Journées

ÉQUATIONS AUX DÉRIVÉES PARTIELLES

Obernai, 11–15 juin 2018

Daniel Tataru

The threshold theorem for geometric nonlinear wave equations

J. É. D. P. (2018), Exposé n° X, 15 p.

<http://jedp.cedram.org/item?id=JEDP_2018____A10_0>

Journées *Équations aux dérivées partielles*
Obernai, 11–15 juin 2018
GDR 2434 (CNRS)

The threshold theorem for geometric nonlinear wave equations

Daniel Tataru

Abstract

The aim of these notes is to provide a brief overview of a large body recent work whose aim was to prove the Threshold Theorem for energy critical geometric nonlinear wave equations. Within the class of geometric wave equations we include nonlinear wave evolutions which have a geometric structure and origin, including a nontrivial gauge group. The problems discussed here include Wave Maps, Maxwell–Klein–Gordon, as well as the hyperbolic Yang–Mills flow. In a nutshell, the Threshold theorem asserts that these problems are globally well-posed for initial data below the ground state energy.

1. Introduction

1.1. An overall picture

The equations considered in these notes can be broadly described as semilinear wave equations on the Minkowski space-time $\mathbb{R}^{n+1}$, schematically of the form

$$\Box u = N(u), \qquad u(0) = u_0, \quad u_t(0) = u_1 \tag{1.1}$$

where N may also depend on ∇u. They can be seen as an evolution equation for the position/velocity pair

$$u[t] = (u(t), \partial_t u(t))$$

in either homogeneous or inhomogeneous Sobolev spaces $H^s \times H^{s-1}$. They all share several standard features:

- *Lagrangian description:* the equations can be described as critical points for a Lagrangian

$$\mathcal{L}(u) = \int_{\mathbb{R}^{n+1}} L(u, \nabla u)\, \mathrm{d}x\mathrm{d}t.$$

- *Conserved energy $E(u)$:* this roughly measures the $\dot{H}^1 \times L^2$ norm of the solutions

$$E(u(t)) \approx \|u(t)\|_{\dot{H}^1}^2 + \|u_t(t)\|_{L^2}^2$$

- *Scaling law:* the equations are invariant with respect to a scaling law

$$u(x,t) \to \lambda^a u(\lambda x, \lambda t)$$

- *Lorentz invariance:* The equations (and the Lagrangian) are invariant with respect to the Lorentz group of linear transformations.

D.T. was supported by the NSF grant DMS-1800294 as well as by a Simons Investigator grant from the Simons Foundation.
Keywords: nonlinear wave equation, energy critical, threshold theorem, renormalization.

The scaling law allows one to select a distinguished initial data space, called the *critical Sobolev space*, $\dot{H}^{s_c} \times \dot{H}^{s_c-1}$, where the critical Sobolev index depends on the dimension and the scaling law,

$$s_c = \frac{n}{2} - a$$

This is often the most natural space to study such equations, and is also the setting where the local in time and the global in time properties of the evolution coincide. In this space it is meaningful to distinguish between

- *Small data*, for which the analysis is at least to a certain extent perturbative.
- *Large data*, where genuinely nonlinear dynamics may play a major role.

The energy functional $E(u)$ plays an important role in this story. It corresponds to the Sobolev index $s = 1$, and thus it is invariant with respect to scaling if and only if $s_c = 1$. We call such problems *energy critical*. By extension, problems with $s_c > 1$ are called energy supercritical, and those with $s_c < 1$ are called energy subcritical.

The small data problems share similar features regardless of energy criticality. However, for large data problems, which is our objective here, there is not much one can say at this point in the energy supercritical case, because there is no good strategy to control the growth of the critical Sobolev norm; one potential phenomena here is that of *frequency cascade* where the energy possibly migrates to higher frequencies at a rate which may cause growth of the critical Sobolev norm. A similar phenomena could potentially occur in the energy subcritical case, but where the energy may possibly migrate to lower frequencies. For these reasons in what follows we will focus on the energy critical case.

Consider now an energy critical nonlinear wave equation, where the goal is to describe the global dynamics in terms of its energy. Heuristically one expects that global well-posedness holds for energy below a threshold, whereas at that threshold some obstruction arises. There are two natural obstructions one may immediately think of:

- *Solitons.* The simplest are the steady states, which exist in many such problems, and solve a nonlinear elliptic equation. We will call the smallest energy steady state the *ground state*, and its energy the *ground state energy*. In addition to steady states, one also should consider here their Lorenz transforms, which still maintain a fixed profile, but which moves with a constant velocity v with $|v| < 1$. These are less important in a first analysis, as their energy usually increases with their speed $|v|$.
- *Self-similar solutions.* These exist in many nonlinear wave models, but in energy critical problems they usually have infinite energy, so they are not so much a part of our story. Still, proving this is always an important part of the process.

Based on the above discussion, one is led to the following broad conjecture, which in some form or another has been around for the better part of 30 years:

Conjecture 1.1 (Threshold conjecture)**.** *Energy critical equations have global well-posedness and scattering below the ground state energy.*

Above the ground state energy, one may imagine that solutions might naturally separate into solitons and a dispersive part which is akin to the above global solutions. The story is not as simple because the ground state is not an isolated object as it comes together with its orbit via the symmetry group (usually scaling and translations). While some orbital stability property often holds, the scaling group is noncompact and there are many examples of finite time blow-up along a scaled family of solitons e.g. [10, 11, 23, 24].

