

# NONLINEAR WAVE EQUATIONS - CLASSICAL METHODS

ARICK SHAO

The wave equation is one of the most fundamental partial differential equations, and it can be found everywhere in the real world. One of the simplest and earliest examples, in  $(1+1)$ -dimensions, is that of modeling a vibrating string. Many additional examples exist in physics, such as in the study of sound and light waves. In fact, the wave equation is ubiquitous in a significant class of dynamical situations. For example, solutions of the Maxwell equations in electromagnetics and the Einstein equations in general relativity exhibit wavelike phenomena; indeed, the wave equation can be found within both of these very fundamental equations of physics.

In these notes, we review some classical methods for treating *nonlinear wave equations* (NLW). More specifically, we accomplish the following:

- (1) First, we list some basic results from the study of linear wave equations.
- (2) Next, we discuss the classical local well-posedness theory for a specific subclass of NLW: the quadratic derivative NLW (abbreviated *dNLW*).
- (3) We then advance into the global well-posedness theory for quadratic dNLW. We first tackle the problem for higher dimensions via the *vector field method*. Then, we discuss the  $(3+1)$ -dimensional case using the *null condition*.

By “classical methods”, we generally mean that we will deal only with functions for which all our differential operations will be defined in the classical sense. For convenience, we will assume that all the functions we will encounter are smooth. In particular, we will avoid the study of low regularity solutions, which has been an active area of research that has yielded a vast amount of literature. Moreover, there are many important classical topics that we will not have the space to discuss. One notable omission is the method of *normal forms*.

**Remark.** *These notes are based on a series of introductory level talks given at the University of Toronto in fall 2012 and fall 2013. Many of the topics discussed here are abridged versions of more detailed discussions found in [10].*

## 1. LINEAR THEORY

Before discussing nonlinear wave equations, which is the focus of these notes, we first review some preliminaries for the *linear* wave equation. Throughout, our background setting is the Minkowski spacetime  $\mathbb{R}^{n+1}$ . Moreover:

- The first (time) component in  $\mathbb{R}^{n+1}$  is denoted by  $t = x^0$ .
- The remaining (spatial) components will be denoted by  $x = (x^1, \dots, x^n)$ .

On  $\mathbb{R}^{n+1}$ , the *wave operator*, or the *d'Alembertian*, is defined

$$(1) \quad \square = -\partial_t^2 + \Delta = -\partial_t^2 + \sum_{i=1}^n \partial_{x^i}^2.$$

Our objective is to solve the *initial value problem* for the wave equation,

$$(2) \quad \square u = F, \quad u|_{t=0} = f, \quad \partial_t u|_{t=0} = g,$$

where:

- $u : \mathbb{R}^{1+n} \rightarrow \mathbb{R}$  is the *unknown* we wish to solve for.
- $f, g : \mathbb{R}^n \rightarrow \mathbb{R}$  represent the *initial data*, at  $t = 0$ .
- $F : \mathbb{R}^{1+n} \rightarrow \mathbb{R}$  represents the *forcing term*.

More specifically, the problem is to solve (uniquely) for  $u$ , given  $f$ ,  $g$ , and  $F$ .

Let  $C_0^\infty(\mathbb{R}^n)$  denote the space of all real-valued, smooth, and compactly supported functions on  $\mathbb{R}^n$ . For convenience, we assume throughout that:

- $f, g \in C_0^\infty(\mathbb{R}^n)$ .
- Letting  $F(t)$  denote the restriction of  $F$  to the time  $t$ , i.e.,

$$F(t) : \mathbb{R}^n \rightarrow \mathbb{R}, \quad [F(t)](x) = F(t, x),$$

then every  $F(t) \in C_0^\infty(\mathbb{R}^n)$  for every  $t \in \mathbb{R}$ .

While much of this theory can be replicated for far less regular functions (or even for distributions), here we wish to avoid such technical considerations.

**1.1. The Homogeneous Case.** Consider first the case  $F \equiv 0$ , i.e., the *homogeneous* wave equation. There are many ways to solve this, which, broadly speaking, we can separate into *physical space* and *Fourier space* methods. While we will not delve into details here, as they can be found in most elementary textbooks in partial differential equations, we do summarize some of the main results below.

For physical space methods, the specifics are highly dependent on the dimension:

- When  $n = 1$ , we can solve for  $u$  by changing variables into *null coordinates*,

$$u = t - x, \quad v = t + x,$$

and integrating directly. The result of this effort is *d'Alembert's formula*:

$$(3) \quad u(t, x) = \frac{1}{2}[f(x-t) + g(x+t)] + \frac{1}{2} \int_{x-t}^{x+t} g(\xi) d\xi.$$

- When  $n = 3, 5, \dots$ , this can be done through the *method of spherical means*, in which one considers the averages of  $u$  on the spheres,

$$S(t_0, r_0) = \{(t, x) \in \mathbb{R}^{n+1} \mid t = t_0, |x| = r_0\}.$$

One then derives a  $(1+1)$ -dimensional wave-type equation for these averages. In the case  $n = 3$ , solving these equations yields *Kirchhoff's formula*:

$$(4) \quad u(t, x) = \frac{1}{4\pi t^2} \int_{\partial B_t(x)} [f(y) + \nabla f(y) \cdot (y - x) + tg(y)] d\sigma(y).$$

Here,  $\nabla f$  denotes the spatial gradient of  $f$ .

- When  $n = 2, 4, \dots$ , this is accomplished via the *method of descent*. In this process, one adds an extra “dummy” dimension (with  $f$ ,  $g$ , and  $F$  all independent of this new dimension). One can then solve this new  $(n+1)$ -dimensional problem using the method of spherical means.

For more details on such formulas, see, e.g., [10, Ch. 1] or [1].

By using such physical space representation formulas, such as (3) and (4), we can obtain the fundamental *finite speed of propagation* property:

**Theorem 1** (Finite speed of propagation). *Suppose  $u$  solves (2), with  $F \equiv 0$ . If  $f$  and  $g$  vanish on the ball  $\{x \in \mathbb{R}^n \mid |x| \leq R\}$ , then  $u$  vanishes on the set*

$$C = \{(t, x) \in \mathbb{R} \times \mathbb{R}^n \mid |x| \leq R - |t|\}.$$

*Proof.* See [10, Ch. 1] or [1].  $\square$

It is often more convenient to use Fourier space methods. By taking a Fourier transform of  $\square u = 0$  in the spatial variables, we obtain

$$-\partial_t^2 \hat{u}(t, \xi) - |\xi|^2 \hat{u}(t, \xi) \equiv 0, \quad \hat{u}|_{t=0} = \hat{f}, \quad \partial_t \hat{u}|_{t=0} = \hat{g}.$$

This yields for every  $\xi \in \mathbb{R}^n$  an ODE in  $t$ , which we can solve explicitly:

$$(5) \quad \hat{u}(t, \xi) = \cos(t|\xi|)\hat{f}(\xi) + \frac{\sin(t|\xi|)}{|\xi|}\hat{g}(\xi).$$

Taking an inverse Fourier transform and using some fancy notation, we can write

$$(6) \quad u(t, x) = \cos(t\sqrt{-\Delta})f + \frac{\sin(t\sqrt{-\Delta})}{\sqrt{-\Delta}}g.$$

Note that in contrast to the physical space solutions, the Fourier-based formula (5) has the added advantage of being independent of dimension.

**Remark.** *The physical and Fourier space formulas highlight rather different aspects of waves. For instance, via Plancherel's theorem, it is easy to obtain  $L^2$ -estimates for  $u$  from (5); on the other hand, such estimates are not apparent from (4). Other properties, such as decay estimates or finite speed of propagation, are readily seen using formulas such as (3) and (4) but are far more difficult to obtain from (5).*

**1.2. The General Case.** Next, we move on to the inhomogeneous equations, that is, for general  $F$ . The basic idea is decompose  $u$  as  $u_1 + u_2$ , where:

- $u_1$  solves the *homogeneous* problem,

$$(7) \quad \square u_1 \equiv 0, \quad u_1|_{t=0} = f, \quad \partial_t u_1|_{t=0} = g.$$

- $u_2$  solves the inhomogeneous problem, but with zero data:

$$(8) \quad \square u_2 = F, \quad u_2|_{t=0} \equiv 0, \quad \partial_t u_2|_{t=0} \equiv 0.$$

$u_1$  can be solved using the formulas in the preceding discussion. For  $u_2$ , one applies *Duhamel's principle*, in which one represents  $u_2$  as an infinite superposition of solutions to the *homogeneous* problem, with initial data given by the restriction of  $F$  to various timeslices. To be more specific, for each fixed time  $s$ , we solve

$$\square u_2^s \equiv 0, \quad u_2^s|_{t=s} \equiv 0, \quad \partial_t u_2^s|_{t=s} = F(s).$$

