if we add the correction term

$$\frac{2i}{3}\int_{\xi_1+\xi_2+\xi_3=0} \frac{\xi_1 m^2(\xi_1) + \xi_2 m^2(\xi_2) + \xi_3 m^2(\xi_3)}{i\xi_1^3 + i\xi_2^3 + i\xi_3^3}\hat{u}(\xi_1)\hat{u}(\xi_2)\hat{u}(\xi_3)$$

to $L(Iu)$, we now get something which when differentiated, has no more trilinear term (thanks to cancellation), but instead picks up a quadrilinear term. At first glance things look bad because we have something in the denominator, but the magic of KdV intervenes at this point: if the denominator vanishes, then one of $\xi_1$, $\xi_2$, or $\xi_3$ must be zero, which then forces the numerator to also vanish! In fact the multiplier is smooth.

The quadrilinear term turns out to be better than the trilinear term to estimate (it is "lower order" in that there are more powers of $\xi$ in the denominator). This is ultimately due to the subcritical regularities we are dealing with: each $u$ has more regularity than scaling would suggest and so every $u$ we add to the multilinear expression should make things better.

By adding this term we can already improve KdV from $s > -3/10$ to $s > -1/2$. We can go further and add a second correction term to go all the way down to $s > -3/4$, which is essentially the best possible from the local theory. (Remember, it is not just enough to have $H^s$ control: we also need a local theory which can use this $H^s$ control). For periodic data, things are a little more difficult (we do not get as much dispersion), but we can get global existence down to $s \geq -1/2$,

which is best possible for the local theory. In particular we have a global continuous flow in the symplectic phase space $\dot{H}^{-1/2}(\mathbf{T})$, which allows one to use infinite dimensional symplectic geometry machinery (e.g. symplectic capacity [25], [5]). A similar scheme for 2D NLS has managed to push regularity down to $s > 1/2$ and perhaps a little below, but there is much difficulty in going down to the optimal $s \geq 0$ because this equation is not completely integrable and we do not enjoy as much cancellation as before. This is still a work in progress.

## 4. Critical regularity global existence for wave maps

Let $\mathbf{R}^{1+n}$ be $n+1$ dimensional Minkowski space with flat metric $\eta := \mathrm{diag}(-1, 1, \ldots, 1)$, and let $S^{m-1} \subset \mathbf{R}^m$ denote the unit sphere in the Euclidean space $\mathbf{R}^m$. Elements $\phi$ of $\mathbf{R}^m$ will be viewed as column vectors, while their adjoints $\phi^\dagger$ are row vectors. We let $\partial_\alpha$ and $\partial^\alpha$ for $\alpha = 0, \ldots, n$ be the usual derivatives with respect to the Minkowski metric $\eta$. We let $\Box := \partial_\alpha \partial^\alpha = \Delta - \partial_t^2$ denote the D'Lambertian. We shall also use $\dot{\phi}$ for $\partial_t \phi$.

Define a *wave map* to be any function $\phi$ defined on an open set in $\mathbf{R}^{1+n}$ taking values on the sphere $S^{m-1}$ which obeys the equation

$$\partial_\alpha \partial^\alpha \phi = -\phi \partial_\alpha \phi^\dagger \partial^\alpha \phi. \tag{2}$$

For any time $t$, we use $\phi[t] := (\phi(t), \dot{\phi}(t))$ to denote the position and velocity of $\phi$ at time $t$. We refer to $\phi[0]$ as the *initial data* of $\phi$. We shall always assume that the initial data $\phi[0] = (\phi(0), \dot{\phi}(0))$ satisfies the consistency conditions

$$\phi^\dagger(0)\phi(0) = 1; \quad \phi^\dagger(0)\dot{\phi}(0) = 0 \tag{3}$$

(i.e. $\phi[0]$ lies on the sphere). It is easy to show (e.g. by Gronwall's inequality) that this consistency condition is maintained in time, for smooth solutions at least.

Let $H^s := (1+\sqrt{-\Delta})^{-s} L^2(\mathbf{R}^n)$ denote the usual[7] $L^2$ Sobolev spaces. Since the equation (2) is invariant under the scaling $\phi \mapsto \phi_\lambda$ defined by

$$\phi_\lambda(t, x) := \phi(t/\lambda, x/\lambda)$$

we see that the critical regularity is $s = n/2$.

