Invariant Gibbs measures and global strong solutions to 2D nonlinear Schrödinger equation

Andrea R. Nahmod

University of Massachusetts Amherst

Joint work with Yu Deng (USC) and Haitian Yue (USC)

25th PDE Seminar via Zoom

Org. by Prof. Quoc-Hung Nguyen
Institute of Mathematical Sciences, ShanghaiTech University

October 8th, 2020

The periodic NLS

Consider the defocusing nonlinear Schrödinger (NLS) equation on torus,

$$(i\partial_t + \Delta)\, u = |u|^{p-1}\, u, \quad (t,x) \in \mathbb{R} \times \mathbb{T}^d. \tag{1}$$

- (1) is an infinite dimensional Hamiltonian system with Hamiltonian

$$H[u](t) := \frac{1}{2} \int_{\mathbb{T}^d} |\nabla u|^2\, dx + \frac{1}{p+1} \int_{\mathbb{T}^d} |u|^{p+1}\, \mathrm{d}x = H[u](0). \tag{2}$$

- It also conserves the mass $m(u) := \int_{\mathbb{T}^d} |u|^2 dx$
- The scaling[1] critical threshold is

$$s_c := \frac{d}{2} - \frac{2}{p-1}$$

- For $s > \max(0, s_c)$ local well-posedness holds in H^s. For $s < s_c$ ill-posedness happens.

[1] $u(t,x)$ solution $\leftrightarrow$ $u_\lambda(t,x) = \lambda^{\frac{2}{p-1}} u(\lambda^2 t, \lambda x)$ solution, with same $\dot{H}^{s_c}$ norm.

NLS with random data

In fact, we'll consider the Wick ordered renormalization of (1) for p odd.

$$(i\partial_t + \Delta)u = :|u|^{p-1}u: \qquad (t,x) \in \mathbb{R} \times \mathbb{T}^d, \tag{3}$$

with random initial data

$$u(0) = f(\omega) = \sum_{k \in \mathbb{Z}^d} \frac{g_k(\omega)}{\langle k \rangle^\alpha} e^{ik\cdot x}, \qquad \alpha > 0. \tag{4}$$

- Here $\{g_k(\omega)\}$ are i.i.d. complex random Gaussian variables on a probability space $(\Omega, \mathcal{B}, \mathbb{P})$, that are normalized, i.e. $\mathbb{E}g_k = 0$ and $\mathbb{E}|g_k|^2 = 1$ and the law of g_k is rotationally symmetric.
- The renormalization $:|u|^{p-1}u:$ is needed when the solution has **infinite mass**. It will be defined later and we will make sense of its associated Hamiltonian[2]:

$$\mathcal{H}[u] := \frac{1}{2}\int_{\mathbb{T}^d} |\nabla u|^2\, dx + \frac{1}{p+1}\int_{\mathbb{T}^d} :|u|^{p+1}:\ \mathrm{d}x$$

- **In this work we'll set** $d = 2$, but many of our discussions work for general (d, p).

[2] Important: it is still a conserved quantity.

Random data

- The behavior of **generic** initial data should be more physically meaningful. A natural way to describe genericity is via a canonically defined measure.
- This is also linked to the **statistical ensemble** point of view: instead of **individual** solutions, we are interested in the **family** of solutions that are distributed according to some canonical law[3].
- Key point: for random initial data that are linear combinations of i.i.d centered Gaussian random variables, for which is well known that one can prove linear and nonlinear estimates that are far better than those available for arbitrary functions of the same regularity thanks to **large deviation estimates.** For example
- For **any** $p > 2$ one has with probability ~ 1 the linear estimate

$$\| \sum_k a_k g_k^\omega e^{ikx} \|_{L^p(\mathbb{T})} \lesssim \Big(\sum_k |a_k|^2 \Big)^{\frac{1}{2}}.$$

- And corresponding multilinear estimates; e.g. with probability ~ 1,

$$\| \sum_{k_1,\dots k_r} a_{k_1,\dots,k_r} \cdot g_{k_1}^\omega \cdots g_{k_r}^\omega e^{i(k_1+\dots k_r)\cdot x}) \|_{L^p} \lesssim \Big(\sum_{k_1,\dots,k_r} |a_{k_1,\dots,k_r}|^2 \Big)^{\frac{1}{2}}.$$

[3] Gaussian law (and alike) are certainly physically relevant ones.

What do we study?

From the PDE point of view, the fundamental question is well-posedness:

- Does the solution almost surely exist locally?
- Does the solution almost surely exist globally/have some long-time behavior?

where we mean existence in the **strong** sense; i.e. with suitable form of **uniqueness**.

From the statistical physics point of view, we are interested in the behavior of statistics ensemble as a whole:

- How does the law of distribution evolve under the flow, are there invariant measures?
- How do the ensemble averages evolve, is there ergodicity or mixing?