Then, one can show (see, e.g., [1, 12]) that  $u_2$  is precisely

$$(9) \quad u_2(t, x) = \int_0^t u_2^s(t-s, x) ds.$$

**Remark.** *From the point of view of a second-order equation, as written here, the initial conditions imposed for  $u_2$  in (8) may seem a bit mysterious. However, by letting  $v = \partial_t u$  and writing (2) as a first-order system,*

$$\partial_t \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 & I \\ \Delta & 0 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} 0 \\ F \end{pmatrix},$$

*this becomes more clear. In particular, the pair  $(0, F)$  is precisely the forcing term for the wave equation when it is written as a first-order system.*

Using the Fourier space approach for the homogeneous problem, i.e., the formula (5), Duhamel's principle yields the following representation formula for  $u$ :

$$(10) \quad \hat{u}(t, \xi) = \cos(t|\xi|)\hat{f}(\xi) + \frac{\sin(t|\xi|)}{|\xi|}\hat{g}(\xi) + \int_0^t \frac{\sin[(t-s)|\xi|]}{|\xi|}\hat{F}(s, \xi).$$

Analogous physical space formulas can be derived, for instance, from (3) and (4). For example, in the case  $n = 3$ , we have the identity

$$(11) \quad u(t, x) = \frac{1}{4\pi t^2} \int_{\partial B_t(x)} [f(y) + \nabla f(y) \cdot (y - x) + tg(y)] d\sigma(y) \\ + \frac{1}{4\pi} \int_0^t \int_{\partial B_s(x)} \frac{F(t-s, y)}{s} d\sigma(y) ds.$$

**1.2.1. Energy Estimates.** One fundamental feature of wave equations is energy estimates, which control  $L^2$ -type norms for  $u$  in terms of analogous  $L^2$ -type norms of the initial data. In particular, with respect to  $L^2$ -type norms, solutions of wave equations at any time preserve the regularity of the initial data.

Recall that for any integer  $s \geq 0$ , one can characterize the standard  $L^2$ -based homogeneous and inhomogeneous Sobolev norms, with  $s$  derivatives, by

$$(12) \quad \|\phi\|_{\dot{H}^s} = \|\partial^s \phi\|_{L^2}, \quad \|\phi\|_{H^s} \simeq \sum_{k=0}^s \|\partial^k \phi\|_{L^2}, \quad \phi \in C_0^\infty(\mathbb{R}^n),$$

where  $\partial\phi$  refers to the gradient of  $\phi$  in  $\mathbb{R}^n$ . Moreover, (12) can be extended to *fractional derivatives*, that is, to *real*  $s$ . This is most easily done using a bit of Fourier magic, courtesy of Plancherel's theorem: for  $s \geq 0$ , we define

$$(13) \quad \|\phi\|_{\dot{H}^s} = \int_{\mathbb{R}^n} |\xi|^{2s} |\hat{\phi}(\xi)|^2 d\xi, \quad \|\phi\|_{H^s}^2 = \int_{\mathbb{R}^n} (1 + |\xi|^2)^s |\hat{\phi}(\xi)|^2 d\xi.$$

For more details, the reader is referred to [10, Ch. 4] or to [1, 12].

The formulas (13) are quite compatible with our Fourier representation formula (10). For instance, multiplying (10) by  $|\xi|^s$  and integrating, we obtain:

**Theorem 2** (Energy inequality). *If  $u$  solves (2), then for any  $s \geq 1$  and  $t \geq 0$ ,*

$$(14) \quad \|u(t)\|_{\dot{H}^s} + \|\partial_t u(t)\|_{\dot{H}^{s-1}} \lesssim \|f\|_{\dot{H}^s} + \|g\|_{\dot{H}^{s-1}} + \int_0^t \|F(s)\|_{\dot{H}^{s-1}} ds.$$

*In particular, when  $s = 1$ , we have*

$$(15) \quad \|\partial_x u(t)\|_{L^2} + \|\partial_t u(t)\|_{L^2} \lesssim \|\partial_x f\|_{L^2} + \|g\|_{L^2} + \int_0^t \|F(s)\|_{L^2} ds.$$

We can also derive an inhomogeneous version of the above. Multiplying (10) by the weight  $(1 + |\xi|^2)^{s/2}$  and integrating again (a bit of extra care is required due to the factors  $|\xi|^{-1}$ ) yields the following estimate:

**Theorem 3** (Energy inequality). *If  $u$  solves (2), then for any  $s \geq 1$  and  $t \geq 0$ ,*

$$(16) \quad \|u(t)\|_{H^s} + \|\partial_t u(t)\|_{H^{s-1}} \lesssim (1 + |t|) \left[ \|f\|_{H^s} + \|g\|_{H^{s-1}} + \int_0^t \|F(s)\|_{H^{s-1}} ds \right].$$

## 2. LOCAL SOLUTIONS

We now advance to our model *nonlinear* initial value problem,

$$(17) \quad \square u = (\partial u)^2, \quad u|_{t=0} = f \in C_0^\infty(\mathbb{R}^n), \quad \partial_t u|_{t=0} = g \in C_0^\infty(\mathbb{R}^n),$$

which we refer to as the *quadratic nonlinear wave equation*, or the *quadratic dNLW*. The schematic notation  $(\partial u)^2$  in (17) refers to any sum of terms quadratic in the *spacetime* gradient of  $u$ . Examples of this include the following:

$$(18) \quad (\partial u)^2 = (\partial_t u)^2, \quad (\partial u)^2 = -(\partial_t u)^2 + \sum_{i=1}^n (\partial_i u)^2.$$

As we shall see, in the case  $n = 3$ , the two nonlinearities in (18) have radically different properties with regards to the existence of global solutions.

In this section, we affirmatively answer the following questions:

- *Does the problem (17) have unique solutions that exist locally in time?*

The precise result, which is our main topic of discussion in this section, is below:

**Theorem 4** (Local existence and uniqueness). *Let  $s > n/2 + 1$ , and suppose*

$$(19) \quad R = \|f\|_{H^s} + \|g\|_{H^{s-1}} < \infty.$$

*Then, there exists  $T > 0$ , depending on  $n$  and  $R$ , and a unique solution*

$$(20) \quad u \in C^0([0, T]; H^s(\mathbb{R}^n)) \cap C^1([0, T]; H^{s-1}(\mathbb{R}^n))$$

*to the initial value problem (17).*

Much of this discussion will be based on the contents of [10, Ch. 5].

**2.1. Existence and Uniqueness.** Before proving Theorem 4, we first discuss some of the main ideas. We begin by addressing the most basic question:

- *What exactly do we mean by a local solution to (17)?*

Consider a system of ordinary differential equations,

$$(21) \quad \partial_t \vec{v} = \vec{F}(t, \vec{v}), \quad \vec{F} : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n,$$

for which the unknown can be viewed as a curve  $\vec{v} : I \rightarrow \mathbb{R}^n$  lying in  $n$ -dimensional space, with  $I$  an interval in  $\mathbb{R}$ . We can adopt a similar perspective for our partial differential equation (17). Indeed, for each fixed time  $t$ , the restriction  $u(t)$  to this time is a function on  $\mathbb{R}^n$ . Therefore, we can then think of a solution  $u$  to (17) as a curve lying in some *infinite-dimensional* space  $Y$  of functions on  $\mathbb{R}^n$ .

The next task, then, is identify this space  $Y$ . While there are several viable candidates, here we choose the simplest. Recall that in the linear setting, the energy inequality of Theorem 3 implies that solutions to the wave equation preserve  $H^s$ -regularity over all times. One could then hope that a similar phenomenon holds for the nonlinear problem (17), at least for certain  $s$ . This is in fact true for sufficiently large  $s$ , hence we take  $Y = H^s$ . In other words, *we view a solution of (17) as a curve  $u : I \rightarrow H^s(\mathbb{R}^n)$  lying in the Banach space  $H^s(\mathbb{R}^n)$* . Furthermore, by a *local solution*, we mean such a curve  $u$  for which  $I$  is a finite interval. In particular, this view of  $u$  as a curve explains the characterization of  $u$  used in (20).

2.1.1. *The Iteration Scheme.* Though we have decided what we mean by a “solution”, we still have not described how we will obtain this solution. Of course, for general nonlinearities of the form  $(\partial u)^2$ , we will not be able to obtain explicit solutions to (17). Instead, we approach this problem as a *fixed point problem*.