The Cauchy problem for wave maps has been extensively studied; we refer the interested reader to the surveys in [19], [24], [29], [31].

[7]Strictly speaking, one cannot use $H^s$ spaces for functions on the sphere, since they cannot globally be in $L^2$. To get around this we shall abuse notation and allow constant functions to lie in $H^s$ with zero norm whenever the context is for functions on the sphere. Thus when we say that $\phi(t)$ is in $H^s$, we really mean that $\phi(t) - c$ is in $H^s$ for some constant $c$.

For sub-critical regularities $s > n/2$ it is known (see [21], [23], [16] for the $n \geq 4$, $n = 2, 3$, and $n = 1$ cases respectively) that the Cauchy problem for (2) is locally well-posed for initial data $\phi[0]$ in $H^s \times H^{s-1}$, and the solution can be continued (without losing regularity) as long as the $H^s$ norm remains bounded. The critical result however is more subtle. Well-posedness and regularity was demonstrated in the critical Besov space $\dot{B}_1^{n/2}$ in [35] in the high-dimensional case $n \geq 4$ and in [36] for $n = 2, 3$; in the one-dimensional case $n = 1$, a logarithmic cascade from high frequencies to low frequencies causes ill-posedness in the critical Besov and Sobolev spaces [32], although global smooth solutions can still be constructed thanks to the sub-critical nature of the energy norm (see e.g. [28]). As is to be expected at the critical regularity, these results give a global well-posedness (and regularity and scattering) when the norm of the initial data is small.

The question still remains as to whether the wave map equation (2) is well-posed in the critical Sobolev space $\dot{H}^{n/2} := \sqrt{-\Delta}^{-n/2} L^2$ in two and higher dimensions, with global well-posedness and regularity expected for small data. This question is especially interesting in the two dimensional case, since the critical Sobolev space is then the energy norm $H^1$, and one also expects to exploit conservation of energy (and some sort of energy non-concentration result) to obtain global well-posedness and regularity for data with large energy. (In dimensions three and higher one does not have large data global well-posedness for the sphere, even for smooth symmetric data; see [30]). However the Sobolev space $\dot{H}^{n/2}$ is significantly less tractable than its Besov counterpart $\dot{B}_1^{n/2}$; for instance, $\dot{H}^{n/2}$ norm fails to control the $L^\infty$ norm due to a logarithmic pile-up of frequencies. This logarithmic divergence is responsible for failure of any strengthened version of well-posedness (uniform, Lipschitz, or analytic) for wave maps at this regularity, as well as ill-posedness in very similar equations, and is a serious obstacle to any iteration-based argument. See [23], [16], [26] for further discussion.

In [33], [34] it was shown

**Theorem 4.1.** *Suppose that the initial data $\phi[0]$ is smooth and has sufficiently small $\dot{H}^{n/2} \times \dot{H}^{n/2-1}$ norm. Then the solution to the Cauchy problem for* (2) *with initial data $\phi[0]$ stays smooth globally in time.*

Our arguments are heavily based on the geometric structure of the equation (2); in particular, they do not directly apply to the associated difference equation. As a consequence we have not been able to obtain

a well-posedness[8] result at the critical regularity $\dot{H}^{n/2} \times \dot{H}^{n/2-1}$, even for small data.

The main novel ingredient is the use of adapted co-ordinate frames constructed by approximate parallel transport along (Littlewood-Paley regularizations of) $\phi$. The construction presented here is heavily dependent of the geometry of the sphere, although this has recently been extended to general manifolds in spatial dimensions $n \geq 4$ by the work of Klainerman-Rodnianski [22], Uhlenbeck-Nahmod-Stefanov, Krieger, and Shatah-Struwe.

Without the use of these frames, the usual iteration approach for (2) fails at the critical regularity because of a logarithmic pile-up of high-low frequency interactions. The effect of the adapted co-ordinate frame is to transform the high-low frequency interaction into other terms which are more tractable, such as high-high frequency interactions, or high-low interactions in which a derivative has been moved from a high-frequency term to a low-frequency one.