The Gibbs measure

The concept of Gibbs measure lies nicely at the intersection of these two directions. From the Hamiltonian structure of (3), it can be formally defined as

$$\text{`}\; \mathrm{d}\mu \sim e^{-\beta\,\mathcal{H}[u]} \prod_{x\in\mathbb{T}^d} \mathrm{d}u(x) \;\text{'}$$

where $\beta > 0$ is the inverse temperature (which will be normalized to 1).$\mathcal{H}[u]$ is the renormalization of the Hamiltonian,

$$d\mu \sim \underbrace{\exp\left[-\frac{1}{p+1}\int_{\mathbb{T}^2} :|u|^{p+1}:\,\mathrm{d}x\right]}_{\text{weight}} \cdot \underbrace{\exp\left[-\frac{1}{2}\int_{\mathbb{T}^2} |\nabla u|^2\,\mathrm{d}x\right]\prod_{x\in\mathbb{T}^2}\mathrm{d}x}_{\text{Gaussian measure} =:\, d\rho}$$

- This definition is purely *formal*, but as such $\mathrm{d}\mu$ is *formally invariant* under the flow of (3), due to a formal "Liouville's theorem" and conservation of the renormalized Hamiltonian $\mathcal{H}[u]$:
- In some cases this measure can be rigorously defined as a weighted Wiener measure[4].

[4]Strictly speaking the measure involves an additional weight which is an exponential of the L^2 mass. As the mass is also conserved, this does not affect any invariance properties.

Construction of Gibbs measure

Justifying the above definition of Gibbs measure (for some d and p) and studying their properties under various dynamics is a major problem of constructive quantum field theory. This intimately related to the so-called Φ_d^{p+1} model (Φ_d^4 if p=3).

Seminal work by: Glimm and Jaffe, Lebowitz, Rose and Speer, Simon, Nelson, Aizenman, Fröhlich, . . .)

- Dimensions $d = 1, 2$: can be done for any p ($d = 2, p = 3$, Φ_2^4 model)
- Dimension $d = 3$: can be done[5] for $p = 3$ (Φ_3^4 model).
- Dimension $d \geq 4$: cannot be done for any p (recently completed by Aizenman and Duminil-Copin).

In our case, $d = 2$, the Gibbs measure $\mathrm{d}\mu$ is **mutually absolutely continuous** with a Gaussian measure $\mathrm{d}\rho$, which is the law of distribution for the **random initial data** (4) with $\alpha = 1$. More precisely for the $\mathcal{S}'(\mathbb{T}^2)$-valued random variable:

$$f = f(\omega) : \omega \mapsto \sum_{k \in \mathbb{Z}^2} \frac{g_k(\omega)}{\langle k \rangle} e^{ik\cdot x}, \quad \omega \in \Omega. \tag{5}$$

[5] c.f Brydges-Fröhlich-Sokal. It's not expected for any other p

Invariance of Gibbs measure

After constructing the Gibbs measure, the next goal is to prove its invariance under the **dynamics** of (3) together with the almost sure global well-posedness (a.s. GWP) on its statistical ensemble.

The main difficulty is that the support of the Gibbs measure is rough.

- In dimension d, the support of $\mathrm{d}\mu$ is $H^{1-d/2-}$ (negative when $d \geq 2$);
  - This is determined by space where its associated Gaussian measure $\mathrm{d}\rho$ (Gaussian free field) is **countably additive**, which follows from a trace theorem for its covariance operator $Q_s := (Id - \Delta)^{-1+s}$.
- In particular, this regularity admits deterministic local well-posedness (i.e. subcritical) only when $d = 1$.
- The invariance of $\mathrm{d}\mu$ was justified for $d = 1$, all $p \geq 3$ odd (Bourgain '94) and for $d = 2$, $p = 3$ (Bourgain '96).
- There was essentially no progress in this particular problem since 1996, and a.s GWP and invariance (even) for $d = 2$, $p = 5$ remained open.

We solve this problem by proving:

Main Theorem [Yu Deng, A.N.– and Haitian Yue '19]

Let $d = 2$ and $p \geq 3$ odd. Then the renormalized NLS (3) is almost surely globally well-posed on the support of the Gibbs measure $\mathrm{d}\mu$ (which is in H^{0-}). The global flow Φ_t maps a full measure set Σ to itself, forms a one-parameter group (i.e. $\Phi_{t+s} = \Phi_t \Phi_s$), and keeps the Gibbs measure $\mathrm{d}\mu$ invariant under the flow:

$$\mu(E) = \mu(\Phi_t(E))$$

for any Borel set $E \subset \Sigma$.

- The key point is that, the solution we construct is **strong**.
  - It is the **unique** limit of smooth solutions coming from canonically –or smoothly– truncated systems.

Some remarks

- Weak solutions (with no uniqueness) that preserve $\mathrm{d}\mu$ (in some sense) were previously obtained by Oh-Thomann.
- The only case where the measure is constructed but not known to be invariant is $d = 3$, $p = 3$. We expect this to be much harder than $d = 2$, as this is **critical under –what we call– the probabilistic scaling** (more on this later at the end).
- Another open problem for Schrödinger is the invariance of white noise for $d = 1$, $p = 3$. This is also probabilistically critical but here one also has complete integrability.
- There are also results for (the more advantageous case of) wave equations, which we will not discuss in detail today.

Bourgain's invariance argument

The **lack of local well-posedness** due to rough support is the main difficulty in rigorously establishing invariance of the Gibbs measure $\mathrm{d}\mu$; in fact it's the **only** difficulty!