To be more specific, given a “sufficiently nice” function  $v : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}$ , we define  $w = \Phi(v)$  to be the unique solution of the initial value problem,

$$(22) \quad \square w = (\partial v)^2, \quad w|_{t=0} = f, \quad \partial_t w|_{t=0} = g.$$

Note that this is just the initial value problem for an inhomogeneous *linear* wave equation, so that its solution is given explicitly by (10), with  $F$  replaced by  $(\partial v)^2$ . Moreover, note that  $v$  is a *fixed point* of  $\Phi$ , that is,  $\Phi(v) = v$ , if and only if  $v$  is a *solution* of (17). Thus, the goal is now to find such a fixed point of  $\Phi$ .

Now, the mathematical tool we use to rigorously carry out this program is also the same tool that is used to establish existence and uniqueness of solutions to ordinary differential equations: the *contraction mapping theorem*, which is also known as the *Banach fixed point theorem*.

**Theorem 5** (Contraction mapping theorem). *Let  $(X, d)$  be a complete metric space, and suppose  $\Phi : X \rightarrow X$  is a contraction: there is some  $0 < c < 1$  such that*

$$d(\Phi(x), \Phi(y)) \leq c \cdot d(x, y), \quad x, y \in X.$$

*Then,  $\Phi$  has a unique fixed point, i.e., there is a unique  $x \in X$  with  $\Phi(x) = x$ .*

In light of the above contraction mapping theorem, our task is to show that the map  $\Phi$  defined by (22) is indeed a contraction, with respect to an appropriate complete metric space. In particular, note that our choice of space  $Y = H^s(\mathbb{R}^n)$  for our solutions is a Banach space, which will be important, as it will ensure the completeness requirement in the contraction mapping theorem.

Regarding the contraction mapping theorem, its proof (which can be found in many real analysis textbooks, e.g., [9]) is far more important than the result itself. A very brief outline of this argument proceeds as follows:

- Begin with some arbitrary  $x_0 \in X$ .
- Given  $x_k \in X$ , define  $x_{k+1} = \Phi(x_k)$ .
- This sequence  $(x_k)$  is in fact Cauchy, hence it has a limit  $x$ . One can then show that this  $x$  is the unique fixed point of  $\Phi$ .

In particular, this outline can be thought of as a “loop in a computer program”. In each iteration of this loop, in which we obtain  $x_{k+1}$  from  $x_k$ , we obtain a closer approximation to the fixed point than before. If we “run this loop infinitely”, then at the very end, we would expect to arrive at the actual fixed point.

This “computer program” paradigm provides additional intuition for the aforementioned outline for solving (17). Indeed, by running through the proof of the contraction mapping theorem rather than applying the theorem itself, our fixed point problem can be reinterpreted as the following process:

- Begin with some arbitrary function  $u_0 : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}$ .
- Given  $u_k : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}$ , we find  $u_{k+1} : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}$  solving

$$\square u_{k+1} = (\partial u_k)^2, \quad u_{k+1}|_{t=0} = f, \quad \partial_t u_{k+1}|_{t=0} = g.$$

- Show that, for appropriate norms and for  $T$  sufficiently small, the sequence  $(u_k)$  is Cauchy and hence has a limit  $u$ , which will solve (17).

From this viewpoint, the  $u_k$ 's are successively closer approximations to the solution of (17). By running this loop for all of eternity, then at the end of time, the  $u_k$ 's will finally settle at the actual solution of (17).

2.1.2. *Sketch of (most of) the Proof of Theorem 4.* Consider the space

$$\mathcal{X} = C([0, T]; H^s(\mathbb{R}^n)) \cap C^1([0, T]; H^{s-1}(\mathbb{R}^n))$$

of all curves in  $H^s(\mathbb{R}^n)$  which are also  $C^1$ -curves in the larger space  $H^{s-1}(\mathbb{R}^n)$ . Moreover, we can impose a Banach space structure on  $\mathcal{X}$  by defining the norm:

$$\|u\|_{\mathcal{X}} = \sup_{0 \leq t \leq T} [\|u(t)\|_{H^s} + \|\partial_t u(t)\|_{H^{s-1}}].$$

Letting  $A > 0$  (which will be chosen later), then the closed ball

$$X = \{v \in \mathcal{X} \mid \|v\|_{\mathcal{X}} \leq A\}$$

forms a closed metric space (with the metric induced by the  $\mathcal{X}$ -norm).

We wish to apply the contraction mapping theorem to the above space  $X$ , with the map  $\Phi$  defined by (22). To do this, though, we must show the following:

- (1)  $\Phi$  maps the ball  $X$  into itself.
- (2)  $\Phi$  is a contraction: there is some  $0 < c < 1$  such that

$$(23) \quad \|\Phi(v_2) - \Phi(v_1)\|_{\mathcal{X}} \leq c\|v_2 - v_1\|_{\mathcal{X}}, \quad v_1, v_2 \in X.$$

Once these two steps are completed, the contraction mapping theorem furnishes a unique fixed point for  $\Phi$ , which would (almost) complete the proof of Theorem 4.

For the first step, suppose  $v \in X$ . By the definition (22) of  $\Phi$  and the energy estimate (16), we see that  $\Phi(v)$  satisfies the estimate

$$(24) \quad \|\Phi(v)\|_{\mathcal{X}} \lesssim (1 + T) \left[ \|f\|_{H^s} + \|g\|_{H^{s-1}} + \int_0^T \|(\partial v)^2(s)\|_{H^{s-1}} ds \right].$$

Note that the norms of  $f$  and  $g$  in (24) are bounded by  $R$ , defined in (19).

To handle the nonlinear term in (24), we recall the following classical inequalities:

**Theorem 6.** *Let  $\phi, \psi \in C_0^\infty(\mathbb{R}^n)$ .*

- Sobolev estimate: *If  $\sigma > n/2$ , then*

$$(25) \quad \|\phi\|_{L^\infty} \lesssim \|\phi\|_{H^\sigma}.$$

- Product estimate: *For any  $\sigma \geq 0$ ,*

$$(26) \quad \|\phi\psi\|_{H^\sigma} \lesssim \|\phi\|_{L^\infty} \|\psi\|_{H^\sigma} + \|\phi\|_{H^\sigma} \|\psi\|_{L^\infty}.$$

*Proof.* See [10, Ch. 5, 6] or [6, 12]. □

Since  $s - 1 > n/2$  by assumption, we can apply Theorem 6 to estimate

$$\begin{aligned} \int_0^T \|(\partial v)^2(s)\|_{H^{s-1}} ds &\lesssim \int_0^T [\|v^2(s)\|_{H^s} + \|(\partial_t v)^2(s)\|_{H^{s-1}}] ds \\ &\lesssim \int_0^T [\|v(s)\|_{H^s}^2 + \|(\partial_t v)(s)\|_{H^{s-1}}^2] ds \\ &\lesssim T\|v\|_{\mathcal{X}}^2. \end{aligned}$$

Since  $v \in X$ , then by (24), there is some constant  $c > 0$  such that

$$(27) \quad \|\Phi(v)\|_{\mathcal{X}} \leq c(1 + T)(R + T\|v\|_{\mathcal{X}}^2) \leq c(1 + T)(R + TA^2).$$

Thus, by choosing  $A$  sufficiently large (with respect to  $R$ ), and by choosing  $T$  sufficiently small with respect to  $A$ , we see that

$$(28) \quad \|\Phi(v)\|_{\mathcal{X}} \leq A.$$

In other words,  $\Phi$  maps  $X$  into itself, completing the first part of the proof.

The remaining part—the contraction estimate—is proved via analogous applications of Theorem 6. Let  $v_1, v_2 \in X$ , and note that  $\Phi(v_2) - \Phi(v_1)$  satisfies

$$(29) \quad \begin{aligned} \square[\Phi(v_2) - \Phi(v_1)] &= \partial v_2(\partial v_2 - \partial v_1) + (\partial v_2 - \partial v_1)\partial v_1, \\ [\Phi(v_2) - \Phi(v_1)]|_{t=0} &= \partial_t[\Phi(v_2) - \Phi(v_1)]|_{t=0} = 0. \end{aligned}$$

Applying the energy estimate (16) and Theorem 6 to (29), we see that

$$\begin{aligned} \|\Phi(v_2) - \Phi(v_1)\|_{\mathcal{X}} &\lesssim (1+T) \int_0^T \|\partial v_2(\partial v_2 - \partial v_1)(s)\|_{H^{s-1}} ds \\ &\quad + (1+T) \int_0^T \|(\partial v_2 - \partial v_1)\partial v_1(s)\|_{H^{s-1}} ds \\ &\lesssim (1+T) \int_0^T \|\partial v_1(s)\|_{H^{s-1}} \|(\partial v_2 - \partial v_1)(s)\|_{H^{s-1}} ds \\ &\quad + (1+T) \int_0^T \|\partial v_2(s)\|_{H^{s-1}} \|(\partial v_2 - \partial v_1)(s)\|_{H^{s-1}} ds \\ &\lesssim T(1+T)(\|v_1\|_{\mathcal{X}} + \|v_2\|_{\mathcal{X}})\|v_2 - v_1\|_{\mathcal{X}}. \end{aligned}$$

Once again, since  $v_1, v_2 \in X$ , then

$$\|\Phi(v_2) - \Phi(v_1)\|_{\mathcal{X}} \lesssim AT(1+T)\|v_2 - v_1\|_{\mathcal{X}}.$$

By choosing  $T$  to be sufficiently small depending on  $A$ , the above implies

$$(30) \quad \|\Phi(v_2) - \Phi(v_1)\|_{\mathcal{X}} \leq \frac{1}{2}\|v_2 - v_1\|_{\mathcal{X}}.$$

This finishes the second step of the proof and establishes  $\Phi$  as a contraction on  $X$ . The resulting unique fixed point of  $\Phi$  provides the desired solution  $u$  to (17).