Nevertheless, one does expect that some form of the following broad conjecture holds:

Conjecture 1.2 (Soliton resolution conjecture)**.** *Large data solutions for energy critical equations separate into solitons and a global radiation part.*

For now this seems to be out of reach in full generality, and exists only as a broad philosophical goal. To make it more precise one needs to take into account at the very least the following issues:

- Blow up in finite time can occur, in which case the conjecture should apply to the solution near blow-up time.

- The solitons should not be taken individually, but rather within their orbit provided by the group of symmetries, with a time dependent modulation parameter describing the evolution along the orbit.

- If the energy functional is not coercive then the scenario where the critical norm of the solution blows up is also possible; then the conjecture should apply only to solutions which remain bounded in the critical norm.

Nevertheless, a weaker form of the conjecture is more tractable:

Conjecture 1.3 (Dichotomy conjecture)**.** *Large data solutions for energy critical equations are either global and scattering or they bubble off a soliton.*

Bubbling off means that along a sequence of times one obtains a soliton in the limit up to scaling and translation. As above, here the bubbling may occur either at the blow up time, or at infinity for global solutions.

Within the class of nonlinear wave equations, the geometric nonlinear wave equations play a proeminent role. These are equations which arise from Lagrangians which have a geometric nature, and as such the have both a rich array of structural properties and additional layers of complexity in the analysis. We will describe some of these equations shortly, but for now we list some of their features:

- They all have a large gauge group, which in particular implies that before any detailed analysis on has to carry out the additional step of *gauge fixing*.

- The state space for the evolution is not linear, and instead it can be seen (after gauge fixing) as an infinite dimensional manifold. Because of this, many perturbative methods cannot be easily applied.

- The state space is often not connected. The flow stays within the same connected component, called *topological class*. In particular the ground state does not belong to the trivial topological class, which affects the ground state energy threshold in the threshold conjecture.

- Even though at high regularity the equations are semilinear, at critical regularity they become quasilinear, meaning that the solutions are only expected to have continuous dependence on the initial data.

1.2. Three linear wave equations

Before we describe our nonlinear wave equations, we begin with three linear models which will play a role in the nonlinear case:

The wave equation for real valued functions $u : \mathbb{R}^{n+1} \to \mathbb{R}$. The Lagrangian has the form:

$$\mathcal{L}(u) = \int \partial^\alpha u \cdot \partial_\alpha u \, \mathrm{d}x\mathrm{d}t,$$

and its Euler–Lagrange equation is the linear wave equation

$$\Box u = 0.$$

where the D'Allembertian is $\Box = \partial^\alpha \partial_\alpha$ with the standard summation convention and indices raised using the Minkowski metric.

The Maxwell equation for 1*-forms.* Here the object of study is the electromagnetic potential

$$A_\alpha : \mathbb{R}^{n+1} \to \mathbb{R}.$$

This induces a covariant differentiation on complex valued functions

$$D_\alpha = \partial_\alpha + iA_\alpha,$$

with associated curvature

$$F_{\alpha\beta} = \partial_\alpha A_\beta - \partial_\beta A_\alpha.$$

The associated Lagrangian is

$$\mathcal{L}(A) := \frac{1}{2}\int_{\mathbb{R}^{4+1}} F_{\alpha\beta}\cdot F^{\alpha\beta}\ \mathrm{d}x\mathrm{d}t.$$

and as Euler–Lagrange equations we get the Maxwell system:

$$\partial^\alpha F_{\alpha\beta} = 0.$$

Here we have our first example of gauge independence,

$$A \to A + db, \qquad b : \mathbb{R}^{n+1} \to \mathbb{R}.$$

The covariant wave equation for functions $\phi : \mathbb{R}^{n+1} \to \mathbb{C}$. For a given electromagnetic potential A and its associated covariant differentiation we define the Lagrangian:

$$\mathcal{L}(\phi) = \int D^\alpha\phi\cdot\overline{D_\alpha\phi}\ \mathrm{d}x\mathrm{d}t.$$

The corresponding Euler–Lagrange equation is the covariant wave equation

$$\Box_A\phi = 0, \qquad \Box_A = D^\alpha D_\alpha.$$

Here we note that changing the gauge for the electromagnetic potential corresponds to a matching gauge change for ϕ,

$$\phi \to \phi e^{ib}$$

1.3. The nonlinear wave equation

For $p > 1$ and real valued functions

$$u : \mathbb{R}^{n+1} \to \mathbb{R}$$

we define the Lagrangian

$$L(\phi) = \int \frac{1}{2}\partial^\alpha u\cdot\partial_\alpha u \pm \frac{1}{p+1}|u|^{p+1}\ \mathrm{d}x\mathrm{d}t$$

The Euler–Lagrange equation is the nonlinear wave equation

$$\Box u = \pm u|u|^{p-1} \tag{NLW}$$

which has the conserved energy

$$E(u) = \int \frac{1}{2}(|\partial_t u|^2 + |\nabla_x u|^2) \pm \frac{1}{p+1}|u|^{p+1}\mathrm{d}x$$

The energy critical case corresponds to the exponent

$$p = \frac{n+2}{n-2}, \qquad n \geq 3$$

The $+$ sign is the *defocusing* case, where the energy is convex and coercive, while the $-$ sign is the *focusing* case, where the energy is no longer coercive.

In the defocusing case there is no ground state, whereas in the defocusing case the ground state is obtained as the first critical (saddle) point for the energy, and solves

$$-\Delta Q = Q|Q|^{p-1}$$

This is an explicit algebraic function, which is unique up to symmetries.