Once almost sure local well-posedness is proved on the support of $\mathrm{d}\mu$, invariance follows from the classical argument of Bourgain roughly as follows :

- First choose a parameter N and consider the truncated system

$$(i\partial_t + \Delta)u_N = \Pi_N(:|u_N|^{p-1}u_N:), \quad u_N(0) = \Pi_N f(\omega) \tag{6}$$

  where Π_N is the projection to frequencies not exceeding N.

- The above is a finite dimensional Hamiltonian ODE and possesses an **invariant** Gibbs measure $\mathrm{d}\mu_N$, hence well suited for global in time arguments for the truncated system.

- By almost sure local well-posedness, with high probability one can bound solutions to (6) uniformly in N on short time intervals.

- Invariance of $\mathrm{d}\mu_N$ then allows one to extend the above bounds to arbitrarily long time.

- These long-time bounds (which are still uniform in N) then allow one to define the global strong solution to (3) by

$$u = \lim_{N\to\infty} u_N, \qquad \text{in } H^{0-}(\mathbb{T}^2)$$

- Finally, as μ_N converges to μ in a suitable sense[6], one can take limits in the invariance for $\mathrm{d}\mu_N$ and obtain the invariance of $\mathrm{d}\mu$.

- We remark that the projection Π_N can be replaced by any (smooth or sharp) truncations, and the limit u does not depend on the choice of truncation sequence; i.e. this u is the **unique** limit of canonical approximations.

[6] The total variation of $\mu_N - \mu$ converges to 0

Almost sure local well-posedness?

By Bourgain's argument, to prove our Main Theorem it suffices to prove almost sure local well-posedness for (3) on the support of the Gibbs measure $\mathrm{d}\mu$, which is given by the law of the random variable

$$f(\omega) = \sum_{k \in \mathbb{Z}^2} \frac{g_k(\omega)}{\langle k \rangle} e^{ik\cdot x}. \tag{7}$$

This support is H^{0-} (i.e. in any $H^{-\varepsilon}$ but not L^2). This leads to two main difficulties:

- Infinite L^2 mass
- Supercriticality in deterministic scaling.

- The infinite L^2 mass implies for the potential energy that,

$$\mathbb{E}\Big[\int_{\mathbb{T}^2} |f(\omega)|^{p+1}\,\mathrm{d}x\Big] = \infty$$

with $f(\omega)$ as in (7), and that (for example) the nonlinearity $|u|^{p-1}u$ of (1) does not make sense almost surely as distributions. This "infinity" has to be removed by suitably renormalizing the nonlinearity, through a process called **Wick ordering**.

Such renormalizations by Wick ordering (as well as the more complicated ones) are quite common in constructive QFT.

Idea of renormalization by Wick ordering.

- Consider the truncation $\Pi_N f(\omega)$, we can easily calculate that

$$\sigma_N := \mathbb{E}\|\Pi_N f(\omega)\|_{L^2}^2 = \sum_{|k|\leq N} \frac{1}{\langle k\rangle^2} \sim \log N.$$

- If $p := 2r+1$, the Wick ordering of powers as polynomials is defined by

$$(:|u|^{p-1}u:)_N = \sum_{j=0}^{r} (-1)^{r-j} \binom{r+1}{j+1} \frac{r!{\sigma_N}^{r-j}}{j!} |u|^{2j} u,$$

$$(:|u|^{p-1}:)_N = \sum_{j=0}^{r} (-1)^{r-j} \binom{r}{j} \frac{r!{\sigma_N}^{r-j}}{j!} |u|^{2j}.$$

- For example the first few terms are

$$\begin{aligned}(:|u|^2:)_N &= |u|^2 - \sigma_N \\ (:|u|^2 u:)_N &= (|u|^2 - 2\sigma_N)u \\ (:|u|^4:)_N &= |u|^4 - 4\sigma_N|u|^2 + 2\sigma_N^2\end{aligned}$$

- It can then be proved that the limits

$$:|f(\omega)|^{p-1} f(\omega):= \lim_{N\to\infty} :|\Pi_N f(\omega)|^{p-1}\Pi_N f(\omega):, \quad :|f(\omega)|^{p-1}:= \lim_{N\to\infty} :|\Pi_N f(\omega)|^{p-1}:$$

almost surely exist (say in $H^{-\varepsilon}$). These lead to the renormalization (3) of (1).

Bourgain's method

Before describing our method, we first review the existing approaches of studying random data local well-posedness (Bourgain mid '90); as well as stochastically forced equations which are of a similar nature (Da Prato and Debussche early '00).

- The main idea is to make a linear-nonlinear decomposition, where the linear part is **rough** and **random**, and the nonlinear part is smoother.
- In both cases, the method can be understood as constructing solutions in a (random) **affine submanifold** i.e. a ball in a smoother space, **centered** at the random linear evolution u_{lin}.

- Take the NLS equation (3) with $d = 2$, $p = 3$ (Bourgain '96). Here the initial data (at the Gibbs measure level) is in H^{0-}, and is thus supercritical.
- However Bourgain was able to construct solutions of form

$$u = e^{it\Delta} f(\omega) + v$$

where v has **positive** regularity (Recall $u(0) = f(\omega)$ is the random initial data).