**Remark.** *Although we now have a unique fixed point that is our solution, the proof of Theorem 4 is not quite complete. There is one remaining technical issue regarding uniqueness, which we discuss in the following section.*

**2.2. Related Topics.** In this section, we address some additional important issues that are related to local well-posedness for the problem (17).

**2.2.1. Unconditional Uniqueness.** As mentioned above, there remains one technical issue that must be addressed before fully completing the proof of Theorem 4. While this fixed point  $u$  of  $\Phi$  (defined in (22) in the preceding section) is the desired solution of (17), the contraction mapping theorem we applied does *not* necessarily guarantee that this fixed point  $u$  is unique in the whole space

$$\mathcal{X} = C([0, T]; H^s(\mathbb{R}^n)) \cap C^1([0, T]; H^{s-1}(\mathbb{R}^n)).$$

Indeed, uniqueness is only established for the closed ball

$$X = \{v \in \mathcal{X} \mid \|v\| \leq A\}, \quad A \gg R = \|f\|_{H^s} + \|g\|_{H^{s-1}}.$$

This leaves open the possibility of another solution  $\tilde{u}$  to (17) which begins like  $u$  (due to uniqueness in  $X$ ), but at some point becomes large and diverges from  $u$ .



In other words, from the contraction mapping theorem, we have established *conditional uniqueness* for the solution  $u$  of (17):  *$u$  remains unique as long as its norm remains below the threshold  $A$* . To complete the proof of Theorem 4, we must prove *unconditional uniqueness*, i.e., *that  $u$  remains unique regardless of its size*. The proof of this is rather simple in our case; the main point is that this requires a separate argument apart from the contraction mapping theorem.

Suppose  $u, \tilde{u} \in \mathcal{X}$  are solutions of (17). Without loss of generality, we can assume that  $u$  is the solution obtained in the preceding section, that is, the unique solution in  $X$  obtained via the contraction mapping theorem. The main observation is that one can choose  $A$  large enough in the preceding proof so that  $\|u\|_{\mathcal{X}}$  is at most  $A/2$ . In other words,  $u$  is not on the boundary of the ball  $X$ .

The proof proceeds via a topological argument, which we outline below. Let

$$J = \{t \in [0, T] \mid u(t) = \tilde{u}(t)\}.$$

First, note that  $J$  is nonempty, since  $u$  and  $\tilde{u}$  are assumed to coincide at the initial time  $t = 0$ . Furthermore, by the continuity-in-time properties of the space  $\mathcal{X}$ , we see that  $J$  must be closed. Finally, if  $t \in J$ , then since  $\|u\|_{\mathcal{X}} \leq A/2$ , by continuity,

$$(31) \quad \|\tilde{u}(t')\|_{H^s} + \|\partial_t \tilde{u}(t')\|_{H^{s-1}} \leq A,$$

for  $t'$  sufficiently close to  $t$ . By essentially rerunning the previous fixed point argument on this small subinterval containing  $t$  for which (31) holds, we conclude that  $\tilde{u}(t')$  and  $u(t')$  coincide for  $t'$  close to  $t$ . It follows that  $J$  is also open in  $[0, T]$ . As a result,  $J = [0, T]$ , which completes the proof of unconditional uniqueness.

**Remark.** *One can also establish unconditional uniqueness using the energy estimates essentially as in the previous section; see [10, Ch. 5].*

**Remark.** *While unconditional uniqueness is only a minor technical issue in our setting, there do exist other settings in which unconditional uniqueness becomes nontrivial, or even an open problem. Furthermore, for lower regularities, i.e., values of  $s$  smaller than are considered in these notes, one may have uniqueness only in a special subclass of  $\mathcal{X}$ ; whether solutions are unique in all of  $\mathcal{X}$  may not be known.*

**2.2.2. Maximal Solutions.** Theorem 4 ensures that given appropriate initial data at time  $t = 0$ , one has a unique solution to (17) on some finite time interval  $[0, T]$ . One can then ask whether this solution can be further extended past time  $t = T$ . If so, how far can this be extended? Does there exist a global solution, extending to  $t = \infty$ ? If not, then what happens to prevent this?

It is not difficult to see that the solution  $u$  from Theorem 4 can be extended beyond time  $T$ . To obtain this, we simply take  $(u(T), \partial_t u(T))$  as initial data and apply Theorem 4 again (in fact, we can solve this both forward and backward in time). By uniqueness considerations (similar to the preceding discussion), we can then glue  $u$  together with this new local solution to extend  $u$  to the time interval  $[0, T + \delta]$ . The argument can be repeated at time  $t = T + \delta$ , and so on. By gluing all these local solutions together, we obtain a *maximal solution* of (17),

$$u : [0, T_+) \times \mathbb{R}^n \rightarrow \mathbb{R}, \quad 0 < T_+ \leq \infty,$$

in the sense that  $u$  cannot be further extended to a larger time.

**Remark.** *We can also solve (17) backwards in time to produce a maximal solution*

$$u : (-T_-, T_+) \times \mathbb{R}^n \rightarrow \mathbb{R}, \quad 0 < T_{\pm} \leq \infty.$$

One possibility is that  $T_+ = \infty$ , that is, we have a *global solution* of (17). On the other hand, it is also possible that  $u$  terminates at some finite time, i.e., that  $T_+ < \infty$ . To see how this could occur, suppose that  $\|u(t)\|_{H^s}$  grows as  $t$  increases. Since the time of existence from Theorem 4 depends on the above norm, then in each application of Theorem 4 to extend  $u$ , the new portion of the solution occupies a successively smaller time interval. Thus, it is possible that after the  $k$ -th application of Theorem 4, the solution  $u$  is defined over the interval  $[0, T + \delta_k]$ , where  $\delta_k$  is a strictly increasing sequence with a *finite* limit. In this case, the maximal solution  $u$  could be over only a finite interval  $[0, T_+)$ .

Furthermore, we can use Theorem 4 to show that the only manner that the maximal solution  $u$  can terminate for its Sobolev norms to blow up.

**Corollary 7** (Maximal solutions). *Suppose  $u : [0, T_+) \times \mathbb{R}^n \rightarrow \mathbb{R}$ , is a maximal solution to (17), with  $0 < T_+ \leq \infty$ , and suppose*

$$u \in C^0([0, T]; H^s(\mathbb{R}^n)) \cap C^1([0, T]; H^{s-1}(\mathbb{R}^n))$$

for any  $0 < T < T_+$ . If  $T_+ < \infty$ , then

$$\limsup_{t \nearrow T_+} [\|u(t)\|_{H^s} + \|\partial_t u(t)\|_{H^{s-1}}] = \infty.$$

*Proof.* Suppose  $T_+ < \infty$ , and suppose for a contradiction that

$$(32) \quad \mathcal{E}(t) = \|u(t)\|_{H^s} + \|\partial_t u(t)\|_{H^{s-1}}$$

remains bounded as  $t \nearrow T_+$ . Since the time of existence in Theorem 4 depends only on the  $(H^s \times H^{s-1})$ -norm of the initial data, we can apply Theorem 4 to  $u$  at a time  $t$  sufficiently close to  $T_+$ , which will extend  $u$  beyond time  $T_+$ . This contradicts the assumption that  $u$  was a maximal solution.  $\square$

We can view Corollary 7 as a *continuation criterion* for solutions of (17):

If  $u$  solves (17) over a finite time interval, and if  $(u, \partial_t u)$  has uniformly bounded  $(H^s \times H^{s-1})$ -norms, then  $u$  can be further extended in time.