In the remainder of this section we shall informally motivate the key ideas in the argument. In doing so we shall make frequent use of the following heuristic: if $\phi$, $\psi$ are two functions, and $\psi$ is much rougher (i.e. higher frequency) than $\phi$, then $(\nabla\phi)\psi$ is very small compared to $\phi\nabla\psi$. In other words, we should be able to neglect terms in which derivatives fail to fall on rough functions, and land instead on smooth ones. (Indeed, these terms can usually be treated just by Strichartz estimates). In particular, we expect to have $\nabla(\phi\psi) \approx \phi\nabla\psi$ (which can be viewed as a statement that $\phi$ is approximately constant when compared against $\psi$).

Let us suppose that our wave map $\phi$ has the form $\phi = \tilde{\phi} + \varepsilon\psi$, where $\tilde{\phi}$ is a smooth wave map, $0 < \varepsilon \ll 1$ and $\psi$ is a $H^{n/2}$ function which is much rougher than $\tilde{\phi}$. (In other words, $\phi$ is a small rough perturbation of a smooth wave map). If we ignore terms which are quadratic or better in $\varepsilon$, or which fail to differentiate the rough function $\psi$, we obtain the linearized equation

$$\partial_\alpha \partial^\alpha \psi = -2\tilde{\phi}\partial_\alpha \tilde{\phi}^\dagger \partial^\alpha \psi \tag{4}$$

[8]We should remark at this point that strong versions of well-posedness, such as uniform, Lipschitz, or analytic well-posedness, are known to fail at the critical Sobolev regularity, mainly because the critical norm does not control $L^\infty$. It is similar to the fact that the non-linear map $u \mapsto e^{iu}$ is bounded and even continuous on the critical $\dot{H}^1(\mathbf{R}^2)$, but the power series $1 + iu + (iu)^2/2! + \ldots$ is divergent in $\dot{H}^1(\mathbf{R}^2)$ and hence the map is not analytic.

for $\psi$. Also, since $\tilde{\phi}$ and $\tilde{\phi} + \varepsilon\psi$ both take values on the sphere we see that

$$\tilde{\phi}^\dagger \psi = 0; \quad \tilde{\phi}^\dagger \partial_\alpha \psi = 0 \tag{5}$$

(again ignoring terms quadratic in $\varepsilon$, and terms where the derivative fails to land on $\psi$).

In order to keep the $H^s$ norm of $\tilde{\phi} + \varepsilon\psi$ from blowing up, we need to prevent the $\dot{H}^{n/2}$ norm from being transferred from $\tilde{\phi}$ to $\varepsilon\psi$. In particular, we need $L_t^\infty \dot{H}_x^{n/2}$ bounds on $\psi$ which are independent of $\varepsilon$. We would also like the corresponding Strichartz estimates for $\psi$, in order to control the error terms that we have been ignoring. (This scheme is not restricted to rough perturbations of smooth wave maps, and will be adapted to general wave maps by use of Littlewood-Paley projections).

Despite being linear, the equation (4) is not very well-behaved, having no obvious cancellation structure (beyond the null form, which is not particularly useful in the high-dimensional setting). In order to iterate away the first-order terms on the right-hand side of (4) we would like $\tilde{\phi}\partial_\alpha\tilde{\phi}^\dagger$ to be in $L_t^1 L_x^\infty$. In principle this might be feasible if we had the Strichartz estimate $\nabla^{1/2}\tilde{\phi} \in L_t^2 L_x^\infty$, but this estimate just barely fails to hold because of a logarithmic divergence in the frequencies. However, if we could somehow ensure that the derivative in $\tilde{\phi}\partial_\alpha\tilde{\phi}^\dagger$ always fell on a low-frequency component of $\tilde{\phi}$ and not on a high-frequency component then one would have a chance of iterating away the non-linearity. This will be accomplished by a renormalization using a co-ordinate frame adapted to $\phi$.

We begin by taking advantage of (5) to rewrite (4) in a form reminiscent of parallel transport:

$$\partial_\alpha \partial^\alpha \psi = 2A_\alpha \partial^\alpha \psi \tag{6}$$

where $A_\alpha$ is the matrix

$$A_\alpha := \partial_\alpha \tilde{\phi}\tilde{\phi}^\dagger - \tilde{\phi}\partial_\alpha\tilde{\phi}^\dagger.$$

Note that (6) exhibits more cancellation than (4), as $A_\alpha$ is now anti-symmetric. This type of trick is standard in the study of wave and harmonic maps, see e.g. [16], [13], [7], etc.