1.4. Three geometric wave equations

Wave maps. Here the objects of study are maps into a Riemannian manifold,

$$\phi : \mathbb{R}^{n+1} \to (M, g).$$

with the associated Lagrangian

$$\mathcal{L}(\phi) = \int \langle \partial^\alpha \phi, \partial_\alpha \phi \rangle_g \, \mathrm{d}x\mathrm{d}t.$$

In this context $\partial_\alpha \phi \in T_\phi M$, and the covariant formulation of the Euler–Lagrange equation is

$$D^\alpha \partial_\alpha \phi = 0,$$

where D^α represents the pullback of the Levi-Civita connection. The conserved energy is

$$E(\phi) = \frac{1}{2} \int |\partial_t \phi|_g^2 + |\nabla_x \phi|_g^2 \mathrm{d}x.$$

This problem has dimensionless scaling $\phi(t, x) \to \phi(\lambda t, \lambda x)$ and the critical Sobolev exponent is thus $s_c = \frac{n}{2}$, which yields the energy critical dimension $n = 2$.

This is where we see the first nonlinear example of a gauge invariance, namely the one associated to the choice of coordinates on M. One might be tempted to work with the Euler–Lagrange equation in local coordinates, which has the form

$$\Box \phi + \Gamma(\phi) \partial^\alpha \phi \partial_\alpha \phi = 0$$

However, at critical regularity the embedding $\dot{H}^{\frac{n}{2}} \subset L^\infty$ fails, so one cannot work in local coordinates.

Instead it turns out to be more convenient to work with an extrinsic formulation, assuming that $(M, g) \subset (\mathbb{R}^m, e)$ is an isometric embedding, as in Nash's theorem. Then the equations take the form

$$\Box \phi + S(\phi) \partial^\alpha \phi \partial_\alpha \phi = 0 \tag{WM}$$

where S is now the second fundamental form associated to the embedding.

Another way to write the Wave Map equations is as a system for the vectors $\nabla_{t,x} \phi$ expressed in a suitable orthonormal frame in $T_\phi M$. Then the gauge group becomes the group of orthogonal transformations.

We also remark that $\dot{H}^{\frac{n}{2}}$ maps from $\mathbb{R}^n$ into (M, g) can be split into topological classes according to their homotopy class, where $\mathbb{R}^n$ is viewed as its one point compactification at infinity.

The existence of ground states is connected to the existence of nontrivial topological classes. For instance in the two dimensional case we have two model problems:

- $M = \mathbb{S}^2$, where the topological classes are indexed by integers and the ground state is the stereographic projection.
- $M = \mathbb{H}^2$, where there is only the trivial topological class.

One can view these two cases as the analogues of the focusing/defocusing (NLW).

Maxwell–Klein–Gordon. Here our states have two components:

- A Maxwell field A in $\mathbb{R}^{n+1}$, with curvature $F_{\alpha\beta} = \partial_\alpha A_\beta - \partial_\beta A_\alpha$.
- A complex field ϕ in $\mathbb{R}^{n+1}$, with covariant differentiation $D_\alpha = \partial_\alpha + i A_\alpha$.

The Lagrangian is obtained simply by adding the Maxwell Lagrangian with the covariant wave Lagrangian,

$$\mathcal{L}(A, \phi) = \int_{\mathbb{R}^{n+1}} \frac{1}{2} D^\alpha \phi \overline{D_\alpha \phi} + \frac{1}{4} F^{\alpha\beta} F_{\alpha\beta} \, \mathrm{d}x\mathrm{d}t.$$

The Euler–Lagrange equations are a system for (A, ϕ), namely

$$\begin{cases} \partial^\beta F_{\alpha\beta} = \Im(\phi \overline{D_\alpha \phi}) \\ D^\alpha D_\alpha \phi = 0. \end{cases}$$

This has the conserved energy:

$$E(A,\phi) = \int_{\mathbb{R}^n} \frac{1}{4}|F|^2 + \frac{1}{2}|D_A\phi|^2 \,\mathrm{d}x.$$

The critical Sobolev exponent is now $s_c = \frac{n}{2} - 1$, which yields the energy critical dimension $n = 4$. For this system the gauge invariance has the form

$$(A,\phi) \to (A - db, \phi e^{ib}).$$

Finally, we remark that the Maxwell–Klein–Gordon system has no ground state.

Hyperbolic Yang–Mills flow. This can be viewed as a noncommutative version of the Maxwell equations. We start with a compact noncommutative Lie group $\mathfrak{G}$, its associated Lie algebra $\mathfrak{g}$, and a bi-invariant positive definite inner product $\langle \cdot, \cdot \rangle$ on $\mathfrak{g}$. The objects of study are connection 1-forms

$$A_\alpha : \mathbb{R}^{n+1} \to \mathfrak{g},$$

which define a covariant differentiation on $\mathfrak{g}$ valued maps,

$$D_\alpha B := \partial_\alpha B + [A_\alpha, B].$$

Commuting these we obtain the curvature tensor

$$F_{\alpha\beta} := \partial_\alpha A_\beta - \partial_\beta A_\alpha + [A_\alpha, A_\beta].$$

The Lagrangian is similar to the Maxwell Lagrangian,

$$\mathcal{L}(A) := \frac{1}{2}\int_{\mathbb{R}^{4+1}} \langle F_{\alpha\beta}, F^{\alpha\beta}\rangle \,\mathrm{d}x\mathrm{d}t.$$

In covariant form the Euler–Lagrange equations are also similar to Maxwell,

$$D^\alpha F_{\alpha\beta} = 0 \tag{YM}$$

The conseved energy is now

$$E(A) = \frac{1}{4}\int_{\mathbb{R}^4} |F|^2 \mathrm{d}x$$

As for Maxwell–Klein–Gordon, the critical Sobolev exponent is now $s_c = \frac{n}{2} - 1$ which yields the energy critical dimension $n = 4$.