In addition to Corollary 7, one can find other continuation criteria for (17) that, for example, may be easier to check. One such classical result is the following:

**Theorem 8** (Continuation criterion). *Suppose  $u : [0, T) \times \mathbb{R}^n \rightarrow \mathbb{R}$  is a solution to (17), for some  $0 < T < \infty$ , and suppose*

$$u \in C^0([0, T']; H^s(\mathbb{R}^n)) \cap C^1([0, T']; H^{s-1}(\mathbb{R}^n))$$

for any  $0 < T' < T$ . If

$$\|\partial_{t,x} u\|_{L^\infty([0, T] \times \mathbb{R}^n)} < \infty,$$

then  $u$  can be extended beyond time  $T$ , i.e., there is some  $\varepsilon > 0$  with

$$u \in C^0([0, T + \varepsilon]; H^s(\mathbb{R}^n)) \cap C^1([0, T + \varepsilon]; H^{s-1}(\mathbb{R}^n)).$$

*Proof.* Let  $\mathcal{E}(t)$  be defined as in (32). Applying the energy inequality (16) and the product estimate (26), we obtain for any  $t \in [0, T)$  that

$$\begin{aligned} \mathcal{E}(t) &\lesssim (1 + T) \left[ \|f\|_{H^s} + \|g\|_{H^{s-1}} + \int_0^t \|\partial u(s)\|_{L^\infty} \|\partial u(s)\|_{H^{s-1}} ds \right] \\ &\lesssim \|f\|_{H^s} + \|g\|_{H^{s-1}} + \int_0^t R(s) ds. \end{aligned}$$

By Gronwall's inequality (see [10, Ch. 5] or [12]), one uniformly bounds  $\mathcal{E}(t)$  for every  $t \in [0, T)$ . Corollary 7 then implies that  $u$  can be further extended in time.  $\square$

**2.2.3. Finite Speed of Propagation.** Next, we mention that an analogue of the finite speed of propagation property for homogeneous linear wave equations, Theorem 1, holds for many nonlinear wave equations, including (17).

**Theorem 9** (Finite speed of propagation). *Suppose  $u$  solves (17). If  $f$  and  $g$  vanish on the ball  $\{x \in \mathbb{R}^n \mid |x| \leq R\}$ , then  $u$  vanishes on the set*

$$C = \{(t, x) \in \mathbb{R} \times \mathbb{R}^n \mid |x| \leq R - |t|\}.$$

*Proof.* See [10, Ch. 5].  $\square$

We sketch here the main points of the proof of Theorem 9. Suppose  $u$  and  $v$  both solve (17), so that  $v - u$  satisfies the nonlinear wave equation

$$\square(v - u) = (\partial u + \partial v)\partial(v - u),$$

with zero initial data. We then derive a variant of the energy estimate (16) for  $v - u$ , but *localized to the solid cone  $C$* . From this estimate and Gronwall's inequality, we can control the local energy for  $v - u$  at any time (localized to  $C$ ) by the localized initial energy, which vanishes. It follows that  $v - u$  vanishes on all of  $C$ .

Now, while the preceding discussions fail to prove global existence for all nice initial data, they leave unanswered the question of whether exists nice data for which the solution blows up in finite time. Using Theorem 9, though, we can very easily construct such blowup solutions. Consider the special case

$$\square u = (\partial_t u)^2.$$

By first letting our initial data  $f$  and  $g$  to be constant functions, and by assuming  $u$  to be independent of the spatial variable, the above reduces to the ODE

$$y' = -y^2, \quad y = \partial_t u.$$

By solving this explicitly, we see that, for appropriately chosen  $g$ , solutions will blow up in finite time. Next, consider smooth and compactly supported  $f$  and  $g$  such that both are constant as above on a ball of sufficiently large radius. Recalling the finite speed of propagation, Theorem 9, the solution arising from this new  $f$  and  $g$  will be identical to the preceding solution in the corresponding solid cone  $C$ . As a result, we can construct smooth and compactly supported initial data such that the corresponding solutions blow up in finite time. Therefore, unlike the linear case, we cannot hope to achieve global solutions in general.

**2.2.4. Lower Regularity.** For Theorem 4 and the discussions above, we required that the initial data for (17) lie in  $H^s \times H^{s-1}$  for  $s > n/2 + 1$ . One can then ask is whether there exist local solutions for less regular initial data, that is, when  $s \leq n/2 + 1$ . Note that to find such solutions, one would require a proof that is different from that of Theorem 4, since one can no longer rely on Theorem 6.

Thus, any results that push  $s$  down to and below  $n/2 + 1$  would require some mechanism to make up for the lack of spatial derivatives. The key observation is the time integral on the right-hand side of the energy estimate (16), which one can interpret as an antiderivative with respect to  $t$ . This time integral can be shown to have a smoothing effect on solutions of wave equations (note that the wave equation  $\partial_t^2 u = \Delta u$  can be interpreted as being able to trade space derivatives for

time derivatives, and vice versa). The smoothing effect is manifested in a class of estimates for the wave equation known as Strichartz estimates.

Using such Strichartz estimates, one can reduce the regularity required in Theorem 4 to  $s > n/2 + 1/2$  whenever  $n \geq 3$  (when  $n = 2$ , Strichartz estimates reduce the required regularity to  $s > n/2 + 3/4$ ). For further expositions on Strichartz estimates for wave equations, see [6, 10]. To further reduce the needed regularity, one requires instead *bilinear estimates*; see [2, 13]. Furthermore, in many cases, if one has particularly favorable quadratic nonlinearity (in particular, if  $(\partial u)^2$  in (17) is given by so-called “null form”), then one can further push down the required regularity; see [7, 4]. Lastly, for sufficiently small  $s$ , local existence of solutions to (17) is false; see [10, Ch. 9] and [8].

### 3. GLOBAL WELL-POSEDNESS: THE VECTOR FIELD METHOD

Although the preceding discussion demonstrated that global well-posedness for the quadratic DNLW is impossible for general data in  $C_0^\infty(\mathbb{R}^N)$ , one can still ask whether this could occur for *sufficiently small* initial data, for which one would expect the linear part of the equation to dominate in the evolution. To investigate this, we return once again to our model problem

$$(33) \quad \square u = (\partial u)^2, \quad u|_{t=0} = \varepsilon f, \quad \partial_t u|_{t=0} = \varepsilon g,$$

where we assume  $f, g \in C_0^\infty(\mathbb{R}^n)$ . The additional parameter  $\varepsilon$ , in contrast to (17) in the previous section, signifies that we are to consider only small initial data.

**Theorem 10** (Small-data global and long-time existence and uniqueness, [5]). *Suppose  $\varepsilon$  is sufficiently small, with respect to  $f$ ,  $g$ , and  $n$ .*

- *If  $n \geq 4$ , then (33) has a unique global solution.*
- *Otherwise, the maximal time of existence  $T$  of the solution to (33) satisfies:*
  - *If  $n = 3$ , then  $T \geq e^{C\varepsilon^{-1}}$ .*
  - *If  $n = 2$ , then  $T \geq C\varepsilon^{-2}$ .*
  - *If  $n = 1$ , then  $T \geq C\varepsilon^{-1}$ .*

This section is dedicated to the proof of Theorem 10. For a more detailed exposition on this topic, see [10, Ch. 7] or [3, 5, 11].

**Remark.** *Note that uniqueness automatically follows from the arguments of the preceding section, so that we only need to consider existence here.*

**3.1. Main Ideas.** Recall the proof of local existence and uniqueness, i.e., Theorem 4, is based on the energy estimate. Indeed, by (16) and standard product estimates (see Theorem 6), we obtained for any solution  $u$  the estimate

$$(34) \quad \mathcal{E}(t) \lesssim \mathcal{E}(0) + \int_0^t [\mathcal{E}(s)]^2 ds,$$

where  $\mathcal{E}(t)$  denotes the usual “energy” quantity defined by (32):

$$\mathcal{E}(t) = \|u(t)\|_{H^s} + \|\partial_t u(t)\|_{H^{s-1}}, \quad s > \frac{n}{2} + 1.$$

From this estimate, we generated the contraction needed to obtain our solution.

For this, we required that the maximum time of consideration  $T$  is small, so that the integral of the nonlinearity  $\mathcal{E}^2$  in (34) does not sufficiently interfere with the linear evolution. If  $T$  becomes large enough, however, then the entire argument

breaks down, since the nonlinear evolution may dominate. Moreover, the map  $\Phi$  defined in (22) is no longer guaranteed to be a contraction. In particular, (34) is not enough to ensure that  $\mathcal{E}(t)$  does not blow up at a finite time.