- That is the multilinear large deviation estimates we mentioned above, allow one to **re-center** the solution around **the linear evolution of random data and/or around higher order iterates**, and conclude that the difference between the two belongs to a Banach space of **higher regularity** than the one dictated by the (weaker) regularity of the random initial data.
  - More precisely, due to the exact Gaussian structure of the rough random part $u_{\mathrm{lin}} := e^{it\Delta} f(\omega)$, the interaction of u_{lin} with itself has much better regularity than one would naively expect; as we saw above.
  - The interaction of u_{lin} with v, and of v with itself, are under control as v has positive regularity.
- One may also add higher order Picard iterations (interactions of u_{lin} with itself) to the rough random part and hope for better regularities for v, but there is always a limit after which things stop improving!
  - Because of this, this approach by itself does not give optimal results in most cases.

Regularity structures and para-controlled calculus

Recently, two powerful methods have emerged in the context of stochastically forced **parabolic** equations:

- the **regularity structures** of Hairer, and
- the **para-controlled calculus** of Gubinelli, Imkeller and Perkowski.

Starting with the KPZ equation and the parabolic Φ_3^4 model (i.e. stochastic cubic heat equation), these have now become general theories that can solve a large class of subcritical equations.

- The regularity structures method is based on the local-in-space properties of solutions at fine scales (so it is particularly suitable for parabolic equations); it builds a general theory of distributions which includes the profiles coming from the noise, and allows one to perform multiplications and thus analyze the nonlinearity.

Para-controlled calculus

The theory of para-controlled distributions is a closer precursor of our approach, so we discuss in more detail.

The idea is to go back to the method of Bourgain and of Da Prato-Debussche, and identify the contributions in the smoother term v that have the lowest regularity.

- Such terms only come from the interactions between the **high** frequency of u_{lin} (and its interactions with itself) and the **low** frequency of v.

Thus, roughly speaking:

- In addition to the (random) linear evolution and possibly (suitably renormalized) higher order (random) expressions of the linear evolution (formally trees), one moves to the new 'center' terms that are 'para-controlled' by such expressions.

- That is one does not need the structured term to have an explicit multilinear Gaussian expression in order for the right estimates to hold. Rather one only needs them to be **para-controlled** by some explicit multilinear Gaussian expression.

Quick Definition

- f is said to be para-controlled by g if, up to some smoother 'remainder' terms, f looks like the **Bony para-product** between high frequencies of g and low frequencies of some auxiliary h.

- Namely,

$$f = \Pi_{>}(g,h) + R := \sum_{N} P_N g \cdot P_{\ll N} h + R,$$

where for dyadic frequencies N, P_N and $P_{\ll N}$ are the standard Littlewood-Paley operators projecting onto frequencies $\sim N$ and $\ll N$ respectively, and R is smoother than f.

- Roughly speaking: on small scales, f is expected to "behave like" g.

Recall:

$$FG = \underbrace{\sum_{N} P_N(F) P_{\ll N}(G)}_{\text{high-low paraproduct}} + \underbrace{\sum_{N} P_{\ll N}(F) P_N(G)}_{\text{low-high paraproduct}} + \underbrace{\sum_{N} P_N(F) P_N(G)}_{\text{diagonal/resonant term}}$$

A Stochastic PDE Example

- Consider the Φ_3^4 model (cubic heat equation on $\mathbb{T}^3$ with white noise forcing):

$$(\partial_t - \Delta)u = u^3 - 3Cu + \xi$$

  $\xi =$ space-time white noise. We ignore the linear renormalization term below:

- For $\mathcal{I} = (\partial_t - \Delta)^{-1}$ (Duhamel) and $Z = \mathcal{I}\xi$, modulo details, one constructs:

$$u = \underbrace{Z + \mathcal{I}(P_3(Z))} + \mathcal{I}\,\Pi_{>}(P_2(Z), u - Z) + R,$$

  $P_m(Z) =$ explicit multilinear Gaussian components given in terms of (suitably renormalized) powers of Z (trees); and $R =$ remainder (higher regularity).

- $Z \in C^{-1/2-}$ (noise is at $C^{-5/2-}$ level, heat kernel gains 2 derivatives)
- The self interaction $\mathcal{I}(P_3(Z)) \in C^{1/2-}$.
  - These are terms in the method of Bourgain and Da Prato-Debussche.
- The term $\mathcal{I}\,\Pi_{>}(P_2(Z), u - Z)$ is para-controlled by an explicit multilinear Gaussian and is in C^{1-}

- The **para-controlled structure** ensures that $\mathcal{I}\pi_{>}(Z^2, u - Z)$ behaves essentially like $\mathcal{I}(Z^2)$, which has the exact Gaussian structure, so the product makes sense due to Gaussian estimates, modulo renormalization.
- In practice, the smooth(er) remainder R is constructed in a smooth space $C^{(5/4)-}$.
- **Note**: Here then the solution belongs to a (random) sub-manifold which is much more nonlinear than in Bourgain's or DaPrato-Debussche methods

We remark that a recent work of Gubinelli-Koch-Oh applied this method to a stochastic wave equation.