The gauge invariance is more complicated for Yang–Mills,

$$A_\alpha \to OAO^{-1} - \partial_\alpha O O^{-1}, \qquad O : \mathbb{R}^{n+1} \to \mathfrak{G},$$

which makes the choice of a good gauge more complicated.

Finite energy Yang–Mills connections are not necessarily trivial at infinity, so the topological classes are related to the homotopy classes for maps $O : \mathbb{S}^3 \to \mathfrak{G}$. The model case here is $\mathfrak{G} = SU(2)$, where the topological classes are indexed by integers, precisely the second Chern numbers c_2. The energy minimizers in each topological class are called instantons, and must be either self-dual or anti-self dual.

2. Small data results

Before discussing the large data problems, it is very useful to start with the small data problem, which already brings out some key difficulties. We begin by stating the main results for the three problems:

Theorem 2.1 (Tao [28] ($\mathbb{S}^n$), Krieger [7] ($\mathbb{H}^2$), Tataru [32] ((M,g) compact))**.** (WM) *is globally well-posed for small data in* $\dot{H}^{\frac{n}{2}} \times \dot{H}^{\frac{n}{2}-1}$, $n \geq 2$.

Theorem 2.2 (Rodnianski–Tao [25] ($n \geq 6$), Krieger–Sterbenz–Tataru [13]($n \geq 4$))**.** (MKG) *is globally well-posed for small data in* $\dot{H}^{\frac{n}{2}-1} \times \dot{H}^{\frac{n}{2}-2}$, $n \geq 4$.

Theorem 2.3 (Krieger–Sterbenz [12] ($n \geq 6$), Krieger–Tataru [14] ($n \geq 4$))**.** (YM) *is globally well-posed for small data in* $\dot{H}^{\frac{n}{2}-1} \times \dot{H}^{\frac{n}{2}-2}$, $n \geq 4$.

Two new features for all of these results, unlike for more regular solutions, are as follows:

- quasilinear well-posedness (continuous dependence on data)
- modified scattering

To provide a brief overview of the ideas involved in the proofs, let us consider first a perturbative argument to solve an equation such as (1.1). What one needs are two function spaces, call them S and N, so that the following two properties hold:

1. Linear mapping property: The solution map for
$$\Box u = f, \qquad u(0) = u_0, \quad u_t(0) = u_1$$
satisfies
$$\|u\|_S \lesssim \|f\|_N + \|v\|_{\dot{H}^1 \times L^2}.$$
2. Nonlinear mapping property: The map
$$S \ni u \to N(u) \in N \qquad \text{is Lipschitz continuous.}$$

Then the solution is obtained using the contraction principle, where the small Lipschitz constant comes from the smallness of the initial data.

The above strategy works easily for (NLW), where S and N are simply Strichartz type spaces. Unfortunately, regardless of how well the function spaces are chosen, such properties cannot hold for our three geometric nonlinear wave equations.

The next best thing, then, is to still choose the S and N spaces in the most favourable way so that the bulk of the nonlinearity can be estimated perturbatively, while the rest exhibits only a mild (logarithmic) failure. This leads one to the paradifferential formulation of the equations, which are viewed now as an infinite system for the Littlewood–Paley dyadic pieces of u, namely (here u_k is localized at frequency 2^k)
$$\Box u_k = DN(u_{<k})u_k + R_k(u).$$
Here DN stands for the differential of N, which is applied to the lower frequency part of u. Then the idea is to replace the above perturbative argument with a nonperturbative argument as follows:

1. Linear mapping property: The solution map for
$$(\Box - DN(u_{<k}))v_k = f_k, \qquad v_k(0) = v_0, \quad v_t(0) = v_1$$
satisfies
$$\|v\|_S \lesssim \|f\|_N + \|(v_0, v_1)\|_{\dot{H}^1 \times L^2}$$
2. Nonlinear mapping property:
$$S \ni u \to R(u) \in N \qquad \text{is Lipschitz continuous}$$

In principle this is the strategy that works, but none of the steps are straightforward. One could identify three main hurdles:

- The gauge choice, which is made to insure that (i) the equations are indeed nonlinear wave equations of the form (1.1), and (ii) that the quadratic nonlinearities satisfy the so-called null condition.
- The function spaces S and N, whose choice needs to insure not only the bound for the perturbative part R, but also to be enough for the next step.
- Renormalization: The key to proving the linear paradifferential mapping property is to show that the operators $\Box$ and $\Box - DN(u_{<k})$ are conjugate up to perturbative errors.

The gauge choice is perhaps easiest to address. This is most easily done by using the extrinsic setting for (WM), respectively the Coulomb gauge $\partial^j A_j = 0$ for (MKG) and (YM). The latter choice has a price to pay, namely that the temporal components A_0 and $\partial_t A_0$ become dependent variables solving some nonlinear elliptic equations, but this is only a technical issue in the small data case.

The two remaining matters are more delicate, and are discussed in what follows.

2.1. Function spaces

The difficulty here is two-fold. First of all, one needs to best capture the dispersive properties of the free waves, both at the linear and at the bilinear level (for wave interactions). Secondly, these properties need to carry over to the solutions to our nonlinear flows, which are certainly not free waves. A brief enumeration of the succession of ideas is as follows:

- *Strichartz norms.* These are mixed $L^p_t L^q_x$ norms. They are scale invariant, however do not capture any of the null structure of the equations.

- *Bourgain $X^{s,b}$ spaces.* They were first introduced by Klainerman–Machedon [5] in the wave equation context. These do capture null structure but do not have good scaling properties.