Suppose now that we have a stronger “energy estimate”,

$$(35) \quad \mathcal{E}(t) \lesssim \mathcal{E}(0) + \int_0^t \frac{[\mathcal{E}(s)]^2}{(1+s)^p} ds, \quad p > 0,$$

where  $\mathcal{E}(t)$  now denotes some alternate “energy quantity”. In other words, suppose the nonlinear estimate comes with an additional decay in time. If this power  $p$  of decay is sufficiently large (i.e.,  $(1+s)^{-p}$  is integrable), then the largeness of  $T$  is no longer the devastating obstruction it once was. In this case, the smallness of  $\mathcal{E}(t)$  itself is sufficient to show that the nonlinear evolution is dominated by the linear evolution, regardless of the size of  $t$ . More specifically, as long as  $\mathcal{E}(0)$  is sufficiently small, one can control  $\mathcal{E}(t)$  by  $\mathcal{E}(0)$  for all  $t > 0$ .

Once we have such a global-in-time bound for  $\mathcal{E}(t)$ , assuming that  $\mathcal{E}(t)$  is a “reasonable” energy quantity, we can apply standard Sobolev estimates to uniformly bound  $\partial_{t,x}u$ , independently of  $t$ . The classical continuation criterion, Theorem 8, then implies one can *always* further continue this local solution. Therefore, one has global solutions to (33) for sufficiently small initial data (that is, for  $\varepsilon$  sufficiently small). This is the idea behind the proof of Theorem 10 when  $n \geq 4$ .

Furthermore, when  $n \leq 3$ , the power  $p$  will be small enough such that  $(1+s)^{-p}$  is no longer integrable. In this case, one can no longer obtain global existence, since the nonlinear evolution will again dominate for large times. However, we can estimate how large  $T$  can be before the nonlinear evolution can dominate. This is responsible for lower bounds for the time of existence in Theorem 10 when  $n \leq 3$ .

The remaining questions, then, are the following:

- What is this alternate energy quantity  $\mathcal{E}(t)$ ?
- How does one obtain this improved energy estimate (35) for  $\mathcal{E}(t)$ ?

The answers to these are also the main ideas behind Theorem 10.

**3.2. The Invariant Vector Fields.** The key innovations behind Theorem 10 are *geometric* in nature. To arrive at the main ideas, we must consider the geometry of Minkowski spacetime, which is fundamentally tied to the wave operator. To fully explain these ideas, we must invoke some basic notions from differential geometry.

**Remark.** *While the intuitions here are best described in geometric terms, readers who do not have a firm background in differential geometry should still be able to follow the details of the proof of Theorem 10 in its entirety.*

We begin by defining the following set of vector fields on  $\mathbb{R}^{n+1}$ :

- *Translations:* The coordinate vector fields

$$(36) \quad \partial_\alpha, \quad 0 \leq \alpha \leq n,$$

which generate translations on  $\mathbb{R}^{n+1}$ , that is, the (Lie) transport along the integral curves of the  $\partial_\alpha$ ’s are translations on  $\mathbb{R}^{n+1}$ .

- *Spatial rotations:* The vector fields,

$$(37) \quad \Omega_{ij} = x^j \partial_i - x^i \partial_j, \quad 1 \leq i < j \leq n,$$

which generate spatial rotations on each level set of  $t$ .

- *Lorentz boosts*: The vector fields,

$$(38) \quad \Omega_{0j} = x^j \partial_t + t \partial_j, \quad 1 \leq j \leq n,$$

which generate *Lorentz boosts* on  $\mathbb{R}^{n+1}$ .

- *Scaling/dilation*: The vector field

$$(39) \quad S = t \partial_t + \sum_{i=1}^n x^i \partial_i,$$

which generate the (spacetime) dilations on  $\mathbb{R}^{n+1}$ .

Note that (36)-(39) define exactly

$$\gamma_n = (n+1) + \frac{n(n-1)}{2} + n + 1 = \frac{(n+2)(n+1)}{2} + 1$$

vector fields, which for convenience we label as  $\Gamma_1, \dots, \Gamma_{\gamma_n}$ .

Furthermore, we will use multi-index notation to denote successive applications of various such  $\Gamma_a$ 's. More specifically, given a multi-index

$$I = (I_1, \dots, I_d), \quad 1 \leq I_i \leq \gamma_n,$$

we let  $\Gamma^I$  denote the operator

$$(40) \quad \Gamma^I = \Gamma_{I_1} \Gamma_{I_2} \dots \Gamma_{I_d}.$$

Note that since the  $\Gamma_a$ 's generally do *not* commute with each other, the ordering of the coefficients in such a multi-index  $I$  carries nontrivial information.

**3.2.1. Geometric Ideas.** We now discuss the geometric significance of the  $\Gamma_a$ 's, as well as their relation to the wave equation. Recall that *Minkowski spacetime* can be described as the *Lorentzian manifold*  $(\mathbb{R}^{n+1}, m)$ . Here,  $m$  is the *Minkowski metric*, i.e., the symmetric covariant 2-tensor on  $\mathbb{R}^{n+1}$  given by

$$m = -(dt)^2 + (dx^1)^2 + (dx^n)^2.$$

**Remark.** *In particular, the Minkowski metric is analogous to the Euclidean metric, except that the sign associated to one of its components is reversed. This, in general, is the distinction between Riemannian and Lorentzian metrics.*

Next, recall that the wave operator  $\square$  on  $\mathbb{R}^{n+1}$  arises entirely from the geometry of  $(\mathbb{R}^{n+1}, m)$ . Indeed,  $\square$  is the Laplace-Beltrami operator associated with  $(\mathbb{R}^{n+1}, m)$ : letting  $\nabla$  denote the  $(m)$ -covariant derivative then

$$\square = m^{ij} \nabla_{ij}^2 = -\partial_t^2 + \sum_{i=1}^n \partial_i^2.$$

Observe now that any vector field  $\Gamma$  given by (36)-(38) is a *Killing vector field* on  $(\mathbb{R}^{n+1}, m)$ , that is,  $\Gamma$  *generates symmetries of Minkowski spacetime*. In terms of the language of differential geometry, this is given by  $\mathcal{L}_\Gamma m = 0$ , where  $\mathcal{L}$  denotes the Lie derivative. In other words, transport along the integral curves of  $\Gamma$  does not change the Minkowski metric, so that such a transformation preserves the Minkowski geometry. Since  $\square$  arises entirely from the Minkowski geometry, then transporting along  $\Gamma$  also preserves the wave operator. The most important consequence, in our setting, is that  $\Gamma$  *commutes with the wave operator*:

$$(41) \quad [\square, \Gamma] = \square \Gamma - \Gamma \square \equiv 0, \quad \Gamma \in \{\partial_\alpha, \Omega_{ij}, \Omega_{0j}\}.$$

**Remark.** *In fact, the vector fields given in (36)-(38) generate the Lie algebra of all Killing vector fields on Minkowski spacetime.*

On the other hand,  $S$ , defined in (39) is *not* a Killing vector field and hence does not generate a symmetry of Minkowski spacetime. However,  $S$  is a *conformal Killing vector field*— $S$  generates a conformal symmetry of Minkowski spacetime. As this is not a full symmetry,  $S$  will not commute with  $\square$ . But, the conformal symmetry property ensures that this commutator will be manageable:

$$(42) \quad [\square, S] = 2\square.$$

Combining the above results in the following lemma:

**Lemma 11.** *For any vector field  $\Gamma_a$  and  $0 \leq \alpha \leq n$ , we have*

$$(43) \quad [\square, \Gamma_a] = c_a \square, \quad c_a \in \mathbb{R},$$

$$(44) \quad [\partial_\alpha, \Gamma_b] = \sum_{\beta=0}^n c_{\alpha b}^\beta \partial_\beta, \quad c_{\alpha b}^\beta \in \mathbb{R}.$$

*Proof.* (43) follows immediately from (41) and (42), while (44) is an immediate consequence of the fact that for any such  $\Gamma_a$ , its coefficients (expressed in Cartesian coordinates) are always either constant or one of the coordinate functions.  $\square$

For our purposes, the main properties we will need are the identities in Lemma 11, which can be checked explicitly without reference to any geometric discussion. Thus, one does not require any actual background in differential geometry to prove Theorem 10. However, to understand the full intuition behind such vector fields, one requires familiarity with Minkowski geometry.