Difficulties for Schrödinger

Now return to the 2D Schrödinger problem with $p \geq 5$.

Compared to heat equations, Schrödinger can only be studied in Sobolev H^s spaces (as opposed to Hölder C^s spaces since Schrödinger kernel is not even bounded there), and the Duhamel evolution has **no smoothing**.

- Suppose one were to follow the approach of Bourgain and write

$$u = u_{\mathrm{lin}} + v = \text{linear evolution of r.d.} + v$$

  then, as observed by Bourgain, v can only be put at best in $H^{\frac{1}{2}-}$, which for us is (better that $H^{-\varepsilon}$ but still) supercritical, and cannot close the estimate by itself.

- One can show however that this poor regularity comes **only from high-low interactions** (like the one discussed in para-controlling theory)[7]

[7] If the two highest frequencies in the input factors of (say) $\mathcal{I}(|u_{\mathrm{lin}}|^{p-1} u_{\mathrm{lin}})$ are comparable then we can place it in $X^{1-,\frac{1}{2}+}$

- So one may try to identify a term X from v of the form:

$$X = \mathcal{I}\pi_{>}(\, u_{\mathrm{lin}}\ ,\ :|u|^{p-1}:\,),$$

where now $\mathcal{I} = (i\partial_t - \Delta)^{-1}$, and hope that X has a random structure "like" u_{lin} and that $Y = v - X$ is smoother.

- But in order to expect X to be "para-controlled" by u_{lin} one would need to have reasonable bounds on the **low frequency** components of $:|u|^{p-1}:$ which **themselves contain** a part of $:|X|^{p-1}:$

- The regularity of X is a priori only $H^{1/2-}$ which is supercritical, and there is **no way** of controlling $:|X|^{p-1}:$ assuming only this (its regularity)[8].
  - This is in contrast with heat and wave equations, where X has higher regularity due to smoothing (order two for heat and order one for wave), so the low-frequency part is always a nice function.

- Instead one needs to **zoom in/unveil and invoke the structure** of X.

- But how can this be done?

[8] If its high is very high, then there is no way of controlling high to a (big) power.

Induction on frequency

We will perform an induction on frequency, and validate the behavior of lower frequency parts of X before validating the high frequency part of X.

- By validate we mean that if X is as above we expect X to have a random structure "like" u_{lin} and that we are proving certain estimates as we will see below.

- For simplicity let us write (where P_N are Littlewood-Paley projections)

$$P_N X = \sum_{L \ll N} \mathcal{I}\big(P_N u_{\text{lin}} \cdot :|P_L u|^{p-1}: \big).$$

- In order to validate the behavior of $P_N X$, the main difficulty comes from $:|P_L u|^{p-1}:$ or more precisely, from the $:|P_L X|^{p-1}:$ piece in it.

- So one needs to first understand and validate the behavior of $P_L X$.

Gaining independence

In the above arguments, in order to apply multilinear Gaussian estimates, one would like to have **independence** between **the high frequency part** $P_N u_{\mathrm{lin}}$ and **the low frequency part** $:|P_L u|^{p-1}:$

- This is not needed in the para-controlled theory (applied to heat and wave equations), again due to smoothing.
- The idea is to replace $|P_{\ll N} u|^{p-1}$ by $|u_{\ll N}|^{p-1}$, where $u_{\ll N}$ is the solution to (3) with initial data $P_{\ll N} f(\omega)$ instead of $f(\omega)$.
  - The difference produced will be a "high-high" interaction and this produces acceptable terms.
- The point is that $|u_{\ll N}|^{p-1}$ is a measurable function of $\{g_k(\omega)\}_{\langle k \rangle \lll N}$. Thus it is automatically independent with $P_N u_{\mathrm{lin}}$.
- This idea also appeared in a recent work of Bringmann, on derivative nonlinear wave equations.

Recursive reduction?

At this point, one could try to make a recursive reduction. Namely write $P_N X$ as a (multilinear) expression of $P_L X$ for $L \ll N$ (up to some remainder terms that presumably can be controlled), then iterate and write $P_L X$ as an expression of $P_R X$ for $R \ll L$, and so on, until the frequency reaches 1.

- **However**, this iteration process will lead to some complicated tree expansion, and the amount of combinatorics will be overwhelming.
- Moreover, this tree expansion involves all different frequency scales from N to 1, and it will be hard to treat them simultaneously.
- This approach may not be impossible for -say- $p = 5$, but we will seek another unified method.

Capturing the implicit randomness

The goal is to capture the implicit randomness structure of $P_N X$ in some norm or quantity that **propagates**[9]; i.e. that allows for an induction from frequencies $L \ll N$ to frequency N.