- *Null frame spaces.* These were first introduced in Tataru [31], and then modified by Tao [28]) These combine Strichartz norms with multiscale frequency localizations adapted to the null cone. In addition, certain norms are measured in suitable null frames.

- *U^2 and V^2 spaces.* They were first introduced in an unpublished initial version of [31], and then developed in Koch–Tataru [6]. One can view then as scale invariant refinements of $X^{0,\frac{1}{2}}$ spaces.

The proofs of all the above theorems use the null frame spaces at least in the more difficult low dimensional setting. The U^2 spaces were also good initial candidates, but at the time there was one stumbling block in the proof of certain bilinear estimates. Very recently that stumbling block seems to have been removed by Candy–Herr [2], so there is renewed hope that these spaces can be alternatively used for the nonlinear wave problems.

2.2. Renormalization

Here the goal is to have a linear bound of the form

$$\|u_k\|_S \lesssim \|u_k[0]\|_E + \|\Box u_k + DN(u_{<k})u_k\|_N$$

for the truncated paradifferential evolution. The *renormalization* method, first introduced by Tao [28] for (WM), is to approximately conjugate the paradifferential flow to the flat wave flow.

$$R(A_k + 2[A^\alpha_{<k}, \partial_\alpha A_k]) - \Box R A_k = err, \qquad \text{(perturbative)}$$

with good mapping and invertibility properties

$$R : S \to S, \qquad R : N \to N$$

This has been developed in several stages as follows:

1. $(\mathrm{WM})_{\geq 2}$. This is the setting where renormalization was first introduced by Tao [28], and further refined by Tataru [31]. Here the renormalization is multiplicative,

$$R = R(t,x) : \mathbb{R}^{n+1} \to SO(d)$$

2. $(\mathrm{MKG})_{\geq 6}$ (Rodnianski–Tao [25]) and $(\mathrm{YM})_{\geq 6}$ (Krieger–Sterbenz [12]). Here $R = R(t,x,D)$ is a pseudodifferential operator with rough symbol, but one only needs $L^p \to L^q$ bounds for err.

3. $(\mathrm{MKG})_{\geq 4}$ (Krieger–Sterbenz–Tataru [13]) and $(\mathrm{YM})_{\geq 4}$ (Krieger–Tataru [14]). Here one additional difficulty is that one needs to use null frame spaces S, N and error bounds in these spaces. In addition, in the (YM) setting the standard pdo calculus is replaced with a noncommutative Lie algebra based counterpart.

3. The large data energy critical problem

We begin with (WM), which is energy critical in dimension $n = 2$. There the main results are as follows:

Theorem 3.1 (Threshold Theorem for (WM), Sterbenz–Tataru [26, 27], (M, g) compact)**.** (WM) *in 2+1-d is globally well-posed for data below the ground state energy.*

Independently Krieger–Schlag [9] considered the case $M = \mathbb{H}^2$, and Tao [30] the case $M = \mathbb{H}^n$. Then there is no ground state, and indeed their result asserts that the problem is globally well-posed.

The above theorem can be seen as a consequence of

Theorem 3.2 (Dichotomy Theorem for (WM), Sterbenz–Tataru [26, 27])**.** *For any finite energy data* (WM) *in 2+1-d we have the following dichotomy. Either*

1. *The solution is global, topologically trivial and scatters at infinity, or*
2. *The solution bubbles off a soliton either*
   (a) *in finite time (blow-up), or*
   (b) *in infinite time.*

The situation is somewhat simpler for (MKG), where no ground states exist:

Theorem 3.3 (Oh–Tataru [16, 17, 22], Krieger–Luhrmann [8])**.** (MKG) *in 4+1-d is globally well-posed for finite energy data.*

Turning our attention to (YM), the topology plays an important role in obtaining the correct threshold. Precisely, the ground state is not in the topologically trivial class, whereas the solutions which are the object to the Threshold Theorem are topologically trivial. By topologically subtracting two ground state solutions with energy $\mathcal{E}_{GS}$, one obtains a topologically trivial near steady state with energy just above $2\mathcal{E}_{GS}$. Is then $2\mathcal{E}_{GS}$ the correct threshold? Fortunately we have

Theorem 3.4 (Gursky–Kelleher-Streets [3])**.** *Any harmonic Yang–Mills connection A in 4-d which is topologically trivial has energy $\mathcal{E}(A) > 2\mathcal{E}_{GS}$.*

This leads us to the following better formulation of the result:

Theorem 3.5 (Threshold Theorem for (YM), Oh–Tataru [18, 19, 20, 21])**.** (YM) *in $4 + 1 - d$ is globally well-posed and scatters in the caloric gauge for topologically trivial data below energy $2\mathcal{E}_{GS}$.*

Correspondingly we have

Theorem 3.6 (Dichotomy Theorem for (YM), Oh–Tataru [18, 19, 20, 21])**.** *For any finite energy data for* (YM) *in 4+1-d we have the following dichotomy. Either*

1. *The solution is global, topologically trivial and scatters at infinity, or*
2. *The solution bubbles off a soliton either*
   (a) *in finite time (blow-up), or*
   (b) *in infinite time.*

We now review several approaches which have led to proofs of Threshold Theorems in several energy critical problems. The goal is to prove an estimate

$$\|A\|_S \lesssim F(E(A)), \qquad E < 2\mathcal{E}_{GS}.$$

in the (YM) context.

1. Direct induction on energy (Bourgain [1]) combines perturbative and nonperturbative elements in a single induction step.

2. Concentration compactness method (Kenig–Merle [4]). This is a two step approach, by contradiction:

   - prove the existence of a minimal energy blow-up solution, with good compactness properties.
   - disprove the existence of a minimal energy blow-up solution, by Morawetz style (non-concentration) estimates.