**3.2.2. The Modified Energy.** Suppose  $u$  satisfies (33). Because of the commutation identities (43), then  $\Gamma_a u$ ,  $1 \leq a \leq \gamma_n$ , satisfies a similarly “nice” wave equation:

$$(45) \quad \square \Gamma u = \Gamma \square u + c \square u = \Gamma \partial u \cdot \partial u + c(\partial u)^2, \quad c \in \mathbb{R}.$$

As a result, we can also apply the energy estimate (16) to control  $\partial \Gamma_a u$  in the same manner that we previously controlled  $\partial u$ . Moreover, the same observations hold for any number of  $\Gamma_a$ ’s applied to  $u$ —for any multi-index  $I$ , we see that  $\Gamma^I u$  satisfies a “nice” wave equation, with which one can apply energy estimates to control  $\partial \Gamma^I u$ .

This leads us to define the following modified energy quantity for  $u$ :

$$(46) \quad \mathcal{E}(t) = \sum_{|I| \leq n+4} \|\partial_{t,x} \Gamma^I u(t)\|_{L^2}.$$

We now wish to show that this satisfies an improved energy estimate (35).

By using (43) repeatedly, we can estimate

$$|\square \Gamma^I u| \leq |\Gamma^I \square u| + |[\square, \Gamma^I] u| \lesssim \sum_{|J| \leq |I|} |\Gamma^J \square u|.$$

Applying (33) and the product rule, the right-hand side can be further expanded:

$$|\square \Gamma^I u| \lesssim \sum_{|J| \leq |I|} |\Gamma^J (\partial u)^2| \lesssim \sum_{|J|+|K| \leq |I|} |\Gamma^J \partial u| |\Gamma^K \partial u|.$$

In addition, applying (44) and noting that any  $\partial_\alpha$  is one of the  $\Gamma_a$ 's, we observe that we can commute  $\Gamma$  and  $\partial$  without affecting the estimate:

$$(47) \quad |\square \Gamma^I u| \lesssim \sum_{|J| \leq |I|} |\Gamma^J (\partial u)^2| \lesssim \sum_{|J|+|K| \leq |I|} |\partial \Gamma^J u| |\partial \Gamma^K u|.$$

Using the homogeneous energy estimate (15) on  $\Gamma^I u$  and applying (47) yields

$$(48) \quad \begin{aligned} \mathcal{E}(t) &= \sum_{|I| \leq n+4} \|\partial \Gamma^I u(t)\|_{L^2} \\ &\lesssim \sum_{|I| \leq n+4} \left[ \|\partial \Gamma^I u(0)\|_{L^2} + \int_0^t \|\square \Gamma^I u(s)\|_{L^2} ds \right] \\ &\lesssim \mathcal{E}(0) + \sum_{|J|+|K| \leq n+4} \int_0^t \|\partial \Gamma^J u(s)\| \|\partial \Gamma^K u(s)\|_{L^2} ds, \end{aligned}$$

where we note that the initial value for  $\Gamma^I u$  is  $f_I = \Gamma^I u(0)$  and  $g_I = \partial_t \Gamma^I u(0)$ . Now, since  $|J| + |K|$  on the right-hand side of (48) is at most  $n+4$ , we can assume without loss of generality that  $|J| \leq n/2 + 2$ . Using Hölder's inequality, we control the  $\Gamma^J$ -factor in  $L^\infty$  and the  $\Gamma^K$ -factor in  $L^2$ :

$$(49) \quad \begin{aligned} \mathcal{E}(t) &\lesssim \mathcal{E}(0) + \int_0^t \sum_{|J| \leq \frac{n}{2}+2} \|\partial \Gamma^J u(s)\|_{L^\infty} \sum_{|K| \leq n+4} \|\partial \Gamma^K u(s)\|_{L^2} ds \\ &\lesssim \mathcal{E}(0) + \int_0^t \sum_{|J| \leq \frac{n}{2}+2} \|\partial \Gamma^J u(s)\|_{L^\infty} \cdot \mathcal{E}(s) \cdot ds. \end{aligned}$$

Previously, we extracted the energy from an  $L^\infty$ -norm of  $u$  by applying the Sobolev inequality (25). In our setting, this results in the bound

$$(50) \quad \|v(t)\|_{L^\infty} \lesssim \sum_{k \leq \frac{n}{2}+1} \|\partial^k v(t)\|_{L^2} \lesssim \sum_{|I| \leq \frac{n}{2}+1} \|\Gamma^I v(t)\|_{L^2}.$$

Note, however, that we are losing information here, since we are considering all the vector fields  $\Gamma_a$ , not just the  $\partial_\alpha$ 's. By leveraging the fact that the  $\Omega_{\alpha\beta}$ 's and  $S$  have growing weights, we see the possibility of an improvement to (50), with additional weights on the left-hand side that grow. In fact, there does exist such an estimate, which is known as the *Klainerman-Sobolev*, or *global Sobolev, inequality*:

**Theorem 12.** *Let  $v : [0, \infty) \times \mathbb{R}^n \rightarrow \mathbb{R}$  be smooth, with each  $v(t)$  compactly supported. Then, the following estimate holds for each  $t \geq 0$  and  $x \in \mathbb{R}^n$ :*

$$(51) \quad (1+t+|x|)^{\frac{n-1}{2}} (1+|t-|x||)^{\frac{1}{2}} |v(t, x)| \lesssim \sum_{|I| \leq \frac{n}{2}+1} \|\Gamma^I v(t)\|_{L^2}.$$

The main idea behind the proof of (51) is to write  $\partial_\alpha$  as linear combinations of the  $\Gamma_a$ 's, which introduces decaying weights, and to then apply a standard Sobolev inequality (either on  $\mathbb{R}^n$  or on  $\mathbb{S}^{n-1}$ ). One then chooses different combinations of  $\Gamma_a$ 's depending on  $t$  and  $x$ , in order to maximize the decaying weight. For details, the reader is referred to either [10, Ch. 7] or [3, 11].

In particular, when we apply Theorem 12 to  $u$ , the right-hand side of (51) will be uniformly bounded in  $t$  by energy estimates. As a result, (51) implies decay for  $u$  in both  $t$  and  $|x|$ . Furthermore, the weight on the left-hand side of (51) indicates



that  $u$  will decay a half-power *faster* away from the null cone  $t = |x|$ . For our current problem, though, we will not need to consider the decay in  $|x|$  or in  $|t - |x||$ .

Indeed, applying (51) to the  $L^\infty$ -norm in the right-hand side of (49), we have

$$\mathcal{E}(t) \lesssim \mathcal{E}(0) + \int_0^t (1+s)^{-\frac{n-1}{2}} \sum_{|J|+|K| \leq n+4} \|\Gamma^K \partial \Gamma^J u(s)\|_{L^2} \cdot \mathcal{E}(s) \cdot ds.$$

Commuting  $\Gamma_a$ 's and  $\partial_\alpha$ 's again using (44), we obtain that

$$\mathcal{E}(t) \lesssim \mathcal{E}(0) + \int_0^t (1+s)^{-\frac{n-1}{2}} \sum_{|I| \leq n+4} \|\partial \Gamma^I u(s)\|_{L^2} \cdot \mathcal{E}(s) \cdot ds.$$

As a result, we have proved the following:

**Lemma 13.** *Suppose  $u$  is a solution of (33) which exists on a time interval  $[0, T)$ , and let  $\mathcal{E}(t)$  be as in (46). Then, for any  $0 \leq t < T$ ,*

$$(52) \quad \mathcal{E}(t) \lesssim \mathcal{E}(0) + \int_0^t \frac{[\mathcal{E}(s)]^2}{(1+s)^{\frac{n-1}{2}}} ds.$$

**Remark.** *Note that we must prescribe a high enough number of derivatives in the definition of  $\mathcal{E}(t)$ , so that after applying (51) to the  $L^\infty$ -factor containing  $u(t)$ , the resulting  $L^2$ -norms are still controlled by  $\mathcal{E}(t)$ .*

**3.3. Proof of Theorem 10.** We now use the energy estimate (52) to complete the proof of Theorem 10, when  $n \geq 4$ . The main step is the following bound:

**Lemma 14.** *Assume  $n \geq 4$ , and assume  $\varepsilon$  is sufficiently small. Suppose  $u$  is a solution of (33) on a time interval  $[0, T)$ , and let  $\mathcal{E}(t)$  be as in (46). Then,*

$$(53) \quad \mathcal{E}(t) \lesssim \mathcal{E}(0), \quad 0 \leq t < T.$$

Suppose Lemma 14 has been established. Then, if  $u$  is as in the statement of Lemma 14, then by Sobolev estimates—either (50) or (51)—it follows that  $\partial_{t,x} u$  is uniformly bounded on  $[0, T) \times \mathbb{R}^n$ . Thus, Theorem 8 implies that as long as  $\varepsilon$  is small enough,  $u$  can be extended as a solution of (33) beyond time  $T$ . In other words, for sufficiently small data, a local solution can always be further extended in time. This completes the proof of Theorem 10 when  $n \geq 4$ .