To solve (6), let us first consider the ODE analogue

$$\ddot{\psi} = 2A_0\dot{\psi}. \tag{7}$$

The matrix $A_0$ is anti-symmetric. Thus if we let $U(t)$ be the matrix-valued function solving the ODE

$$\dot{U}(t) = A_0 U(t)$$

with $U(0)$ initialized to the identity matrix (say), then we see that $\frac{d}{dt}(UU^\dagger) = 0$ and thus that $U$ remains orthogonal for all time. Indeed, one can view $U$ as the parallel transport of the identity matrix along the trajectory of $\tilde{\phi}$. Furthermore, since $\tilde{\phi}$ is smooth, we see that $U$ is also smooth, and in particular is much smoother than $\psi$. One can then use the linear change of variables $\psi = Uw$, and ignore terms which fail to differentiate the rough function $w$, to rewrite (7) as the trivial equation $\ddot{w} = 0$.

The ODE example of (7) suggests that (6) might be simplified by applying some orthogonal matrix $U$ to the wave $\psi$, or in other words by viewing $\psi$ in a carefully chosen co-ordinate frame. (This fits well with the corresponding experience of harmonic maps in [13]). Ideally, we would like $U$ to be carried by parallel transport by $\tilde{\phi}$ in all directions. More precisely, we would like $U$ to solve the PDE

$$\partial_\alpha U = A_\alpha U \tag{8}$$

for each $\alpha$. If we make the improbable assumption that $U$ obeyed (8) exactly for all $\alpha$, we can then substitute $\psi = Uw$ as before and ignore all terms which fail to differentiate the rough function $w$ to transform (6) to the free wave equation

$$\partial_\alpha \partial^\alpha w = 0$$

which we of course know how to solve.

Unfortunately, the system (8) of PDE is overdetermined, and in general has no solution (since the parallel transport connection induced by $\tilde{\phi}$ will have a small[9] but non-zero curvature). Nevertheless, it is possible to use Littlewood-Paley theory to construct a satisfactory *approximate* solution $U$ to (8). Specifically, we perform the Littlewood-Paley decomposition $\tilde{\phi} = \phi_{-M} + \sum_{-M<k} \phi_k$, where $M$ is a large number, $\phi_{-M}$ is the portion of $\phi$ on frequencies $|\xi| \lesssim 2^{-M}$, and $\phi_k$ is the portion on frequencies $|\xi| \sim 2^k$. We then define $U = U_{-M} + \sum_{-M<k} U_k$, where $U_{-M}$ is the identity matrix, and the $U_k$ are defined recursively by the formula

$$U_k := (\phi_k \phi_{<k}^\dagger - \phi_{<k}\phi_k^\dagger) U_{<k} \tag{9}$$

[9]More precisely, the curvature only contains terms which are quadratic in the first derivatives of $\tilde{\phi}$, as opposed to being linear in the second derivatives of $\tilde{\phi}$. This phenomenon seems specific to the wave maps equation; if one tries to apply the techniques here to (for instance) the Maxwell-Klein-Gordon or Yang-Mills equations at the critical Sobolev regularity, an obstruction arises because the connection $A$ has no reason to have a good curvature, regardless of the choice of gauge. At best one can place these equations in the Coulomb gauge, which was already known to be the most useful gauge to study these equations.

where $\phi_{<k}$, $U_{<k}$ are the functions

$$\phi_{<k} := \phi_{-M} + \sum_{-M<k'<k} \phi_k, \quad U_{<k} := U_{-M} + \sum_{-M<k'<k} U_k.$$

It then transpires that the matrix $U$ is approximately orthogonal and approximately satisfies (8), provided that the $\dot{H}^{n/2}$ norm of $U$ is sufficiently small and $M$ is sufficiently large. The point is that $\phi_k$ is a rougher function than $\phi_{<k}$, and so one can (heuristically) neglect terms where the derivative falls on $\phi_{<k}$ instead of $\phi_k$. Similarly for $U_k$ and $U_{<k}$. Thus we can morally differentiate (9) to obtain

$$\partial_\alpha U_k \approx (\partial_\alpha \phi_k \phi_{<k}^\dagger - \phi_{<k} \partial_\alpha \phi_k^\dagger) U_{<k} \tag{10}$$

and (8) follows by summing the telescoping series (and continuing to neglect the same type of terms as before). The approximate orthogonality of $U$ is based on the observation (from (9)) that $U_k^\dagger U_{<k} + U_{<k}^\dagger U_k = 0$. Summing this in $k$ and telescoping, we obtain