- Clearly the $H^{\frac{1}{2}-}$ norm of $P_N X$ will not be enough to do the job, since this is a supercritical norm (the critical norm is $H^{1-\frac{2}{p-1}}$, $p \geq 5$).
- It turns out that, to find the right quantity, we need to **shift the point of view** from the **term**
$$P_N X = \sum_{L \ll N} \underbrace{\mathcal{I}(P_N u_{\text{lin}} \cdot :|P_L u|^{p-1}:)}$$
to the **operator**
$$\mathcal{P}_{NL} : y \mapsto \mathcal{I}(P_N y \cdot :|P_L u|^{p-1}:).$$
- In the para-controlled theory, thanks to smoothing of heat kernel, one can study X directly as a random variable under some norms. Here, instead, we treat X as a sum of the operator $\mathcal{P}_{NL}$'s acting on $P_N u_{\text{lin}}$ and study the structures of $\mathcal{P}_{NL}$, which will help us run our ansatz by induction.

[9]Propagates means that if you can prove the bounds for the low frequency part then you can prove the bound for the higher frequency part.

Random averaging operators

- This $\mathcal{P}_{NL}$ operator depends on (a multilinear expression of) $P_L u$, which has an implicit random structure. Thus we will view it as a random operator.

- On the other hand, it has (basically) a para-product structure, so its **effect** is essentially a **projection to frequency N, followed by a weighted average over smaller scales $L \ll N$**, say when applied to the Gaussian free field u_{lin}.

- For this reason we will call it a **random averaging operator** (matrix operator), and is the **key** to propagating the right probabilistic bounds.

- The question now becomes, what is the right norm or quantity we should estimate for $\mathcal{P}_{NL}$?

Norms and a priori bounds

- Many candidates, but two norms that precisely capture the randomness structure, are the **operator norm** and **Hilbert-Schmidt** norms for operators from $L^2 \to L^2$.
  - Since we are dealing with operators between spacetime functions, we will replace L^2 by the natural spacetime variant, namely the (Fourier restriction) $X^{s,b}$ spaces.
- The right a priori bounds that propagate, are then

$$\|\mathcal{P}_{NL}\|_{\mathrm{OP}} \lesssim L^{-\delta_0}, \quad \|\mathcal{P}_{NL}\|_{\mathrm{HS}} \lesssim N^{1/2+\delta_1} L^{-1/2}, \tag{8}$$

  for some $1 \gg \delta_0 \gg \delta_1$.

- The Hilbert-Schmidt norm bound guarantees that $P_N X = \sum_L \mathcal{P}_{NL} u_{\mathrm{lin}}$ belongs to $H^{1/2-}$ as expected; but as we mentioned we cannot use this directly. The operator norm bound, on the other hand, is the key norm that allow us to propagate.

- If we had general functions in $X^{\frac{1}{2}-,\frac{1}{2}+}$ as low inputs then we would never be able to prove these two estimates. So these really capture the randomness structure of the low part. They are also tractable and simple to use.

The full ansatz (I)

Now we can write down the full ansatz of the solution to (3):

$$u = u_{\mathrm{lin}} + \sum_{N} \sum_{L \ll N} \mathcal{P}_{NL} u_{\mathrm{lin}} + w, \tag{9}$$

where $\mathcal{P}_{NL}$ are the random averaging operators, whose coefficients are independent with $P_N u_{\mathrm{lin}}$, and satisfy the bounds (8).

The remainder term w belongs to the subcritical space H^{1-}, or more precisely some $X^{1-,b}$ space.

The proof then schematically goes as follows:

- We induct on frequency to show that $\mathcal{P}_{NL}$ satisfies (8) and w is bounded in H^{1-}.

The full ansatz (II)

- Suppose this is true for frequency $\ll N$, then in particular for $L \ll N$, $P_L u$ can be written as:

  a linear term $P_L u_{\mathrm{lin}}$, the high-low term $\sum_{R \ll L} \mathcal{P}_{LR} u_{\mathrm{lin}}$ which now behaves like $\mathcal{I} u_{\mathrm{lin}}$, and a smooth term $P_L w \in H^{1-}$.

- Then, the coefficients of the operator $\mathcal{P}_{NL}$, which are multilinear expressions of $P_L u$, can be bounded using (new) multilinear Gaussian estimates, counting lemmata, and TT^*-type arguments[10] from which (8) follows for N (and L).

- Finally, since $P_N w$ only contains high-high interactions (as the bad high-low interactions are captured by $\mathcal{P}_{NL} u_{\mathrm{lin}}$), we can apply a contraction mapping estimate to prove that $\mathcal{P}_N w \in H^{1-}$. This completes the proof.

[10] à la Bourgain

The full ansatz (III)

More details on the full ansatz:

- In fact, $L \ll N$ means $L \leq N^{1-\delta}$. Then we define an operator:

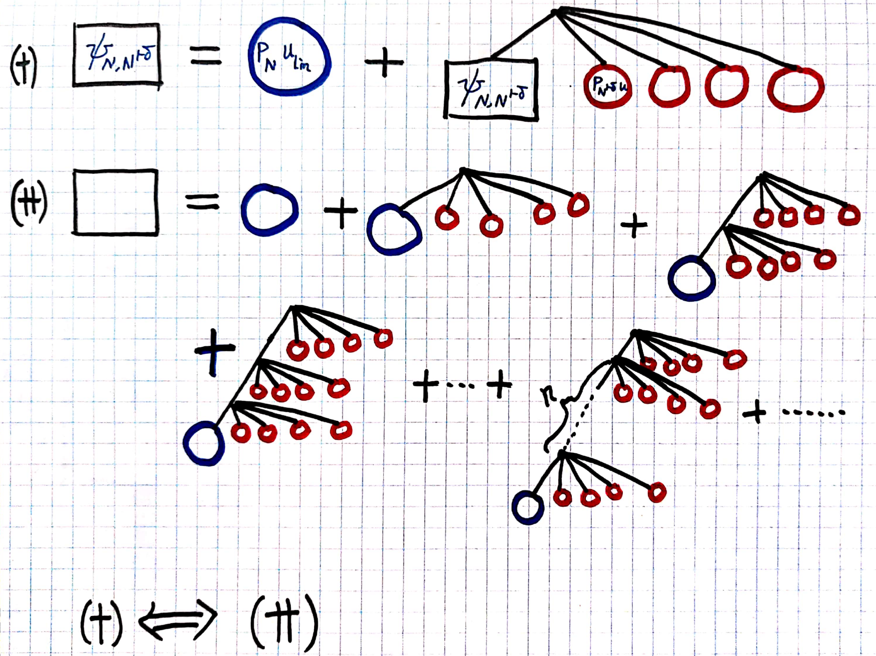
$$y \to \mathcal{L}^N(y) := \Big(\sum_{L \leq N^{1-\delta}} \mathcal{P}_{NL} \Big)(y) \;=\; \mathcal{I}(P_N y \cdot :|P_{<N^{1-\delta}} u|^{2r}:).$$

- The 'true' full ansatz is

$$\begin{aligned} u &= \sum_N \underbrace{\big(I + \mathcal{L}^N + (\mathcal{L}^N)^2 + (\mathcal{L}^N)^3 + \cdots\big)\,(P_N u_{\mathrm{lin}})}_{} + w \\ &= \sum_N \qquad\qquad \psi_{N,N^{1-\delta}} \qquad\qquad + w. \end{aligned} \tag{10}$$

- The above infinite series form of $\psi_{N,N^{1-\delta}}$ can be also written as $(I - \mathcal{L}^N)^{-1} P_N u_{\mathrm{lin}}$ and hence we have an implicit expression:

$$\psi_{N,N^{1-\delta}} \;=\; P_N u_{\mathrm{lin}} + \mathcal{L}^N(\psi_{N,N^{1-\delta}}).$$


(†)
$\psi_{N,N^{1-\delta}}$
$P_N u_{lin}$
(††)
n
(†) ⟺ (††)

Matrix structures in $\psi_{N,N^{1-\delta}}$

By taking Fourier transform in the construction of $\psi_{N,L}$, we have the Fourier coefficients of $\psi_{N,L}$ in the form:

$$\widehat{\psi}_{N,L}(k) = \sum_{k^*} h^{N,L}_{kk^*} \frac{g_{k^*}(\omega)}{\langle k^* \rangle}, \qquad \textit{colored gaussians}$$

where for $\frac{N}{2} < \langle k^* \rangle \leq N$ and $\langle k \rangle \leq N$. In particular, $h^{N,L}_{kk^*} = \widehat{\varphi}(k)$ which is the k-th mode of the solution φ to the equation:

$$\varphi = e^{i(k^* \cdot x + |k^*|^2 t)} + \mathcal{L}^N(\varphi).$$

The full ansatz (10) equipped with the following operator bounds: for any $L \leq N^{1-\delta}$,

$$\|h^{N,L}\|_{\mathrm{OP}} \leq L^{-\delta_0}; \tag{11}$$

$$\|h^{N,L}\|_{\mathrm{HS}} \leq N^{\frac{1}{2}+\gamma_0} L^{-\frac{1}{2}}; \tag{12}$$

$$\left\| \left(1 + \frac{|k - k^*|}{L}\right)^{\kappa} h^{N,L}_{kk^*} \right\|_{\mathrm{HS}} \leq N. \tag{13}$$

can be proved by induction.

Deterministic scaling $\rightarrow$ *Probabilistic scaling*

- The scaling critical threshold for the quintic NLS in $d = 2$ is $H^{\frac{1}{2}}$; in fact, it is $H^{1-\frac{2}{p-1}}$ which approaches H^1, for large p (recall $s_c = \frac{d}{2} - \frac{2}{p-1}$).
  - So in our result in $d = 2$ we gained (at least) $\frac{1}{2}+$ for quintic and essentially **one full derivative** for p large.

Why is such a gain plausible? Heuristic argument (for general d and p) which in fact suggests may be improved:

- Consider the NLS equation (1) in the **deterministic** setting (we omit the renormalization). Suppose the initial data is

$$u_0 = \sum_{|k|\sim N} N^{-\alpha} e^{ik\cdot x},$$

  then u_0 is bounded in H^s for $s = \alpha - \frac{d}{2}$.
- In order for local well-posedness to hold, a guiding principle is that the second Picard iteration should have at least the same regularity as the linear evolution:

$$u^{(2)}(t) = \int_0^t e^{i(t-t')\Delta}(|e^{it'\Delta}u_0|^{p-1}e^{it'\Delta}u_0)\,\mathrm{d}t'$$

  should be bounded in the same H^s.