**3.3.1. The Bootstrap Argument.** When  $n \geq 4$ , it remains only to prove Lemma 14. The main step is a so-called *bootstrap*, or *continuity*, argument, in which *one first assumes the required bound and then proves a strictly better bound*. This may seem peculiar at first, and even uncomfortably like circular logic. However, the justification for this argument can be encapsulated in the following trivial proposition:

**Proposition 15.** *Fix  $A > 0$ , and let  $\mathcal{F} : [0, \infty) \rightarrow \mathbb{R}$  be continuous. Suppose  $|\mathcal{F}(0)| < A$ , and assume the following condition holds for each  $t \in [0, \infty)$ :*

- **(BS)** *If  $|\mathcal{F}(s)| < 2A$  for each  $s \in [0, t]$ , then  $|\mathcal{F}(t)| < A$ .*

*Then,  $|\mathcal{F}(t)| < A$  for all  $t \in [0, \infty)$ .*

We give two extremely elementary proofs of Proposition 15.

*Proof 1.* Suppose, for a contradiction, that  $|\mathcal{F}|$  is not everywhere less than  $2A$ , and let  $t_0$  be the first time such that  $|\mathcal{F}(t_0)| = 2A$  (note that  $t_0 > 0$  by assumption). The condition **(BS)** implies  $|\mathcal{F}(s)| < A$  for all  $s < t_0$ , and it follows via continuity

that  $|\mathcal{F}(t_0)| \leq A$ , a contradiction. As a result,  $|\mathcal{F}(t)| < 2A$  for all  $t \in [0, \infty)$ , and applying **(BS)** again yields that, in fact,  $|\mathcal{F}(t)| < A$  for all  $t$ .  $\square$

The idea in the above is that because of **(BS)**, if we graph of  $\mathcal{F}$  on an  $x$ - $y$ -plane, then this graph could never come close to the vertical line  $y = 2A$ . Now, while this proof is the most intuitive, the following topological proof is more robust.

*Proof 2.* Define the set

$$\mathcal{A} = \{t \in [0, \infty) \mid |\mathcal{F}(s)| < 2A \text{ for all } 0 \leq s < t\},$$

which is nonempty, since  $0 \in \mathcal{A}$ . Since  $\mathcal{F}$  is continuous, it follows that  $\mathcal{A}$  is closed. Also, if  $t \in \mathcal{A}$ , then **(BS)** implies that  $|\mathcal{F}(s)| < A$  for all  $0 \leq s < t$ . By continuity, it follows that  $t+h \in \mathcal{A}$  for sufficiently small  $|h|$ , and hence  $\mathcal{A}$  is open. Since  $[0, \infty)$  is connected, it follows that  $\mathcal{A}$  must be all of  $[0, \infty)$ . Thus, again,  $|\mathcal{F}(t)| < 2A$  for all  $t \in [0, \infty)$ , and another application of **(BS)** implies  $|\mathcal{F}(t)| < A$  for all  $t$ .  $\square$

To apply Proposition 15 toward proving Lemma 14, we let  $\mathcal{F}$  be the modified energy  $\mathcal{E}$  in (46). By (33), we see that  $\mathcal{E}(0) = \varepsilon B$  for some  $B$ . Thus, we set  $A$  in Proposition 15 to be  $A = \varepsilon CB$  for some large  $C > 0$  (note that  $\mathcal{E}(0) < A$ ).

Next, we wish to show that the condition **(BS)** holds in this setting. Fix  $t > 0$ , and suppose  $\mathcal{E}(s) < 2A = 2\varepsilon CB$  for all  $0 \leq s < t$ . Then, applying (52) and recalling that the constant  $C$  is sufficiently large, we have

$$\begin{aligned} \mathcal{E}(t) &\leq \frac{1}{2}C \cdot \mathcal{E}(0) + \frac{1}{2}C \int_0^t \frac{[\mathcal{E}(s)]^2}{(1+s)^{\frac{n-1}{2}}} ds \\ &\leq \frac{1}{2}\varepsilon CB + \frac{1}{2}C(2\varepsilon CB)^2 \int_0^\infty \frac{1}{(1+s)^{\frac{n-1}{2}}} ds \\ &\leq \frac{1}{2}\varepsilon CB + C'(\varepsilon CB)^2, \end{aligned}$$

where  $C'$  is another constant, since the above integral over  $[0, \infty)$  is finite. Now, if  $\varepsilon$  is small enough compared to  $B$  (and hence with respect to  $f$  and  $g$ ) and  $n$ , then

$$C'(\varepsilon CB)^2 < \frac{1}{2}\varepsilon CB,$$

and hence it follows that

$$\mathcal{E}(t) < \varepsilon CB = A.$$

This establishes **(BS)**, and Proposition 15 implies  $\mathcal{E}(t) < \varepsilon CB$  for all  $t \geq 0$ . In particular, this proves Lemma 14, and hence Theorem 10 when  $n \geq 4$ .

**3.3.2. Low Dimensions.** The proof of Theorem 10 in low dimensions (i.e.,  $n \leq 3$ ) is analogous to the previous case  $n \geq 4$ . We briefly sketch the argument here.

The main point is that the bootstrap argument described above can be localized to a finite interval  $[0, T)$ . In particular, Proposition 15 remains valid when  $[0, \infty)$  is replaced by  $[0, T)$ . Retracing previous steps, we obtain from (52) that

$$\mathcal{E}(t) \leq \frac{1}{2}\varepsilon CB + \frac{1}{2}C(2\varepsilon CB)^2 \int_0^T \frac{1}{(1+s)^{\frac{n-1}{2}}} ds.$$

The problem, now, is that  $(1+s)^{-(n-1)/2}$ , on the right-hand side, is no longer integrable over all of  $[0, \infty)$ . Note, however, that as long as

$$(54) \quad \frac{1}{2}C(2\varepsilon CB)^2 \cdot \int_0^T \frac{1}{(1+s)^{\frac{n-1}{2}}} ds < \frac{1}{2}\varepsilon CB,$$

then we can still control  $\mathcal{E}(t)$  by  $\mathcal{E}(0)$  as before whenever  $0 \leq t < T$ .

By direct computations, we have that

$$(55) \quad \int_0^T \frac{1}{(1+s)^{\frac{n-1}{2}}} ds \simeq \begin{cases} \log(1+T) & n=3, \\ (1+T)^{\frac{1}{2}} & n=2, \\ (1+T) & n=1. \end{cases}$$

Combining (54) and (55) results in the lower bounds for  $T$  in Theorem 10.

#### 4. GLOBAL WELL-POSEDNESS: THE NULL CONDITION

To be added later...

#### REFERENCES

1. L. C. Evans, *Partial differential equations*, AMS, 2002.
2. D. Foschi and S. Klainerman, *Bilinear space-time estimates for homogeneous wave equations*, Ann. Sci. École Norm. Sup. **33** (2000), no. 2, 211–274.
3. L. Hörmander, *Lectures on nonlinear hyperbolic differential equations*, Springer-Verlag, 1997.
4. M. Keel and T. Tao, *Local and global well-posedness of wave maps on  $\mathbb{R}^{1+1}$  for rough data*, Internat. Math. Res. Notices **21** (1998), 1117–1156.
5. S. Klainerman, *Uniform decay estimates and the Lorentz invariance of the classical wave equation*, Comm. Pure Appl. Math. **38** (1985), no. 3, 321–332.
6. ———, *Lecture notes. Introduction to analysis*, <https://web.math.princeton.edu/~seri/homepage/courses/Analysis2011.pdf>, 2011.
7. S. Klainerman and S. Selberg, *Remark on the optimal regularity for equations of wave maps type*, Comm. Partial Differential Equations **22** (2006), no. 5–6, 901–918.
8. H. Lindblad, *A sharp counterexample to the local existence of low-regularity solutions to nonlinear wave equations*, Duke Math. J. **72** (1993), no. 2, 503–539.
9. W. Rudin, *Principles of mathematical analysis*, McGraw-Hill, Inc., 1976.
10. S. Selberg, *Lecture notes. Math 632, PDE*, <http://www.math.ntnu.no/~sselberg/HopkinsLectures.pdf>.
11. C. Sogge, *Lectures on nonlinear wave equations*, International Press, 2008.
12. T. Tao, *Nonlinear dispersive equations: Local and global analysis*, American Mathematical Society, 2006.
13. D. Tataru, *On the equation  $\square u = |\nabla u|^2$  in  $5+1$  dimensions*, Math. Res. Lett. **6** (1999), no. 5–6, 469–485.