3. Energy dispersion method (Sterbenz–Tataru [26, 27]). This is also a two step approach, but using a direct method:

   - prove that energy dispersed solutions are global and scatter.
   - prove that all solutions are either energy dispersed or have pockets of energy convergent to a steady state.

It is this last method that has been used for (WM) in [26, 27], (MKG) in [16, 17, 22] and (YM) in [18, 19, 20, 21]. The topological matters briefly described above notwithstanding, there are three major and almost disjoint matters that had to be addressed:

1. *The gauge question.* This is where one fixes the gauge, (ii) prove it applies to all subthreshold, and (iii) write the evolution in the chosen gauge.

2. *The continuation problem.* Here we prove (i) local well-posedness for large data on an energy non-concentration time scale, and (ii) strong global bounds for *energy dispersed* solutions.

3. *Blow-up analysis.* In this final step we use Morawetz type bounds in order to perform a blow-up analysis which leads to the proof of the two theorems above.

We briefly discuss these three items in the last three sections.

4. The gauge choice in (MKG) and (YM)

We recall the gauge invariance properties for (MKG) and (YM).

$$A \to A - db \tag{MKG}$$

respectively

$$A_\alpha \to OAO^{-1} - \partial_\alpha O O^{-1}. \tag{YM}$$

For large data (MKG) the Coulomb gauge $\partial^j A_j$ still works, but additional work is required in order to establish a preliminary short time local well-posedness result. In particular this requires initial data excision and gluing techniques at critical regularity; see [17] for more details.

Hence in what follows we turn our attention to the hyperbolic (YM) flow, where matters are far more delicate but also more interesting. In fixing the gauge there are perhaps four objectives, which turn out not to be mutually compatible:

(i) to preserve hyperbolic structure

(ii) to capture null structure of equations

(iii) globally defined (large subthreshold data)

(iv) consistent with dispersive decay at linear level

There are also three traditional gauge choices:

- Lorenz gauge $\partial^\alpha A_\alpha = 0$. [satisfies (i), (iii), (iv)]
- temporal gauge $A_0 = 0$. [satisfies (i), (ii), (iii)]

- Coulomb gauge $\partial_j A_j = 0$. [satisfies (ii), (iv)]

of which the last one was used in [] for the small data problem. Unfortunately for large data this no longer works, due to competing thresholds connected to harmonic gauge transformations. To remedy this, in [21] we develop a fourth gauge choice for (YM), namely

- Caloric gauge -defined via covariant heat flow. [satisfies (ii), (iii), (iv)]

The caloric gauge is defined in [21] using an additional flow, namely the Yang–Mills heat flow. This idea originates in a different context, namely in Tao's approach [29] to wave maps into the hyperbolic space, where a caloric gauge is constructed using the harmonic heat flow. In the (YM) context, this idea was first implemented by Oh [15] at the local level, and then fully developed in Oh–Tataru [21] for the global problem.

We provide below a very brief overview of the construction. The (YM) heat flow, expressed in a gauge which is akin to the temporal gauge above, has the form

$$\partial_s A_j = D^k F_{kj}, \qquad A_j(0) = A_{j0}$$

where s denotes the parabolic time. This flow has the following properties, see [21]:

- Locally well-posed for $\dot{H}^1$ data.
- Globally well-posed for subthreshold data, with flat limiting connection $F(\infty) = 0$.
- C^1 but *not* C^2 dependence on the initial data.

The limiting flat connection is gauge equivalent to the zero connection. This motivates

Definition 4.1 (Caloric gauge)**.** *A state A_x is called caloric if its parabolic flow satisfies $A(s = \infty) = 0$.*

Which in turn is justified by

Theorem 4.2 ([21])**.** *Each subthreshold connection A admits a unique caloric representation $\tilde{A} = Cal(A)$, which is unique up to constant gauge conjugations.*

The family $\mathcal{C}$ of subthreshold caloric states turns out to be a manifold of class C^1 below the threshold $2\mathcal{E}_{GS}$,

$$\mathcal{C} \subset \mathbf{H} = \{A \in \dot{H}^1, \quad \partial^j A_j \in \ell^1 L^2\}$$

with a good bound in terms of the energy $E < 2\mathcal{E}_{GS}$,

$$\|A\|_{\mathbf{H}} \lesssim_E 1, \qquad A \in \mathcal{C}$$

The hyperbolic (YM) equation in the caloric gauge can be viewed as an evolution for the dynamic Yang–Mills variable:

$$(A_x, \partial_t A_x) \in T^{L^2}\mathcal{C} \subset \mathbf{H} \times L^2$$

whereas the remaining variables $(A_0, \partial_0 A_0)$ are uniquely determined in terms of the dynamic variables via the caloric flow.

Eventually, one is able to express the hyperbolic Yang–Mills equation in the caloric gauge:

$$\Box_A A_j = Q_j(A, \partial A) + R_j(A, A, A)$$

together with the auxiliary equations:

$$\Delta_A A_0 = Q_0(A, \partial A) + R_0(A, A, A)$$

$$\Delta_A(\partial_0 A_0) = R_{00}(A, A)$$

Furthermore, we have a generalized Coulomb condition for caloric states:

$$\partial^j A_j = Q(A, A) + R(A, A, A)$$

It is remarkable that, even though the gauge choice goes through a nonlinear heat flow, we can still split the nonlinearities into a quadratic part which is explicit, and a cubic and higher order part which, while having a complex structure, satisfies nice perturbative estimates.