$$U^\dagger U = I + \sum_{k>-M} U_k^\dagger U_k.$$

The summation on the right-hand side then turns out to be negligible if we assume $\tilde{\phi}$ is small in $\dot{H}^{n/2}$, since this implies from Sobolev embedding that the $L^\infty$ norms of the $\phi_k$ (and hence the $U_k$) are small in $l^2$. (A similar argument can be used to dispose of the error terms which were neglected in (10)). If one then transforms (6) using $\psi = Uw$ as before, we obtain a non-linear wave equation for $w$, but all the terms in the non-linearity either contain expressions such as $\sum_k U_k U_k^\dagger$ which are quadratic[10] in the frequency parameter $k$, or have all derivatives falling on smooth functions rather than rough ones. Both types of terms turn out to be easily controlled by Strichartz estimates in higher dimensions $n \geq 5$. For $n < 5$ one has to use more sophisticated spaces (introduced by Tataru [36]) involving both Fourier space and physical space decompositions.

Klainerman and Rodnianski[22] have reinterpreted the above transform in the following appealing way. We wish to solve the equation (8). We can roughly telescope

$$A_\alpha \approx \sum_k \partial_\alpha \phi_k \phi_{<k}^\dagger - \phi_{<k} \partial_\alpha \phi_k^\dagger \approx \sum_k \partial_\alpha V_k$$

[10]Basically, such quadratic expressions effectively improve the Sobolev space $\dot{H}^{n/2}$ to the Besov space $\dot{B}_1^{n/2}$, which in principle can be treated by the arguments in [35].

where the potential $V_k$ is given by

$$V_k := \phi_k \phi_{<k}^\dagger - \phi_{<k} \phi_k^\dagger.$$

So we now want to solve

$$\partial_\alpha U \approx \sum_k \partial_\alpha V_k U.$$

If everything commuted, this would be solved by

$$U = \exp(\sum_k V_k).$$

However this doesn't quite work. If instead we use the variant

$$U = \prod_k \exp(V_k)$$

where the lower frequencies are on the left and the higher ones on the right, then it does work. This heavily exploits the fact that derivatives falling on low frequencies have small product (and hence mostly commute with) high frequency terms.

## 5. Acknowledgements

The author thanks Jim Colliander for helpful comments.

## References

[1] J. Bona, R. Smith, *The initial-value problem for the Korteweg-de Vries equation*, Philos. Trans. Royal Soc. London Series A **278** (1975), 555–601.

[2] J. Bourgain, *Fourier restriction phenomena for certain lattice subsets and applications to nonlinear evolution equations*, Part I, Geometric and Funct. Anal. **3** (1993), 107–156.

[3] J. Bourgain, *Refinements of Strichartz' inequality and applications to 2D-NLS with critical nonlinearity*, Int. Math. Research Notices, **5** (1998), 253–283.

[4] J. Bourgain, *Periodic Korteweg-de Vries equation with measures as initial data*, Selecta Math. **3** (1997), 115–159.

[5] J. Bourgain, *New global well-posedness results for nonlinear Schrödinger equations*, AMS Publications, 1999.

[6] J. Bourgain, W. Wang, *Construction of blowup solutions for the nonlinear Schrödinger equation with critical nonlinearity*, Dedicated to Ennio De Giorgi. Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **25** (1997), no. 1-2, 197–215.

[7] S.Y.A. Chang, L. Wang, P. Yang, *Regularity of Harmonic Maps*, Comm. Pure. Appl. Math. **52** (1999), 1099–1111.

[8] M. Christ, J. Colliander, T. Tao, *Local well-posedness and ill-posedness for KdV near the endpoint regularity*, preprint.

[9] J. Colliander, G. Staffilani, H. Takaoka, *Global well-posedness of the KdV equation below $L^2$*, Math Res. Letters **6** (1999), 755-778.

[10] J. Colliander, M. Keel, G. Staffilani, H. Takaoka, T. Tao, *Almost conservation laws and global rough solutions to a nonlinear Schrödinger equation*, in preparation.

[11] J. Colliander, M. Keel, G. Staffilani, H. Takaoka, T. Tao, *Sharp global well-posedness for periodic and non-periodic KdV and mKdV*, in preparation.