By making Fourier expansions, we essentially get[11]

$$\mathcal{F}_x u^{(2)}(k) \sim \sum_{\substack{k_1 - \cdots + k_p = k \\ |k_j| \lesssim N}} \frac{1}{\langle \Sigma \rangle} \prod_{j=1}^{p} \widehat{u_0}(k_j), \tag{14}$$

where

$$\Sigma := |k|^2 - |k_1|^2 + \cdots - |k_p|^2.$$

Essentially[12]

$$|\mathcal{F}_x u^{(2)}(k)| \lesssim N^{-p\alpha} \sup_{m \in \mathbb{Z}} \# S_m,$$

where $S_m = \{(k_1, \cdots, k_p) \,:\, k_1 - \cdots + k_p = k,\ |k_j| \lesssim N,\ \Sigma = m\}$.

By dimension counting one expects $\# S_m \lesssim N^{(p-1)d-2}$.

So, in order for $\|u^{(2)}\|_{H^s} \lesssim 1$ we need

$$-p\alpha + (p-1)d - 2 \leq -\alpha,$$

or equivalently $s \geq s_c$, the critical scaling threshold.

[11] We omit complex conjugates

[12] Up to log factors

The probabilistic scaling

Now we add randomness to initial data, say

$$u_0 = \sum_{|k|\sim N} N^{-\alpha} g_k(\omega) e^{ik\cdot x}.$$

- Then $(\mathcal{F}_x u^{(2)})(k)$ would be a sum of products of independent Gaussians.
- It is well-known that in such cases we have a“square root” gain[13] akin to the CLT, i.e. large deviation-type estimates give

$$\left| \sum_{(k_1,\cdots,k_{2r+1})\in S} g_{k_1}(\omega)\cdots g_{k_{2r+1}}(\omega) \right| \sim (\#S)^{1/2}.$$

- This improvement gives

$$|(\mathcal{F}_x u^{(2)})(k)| \sim N^{-p\alpha} \big(\sup_m \#S_m \big)^{1/2} \lesssim N^{-p\alpha + \frac{(p-1)d}{2} - 1}$$

- This leads to the new threshold

$$\alpha \geq \frac{d}{2} - \frac{1}{p-1}, \qquad \text{or} \qquad s \geq -\frac{1}{p-1} := s_p,$$

  which we call the threshold for **probabilistic scaling**. Note $s_p \leq s_c$ and s_p is independent of d. For the quintic in 2D $s_p = -\frac{1}{4}$.

[13] as opposed to $\#S$ if without randomness.

It was thus reasonable to propose the following:

Conjecture in [Deng-N.-Yue 19] $\to$ Theorem [Deng-N.Yue 20]

Given $d \geq 1$, and odd $p \geq 3$, then the renormalized NLS (3) is almost surely locally well-posed with random initial data given by (4), namely

$$u(0) = f(\omega) = \sum_{k \in \mathbb{Z}^d} \frac{g_k(\omega)}{\langle k \rangle^\alpha} e^{ik\cdot x},$$

for any fixed

$$\alpha > s_p + \frac{d}{2}, \qquad s_p := -\frac{1}{p-1}$$

- The random initial data (4) is in H^{s-} where $s := \alpha - (d/2) > s_p$.
- (The local part of) our main theorem is a special case of the conjecture.
- In our case, $d = 2$, as $p \to \infty$, H^{s_p} approaches the support of the Gibbs measure (which is just below L^2), so our theorem is "asymptotically sharp" in this limit.
- Our Theorem above barely misses a.s. LWP for the $d = 3$ cubic NLS in $H^{-\frac{1}{2}-}$. Note in this case $s_p = -\frac{1}{2}$.

Final Remarks: As a byproduct of the proof one has that for smooth well prepared random data (e.g. $\leftrightarrow$ random data arising in derivation of WKE in wave turbulence theory) the time of existence is longer than the deterministic one (no energy cascade until a very **long time** (longer than deterministic time). That is, randomization effectively extends the time of perturbative regime.

Suppose one has well-prepared data at frequency N with H^s norm ~ 1, and assume $p = 3$. Then in the deterministic case, the first energy cascade happens at the timescale of CR equation (see Faou-Germain-Hani and Buckmaster-Germain-Hani-Shatah) which is $N^{2(s-s_c)}$. On the other hand, if the Conjecture is true, then in the randomized case, the first energy cascade only happens at the much later timescale $N^{2(s-s_p)-}$.

Theorem (Deng-N.-Yue 2020)

Fix (s, α) as before, let N be dyadic and φ be Schwartz. Let u solve (3) with *random homogeneous data* defined by

$$u(0) = f(\omega) = N^{-\alpha} \sum_k \varphi\left(\frac{k}{N}\right) g_k(\omega) e^{ik\cdot x}.$$

Then, with high probability, there is no energy cascade between Fourier modes, i.e. $|\widehat{u}(t,k)|^2 \approx |\widehat{u}(0,k)|^2$ with negligible error for large N, up to the time $T = N^{(p-1)(s-s_{pr})-}$. *This is expected to be sharp.*