5. Energy dispersed solutions

To measure pointwise concentration of the solutions in a time interval I, here we use the Besov type *energy dispersion* norm:

$$\|A\|_{ED[I]} = \sup_k 2^{-k} \|A_k\|_{L^\infty[I]},$$

which scales like the energy and is bounded by the energy by Bernstein's inequality.

The main result, stated below for (YM) but which applies equally to (WM) and (MKG), asserts that solutions with small energy dispersion satisfy nice global bounds:

Theorem 5.1 (Energy dispersed solutions, Sterbenz–Tataru [26] (WM), Oh–Tataru [22] (MKG) and [19] (YM))**.** *For each $E < E_0$ there exist $\epsilon(E)$, $F(E)$ so that for each solution A of energy E in a time interval I we have:*

$$\|A\|_{ED} \leq \epsilon(E) \implies \|A\|_S \leq F(E)$$

Broadly speaking, the proof uses an induction on energy, where the goal is to pass from an energy E to a larger energy $E + c$. The key point here, and also the difficulty, is to be able to choose $c = c(E)$, i.e. without any dependence on $\epsilon(E)$ or $F(E)$. Some of the main elements of the proof are as follows:

1. Improved bilinear/multilinear estimates: ED yields gains for all *balanced* frequency interactions.

2. Improved paradifferential bound:
$$\|A_k\|_S \lesssim \|A_k[0]\|_E + \|\Box A_k + 2[A^\alpha_{<k-m}, \partial_\alpha A_k]\|_N$$
with the *frequency gap* $m \gg_{\|\phi\|_S} 1$ as a proxy for smallness.

3. Divisibility estimate: For any solution ϕ of energy E and S size F we can split the time interval into $N \lesssim_F 1$ so that
$$\|\phi\|_{S[I_k]} \lesssim E$$

6. The bubble-off argument

The main result here is stated for (YM), but applies again to all three models (WM), (MKG) and (YM):

Theorem 6.1 (Sterbenz–Tataru [27] (WM), Oh–Tataru [16] (MKG) and [20] (YM))**.** *For any finite energy Yang–Mills wave which either*

(i) *blows up in finite time or*

(ii) *is global but does not scatter we can find a sequence of points (t_n, x_n) and scales*
$$A\left(\frac{x - x_n}{r_n}, \frac{t - t_n}{r_n}\right) \to LA_{steady}$$
with L Lorentz and A_{steady} steady state.

The proofs in the three cases differ in many significant ways, but follow a similar strategy which we briefly outline. The starting points are the finite speed of propagation and the energy flux relation on light cones,

$$\mathcal{E}(t_0) - \mathcal{E}(t_1) = \mathcal{F}(t_0, t_1).$$

These combined with the small data result allow us to identify two unfavourable scenarios:

- Blow up via energy concentration near the tip of a light cone

- Energy concentration inside an outgoing (opening) light cone but no scattering.

Fortunately the two scenarios are dealt with in an almost identical fashion, of which we just enumerate some of the key steps. By translation invariance in what follows we work with the forward light cone C centered at $(0,0)$.

Concentration bubbles. As energy dispersion cannot be small near the tip (or top) of the cone, by the Energy Dispersion theorem we obtain a sequence of energy bubbles centered at (t_n, x_n), on scale r_n and frequency r_n^{-1}.

No null concentration. Using Morawetz estimates with the vector field

$$X_0 = \frac{t\partial_t + x\partial_x}{\sqrt{t^2 - x^2}}$$

and its translates one can control the integrated quantity

$$\int \frac{1}{\sqrt{t^2 - x^2}} |i_{X_0} F|^2 \mathrm{d}x\mathrm{d}t$$

i.e. the scaling component of F. Using weighted (and covariant) versions of Bernstein's inequality, we conclude that the concentration bubbles must lie strictly inside the cone,

$$(x_n, t_n) \in C_\gamma = \left\{ \frac{|x_n|}{t_n} \leq \gamma \right\}, \quad \gamma < 1.$$

Energy persistence in the smaller cone. Here, the same Morawetz estimates allow us to propagate the energy bubbles at times t_n in the smaller cone C_γ, for a time that increases in relative terms as the flux decays to zero.

Final rescaling. This yields a sequence $A^{(n)}$ of rescaled maps with uniformly nontrivial energy in C_γ, but with the following decay properties:

$$\mathcal{F}_{[1,T_n]}(A^{(n)}) \to 0, \qquad \int_{C_{[1,T_n]}} \frac{1}{\sqrt{t^2 - x^2}} |i_{X_0} F^{(n)}|^2 \mathrm{d}x\mathrm{d}t \to 0$$

Concentration scales. Here a further pigeonhole arguments allows us to select cubes $Q_n = Q_{r_n}(x_n, t_n)$ (within C_γ where the energy is nontrivial but $i_X F^{(n)}$ decays to 0 in average.

Compactness. The condition $i_X F^{(n)} \to 0$ yields ellipticity in the limit, and allows extraction of a strongly convergent subsequence in $\dot{H}^1$. Now we need to consider two cases depending on the size of the concentration scales r_n.

Small energy concentration scale. $r_n \to 0$. After scaling and translation we get: $A^{(n)} \to LQ$ where Q is a global harmonic (YM) state. This is our bubbling phenomena.

Large energy concentration scale. $r^n > c > 0$. Then in the limit we get $A^{(n)} \to A_{\text{selfsimilar}}$. But as it turns out, no self-similar finite energy solutions exist.

References

[1] J. Bourgain, *Global solutions of nonlinear Schrödinger equations*, Colloquium Publications, vol. 46, American Mathematical Society, 1999.

[2] T. Candy & S. Herr, "On the Division Problem for the Wave Maps Equation", `https://arxiv.org/abs/1807.02066`, 2018.