[12] C. Gu, *On the Cauchy problem for harmonic maps defined on two-dimensional Minkowski space*, Comm. Pure Appl. Math., **33**,(1980), 727–737.

[13] F. Helein, *Regularite des applications faiblement harmoniques entre une sur face et une varitee Riemannienne*, C.R. Acad. Sci. Paris Ser. I Math., **312** (1991), 591-596.

[14] O.A. Ladyzhenskaya, V.I. Shubov, *Unique solvability of the Cauchy problem for the equations of the two dimensional chiral fields, taking values in complete Riemann manifolds*, J. Soviet Math., **25** (1984), 855–864. (English Trans. of 1981 Article.)

[15] M. Keel, T. Tao, *Endpoint Strichartz Estimates*, Amer. Math. J. 120 (1998), 955–980.

[16] M. Keel, T. Tao, *Local and global well-posedness of wave maps on* $\mathbf{R}^{1+1}$ *for rough data*, IMRN **21** (1998), 1117–1156.

[17] C. E. Kenig, G. Ponce, L. Vega, *The Cauchy problem for the Korteweg-de Vries equation in Sobolev spaces of negative indices*, Duke Math J. **71** (1993), 1–21.

[18] C. Kenig, G. Ponce, L. Vega, *A bilinear estimate with applications to the KdV equation*, J. Amer. Math. Soc. **9** (1996), 573–603.

[19] S. Klainerman, *On the regularity of classical field theories in Minkowski space-time* $\mathbf{R}^{3+1}$, Prog. in Nonlin. Diff. Eq. and their Applic., **29**, (1997), Birkhäuser, 113–150.

[20] S. Klainerman, M. Machedon, *Smoothing estimates for null forms and applications*, Duke Math. J., **81** (1995), 99–133.

[21] S. Klainerman, M. Machedon, *On the optimal local regularity for gauge field theories,* Diff. and Integral Eq. **10** (1997), 1019–1030.

[22] S. Klainerman, I. Rodnianski, *On the global regularity of wave maps in the critical Sobolev norm*, to appear, IMRN.

[23] S. Klainerman, S. Selberg, *Remark on the optimal regularity for equations of wave maps type*, C.P.D.E., **22** (1997), 901–918.

[24] S. Klainerman, S. Selberg, *Bilinear estimates and applications to nonlinear wave equations*, preprint.

[25] S. Kuksin, *Infinite-dimensional symplectic capacities and a squeezing theorem for Hamiltonian PDEs*, CMP **167** (1995), 531–552.

[26] K. Nakanishi, *Local well-posedness and Illposedness in the critical Besov spaces for semilinear wave equations with quadratic forms*, Funk. Ekvac. **42** (1999), 261-279.

[27] J. Shatah, *Weak solutions and development of singularities of the* $SU(2)$ $\sigma$*-model.* Comm. Pure Appl. Math., **41** (1988), 459–469.

[28] J. Shatah, *The Cauchy problem for harmonic maps on Minkowski space*, in Proceed. Inter. Congress of Math. 1994, Birkhäuser, 1126–1132.

[29] J. Shatah, M. Struwe, *Geometric Wave Equations*, Courant Lecture Notes in Mathematics **2** (1998)

[30] J. Shatah, A. Tavildar-Zadeh, *On the Cauchy problem for equivariant wave maps*, Comm. Pure Appl. Math., **47** (1994), 719 - 753.

[31] M. Struwe, *Wave Maps, in Nonlinear Partial Differential Equations in Geometry and Physics*, Prog. in Nonlin. Diff. Eq. and their Applic., **29**, (1997), Birkhäuser, 113–150.
[32] T. Tao, *Ill-posedness for one-dimensional wave maps at the critical regularity*, Amer. J. Math. 122 (2000), 451–463.
[33] T. Tao, *Global regularity of wave maps I. Small critical Sobolev norm in high dimensions*, submitted, IMRN.
[34] T. Tao, *Global regularity of wave maps II. Small energy in two dimensions*, submitted, Comm. Math. Phys.
[35] D. Tataru, *Local and global results for wave maps I*, Comm. PDE **23** (1998), 1781–1793.
[36] D. Tataru, *On global existence and scattering for the wave maps equation*, Preprint, 1999.

Department of Mathematics, UCLA, Los Angeles, CA 90024
*E-mail address*: tao@math.ucla.edu