[3] M. Gursky, C. L. Kelleher & J. Streets, "A conformally invariant gap theorem in Yang–Mills theory", `https://arxiv.org/abs/1708.01157`, 2017.

[4] C. E. Kenig & F. Merle, "Global well-posedness, scattering and blow-up for the energy-critical focusing non-linear wave equation", *Acta Math.* **201** (2008), no. 2, p. 147-212.

[5] S. Klainerman & M. Machedon, "Space-time estimates for null forms and the local existence theorem", *Commun. Pure Appl. Math.* **46** (1993), no. 9, p. 1221-1268.

[6] H. Koch & D. Tataru, "Dispersive estimates for principally normal pseudodifferential operators", *Commun. Pure Appl. Math.* **58** (2005), no. 2, p. 217-284.

[7] J. Krieger, "Global regularity of wave maps from $\mathbf{R}^{2+1}$ to H^2. Small energy", *Commun. Math. Phys.* **250** (2004), no. 3, p. 507-580.

[8] J. Krieger & J. Lührmann, "Concentration compactness for the critical Maxwell–Klein–Gordon equation", *Ann. PDE* **1** (2015), no. 1, article ID 5 (208 pages).

[9] J. Krieger & W. Schlag, *Concentration compactness for critical wave maps*, EMS Monographs in Mathematics, European Mathematical Society, 2012.

[10] J. Krieger, W. Schlag & D. Tataru, *Invent. Math.* **171** (2008), no. 3, p. 543-615.

[11] ———, "Renormalization and blow up for the critical Yang–Mills problem", *Adv. Math.* **221** (2009), no. 5, p. 1445-1521.

[12] J. Krieger & J. Sterbenz, "Global regularity for the Yang–Mills equations on high dimensional Minkowski space", *Mem. Am. Math. Soc.* **223** (2013), no. 1047.

[13] J. Krieger, J. Sterbenz & D. Tataru, "Global well-posedness for the Maxwell–Klein–Gordon equation in $4+1$ dimensions: small energy", *Duke Math. J.* **164** (2015), no. 6, p. 973-1040.

[14] J. Krieger & D. Tataru, "Global well-posedness for the Yang–Mills equation in $4+1$ dimensions. Small energy", *Ann. Math.* **185** (2017), no. 3, p. 831-893.

[15] S.-J. Oh, "Finite energy global well-posedness of the Yang–Mills equations on $\mathbb{R}^{1+3}$: an approach using the Yang–Mills heat flow", *Duke Math. J.* **164** (2015), no. 9, p. 1669-1732.

[16] S.-J. Oh & D. Tataru, "Global well-posedness and scattering of the $(4+1)$-dimensional Maxwell–Klein–Gordon equation", *Invent. Math.* **205** (2016), no. 3, p. 781-877.

[17] ———, "Local well-posedness of the $(4+1)$-dimensional Maxwell–Klein–Gordon equation at energy regularity", *Ann. PDE* **2** (2016), no. 1, article ID 2 (70 pages).

[18] ———, "The hyperbolic Yang–Mills equation for connections in an arbitrary topological class", `https://arxiv.org/abs/1709.08604`, 2017.

[19] ———, "The hyperbolic Yang–Mills equation in the caloric gauge. Local well-posedness and control of energy dispersed solutions", `https://arxiv.org/abs/1709.09332`, 2017.

[20] ———, "The threshold conjecture for the energy critical hyperbolic Yang–Mills equation", `https://arxiv.org/abs/1709.08606`, 2017.

[21] ———, "The Yang–Mills heat flow and the caloric gauge", `https://arxiv.org/abs/1709.08599`, 2017.

[22] ———, "Energy dispersed solutions for the (4+1)-dimensional Maxwell–Klein–Gordon equation", *Am. J. Math.* **140** (2018), no. 1, p. 1-82.

[23] P. Raphaël & I. Rodnianski, "Stable blow up dynamics for the critical co-rotational wave maps and equivariant Yang–Mills problems", *Publ. Math., Inst. Hautes Étud. Sci.*

[24] I. Rodnianski & J. Sterbenz, "On the formation of singularities in the critical O(3) σ-model", *Ann. Math.* **172** (2010), no. 1, p. 187-242.

[25] I. Rodnianski & T. Tao, "Global regularity for the Maxwell–Klein–Gordon equation with small critical Sobolev norm in high dimensions", *Commun. Math. Phys.* **251** (2004), no. 2, p. 377-426.

[26] J. Sterbenz & D. Tataru, "Energy dispersed large data wave maps in $2+1$ dimensions", *Commun. Math. Phys.* **298** (2010), no. 1, p. 139-230.

[27] ———, "Regularity of wave-maps in dimension $2+1$", *Commun. Math. Phys.* **298** (2010), no. 1, p. 231-264.

[28] T. Tao, "Global regularity of wave maps. II. Small energy in two dimensions", *Commun. Math. Phys.* **224** (2001), no. 2, p. 443-544.

[29] ———, "Geometric renormalization of large energy wave maps", *Journ. Équ. Dériv. Partielles* **2004** (2004), article ID XI (32 pages).

[30] ———, "Global regularity of wave maps VII. Control of delocalised or dispersed solutions", `https://arxiv.org/abs/0908.0776`, 2009.

[31] D. Tataru, "On global existence and scattering for the wave maps equation", *Am. J. Math.* **123** (2001), no. 1, p. 37-77.

[32] ———, "Rough solutions for the wave maps equation", *Am. J. Math.* **127** (2005), no. 2, p. 293-377.

Daniel Tataru
University of California
Berkeley
USA
`tataru@math.berkeley.edu`