LOCAL WELL-POSEDNESS FOR QUASILINEAR WAVE EQUATIONS (D'APRÉS SMITH-TATARU)

JASON ZHAO

ABSTRACT. In this note, we outline the work by Smith-Tataru [ST05] concerning the sharp local well-posedness for generic quasi-linear wave equations. That is, given sufficiently regular Lorentzian metrics $\mathbf{g}_{\mu\nu}(\phi)$ and semi-linear terms $\mathcal{N}(\phi)(\partial\phi, \partial\phi)$, we prove that the initial data problem

$$\begin{aligned} \Box_{\mathbf{g}(\phi)}\phi &= \mathcal{N}(\phi)(\partial\phi, \partial\phi), \\ (\phi, \partial_t\phi)_{|t=0} &= (\phi_0, \phi_1), \end{aligned}$$

is locally well-posed in $H^s_x \times H^{s-1}_x(\mathbb{R}^n)$ for $s > \frac{n}{2} + \frac{1}{2}$ when $n = 3, 4, 5$.

CONTENTS

1. INTRODUCTION

In this note, we consider the local well-posedness of *quasilinear wave equations* of the form

$$\begin{aligned} \Box_{\mathbf{g}(\phi)}\phi &= \mathcal{N}(\phi)(\partial\phi, \partial\phi), && \text{on } [0,T] \times \mathbb{R}^n, \\ (\phi, \partial_t\phi) &= (\phi_0, \phi_1), && \text{on } t = 0, \end{aligned} \tag{QNLW}$$

where $\mathbf{g}_{\mu\nu}(\phi)$ is a symmetric matrix with signature $(-, +, \cdots +)$, using the convention[1] $\Box_{\mathbf{g}} := \mathbf{g}^{\mu\nu}\partial_\mu\partial_\nu$ for its associated wave operator, and $\mathcal{N}(\phi)(\partial\phi, \partial\phi) := \mathcal{N}^{\alpha\beta}(\phi)\partial_\alpha\phi\partial_\beta\phi$ is a bilinear form. Without loss of generality, we can take $t = \text{const}$ to be space-like hypersurfaces by reducing to metrics of the form

$$\mathbf{g}_{\mu\nu}dx^\mu dx^\nu = -dt^2 + \mathbf{g}_{ij}dx^i dx^j.$$

We shall also assume sufficient smoothness and boundedness of the metric $\mathbf{g}^{\mu\nu}(\phi)$, its inverse $\mathbf{g}_{\mu\nu}(\phi)$, and of the bilinear form $\mathcal{N}^{\alpha\beta}(\phi)$ as functions of ϕ.

Example. The following can be recast in the form (QNLW),

- the Einstein vacuum equations in wave coordinates,
- the irrotational compressible Euler equations.

For the former, this was observed by Choquet-Bruhat [Fou52], while the later is due to Hughes-Kato-Marsden [HKM77]. The reader may find the lecture notes [Luk] as a more modern reference.

Following the standard set by Hadamard, we say that the initial data problem for the quasi-linear wave equation (QNLW) is *locally well-posed* in $(H^s_x \times H^{s-1}_x)(\mathbb{R}^n)$ if the following hold:

Date: November 19, 2025.

[1] One can equivalently consider the divergence form of the equation, i.e. using $\partial_\mu\mathbf{g}^{\mu\nu}\partial_\nu$ instead of $\mathbf{g}^{\mu\nu}\partial_\mu\partial_\nu$ on the left-hand side, as the lower-order terms are encapsulated by the right-hand side.

(a) *Existence*: for each initial data $\phi[0] \in (H^s_x \times H^{s-1}_x)(\mathbb{R}^n)$, there exists a time $T > 0$ and a solution $\phi[t] \in C^0_t(H^s_x \times H^{s-1}_x)([0,T] \times \mathbb{R}^n)$ to (QNLW).
(b) *(Unconditional) uniqueness*: for each initial data $\phi[0] \in (H^s_x \times H^{s-1}_x)(\mathbb{R}^n)$, the solution $\phi[t]$ to (QNLW) is unique in the space $C^0_t(H^s_x \times H^{s-1}_x)([0,T] \times \mathbb{R}^n)$.
(c) *Continuity of data-to-solution map*: if $\{\phi_k[0]\}_k$ is a sequence of data converging in the $(H^s_x \times H^{s-1}_x)$-topology to $\phi[0]$, then there exists a common time of existence[2] on which the corresponding sequence of solutions $\{\phi_k[t]\}_k$ to (QNLW) converges to $\phi[t]$ in the $L^\infty_t(H^s_x \times H^{s-1}_x)$-topology,

$$\begin{aligned} &\phi_k[0] \xrightarrow{k\to\infty} \phi[0] \qquad \text{in } H^s_x \times H^{s-1}_x \\ \text{implies} \qquad &\phi_k[t] \xrightarrow{k\to\infty} \phi[t] \qquad \text{in } L^\infty_t(H^s_x \times H^{s-1}_x). \end{aligned}$$

This leads us to the following natural question

For which values of $s \in \mathbb{R}$ is the initial data problem for the quasi-linear wave equation (QNLW) locally well-posed in $(H^s \times H^{s-1})_x(\mathbb{R}^n)$?

The sharp answer to this question is due to Smith-Tataru [ST05], and will be the subject of this article. To simplify the presentation, we will only consider high spatial dimensions, i.e. $n \geq 3$; for $n = 2$, the dispersion of waves is weaker, though analogues of various statements made in this note continue to hold with suitable modifications. For the working definition, we will need to slightly modify the existence and uniqueness statements, strengthening the former while weakening the latter, and require an additional property of the data-to-solution map:

(a+) *(Sub-critical) existence*: the time of existence can be taken to depend only on the size of the data

$$T \equiv T(\|\phi[0]\|_{H^s_x \times H^{s-1}_x}).$$

(b') *(Conditional) uniqueness*: uniqueness holds only in the smaller Strichartz space,

$$\left\{\phi[0] \in C^0_t(H^s_x \times H^{s-1}_x) : \partial\phi \in L^2_t L^\infty_x\right\}.$$

(c+) *Weak Lipschitz continuity of data-to-solution map*: there exists a regularity $s_{\text{Lip}} < s$ such that the data-to-solution map is Lipschitz continuous on bounded sets in $(H^s \times H^{s-1})_x(\mathbb{R}^n)$ with respect to the weaker $(H^{s_{\text{Lip}}} \times H^{s_{\text{Lip}}-1})_x(\mathbb{R}^n)$-topology, i.e. for solutions $\phi[t], \psi[t] \in C^0_t(H^s_x \times H^{s-1}_x)([0,T] \times \mathbb{R}^n)$ to (QNLW) satisfying

$$\|\phi[0]\|_{H^s \times H^{s-1}}, \|\psi[0]\|_{H^s \times H^{s-1}} \leq R,$$

the following stability estimate holds:

$$\|\phi[t] - \psi[t]\|_{L^\infty_t(H^{s_{\text{Lip}}} \times H^{s_{\text{Lip}}-1})_x} \leq C(R) \cdot \|\phi[0] - \psi[0]\|_{(H^{s_{\text{Lip}}} \times H^{s_{\text{Lip}}-1})_x}.$$

In this setting, the sharp result, due to Smith-Tataru, may be stated as follows,

Theorem 1.1 (Sharp local well-posedness for (QNLW) [ST05])**.** *In dimensions $n = 3, 4, 5$, the initial data problem for the quasi-linear wave equation (QNLW) is locally well-posed in $(H^s \times H^{s-1})_x(\mathbb{R}^n)$ for $s > \frac{n}{2} + \frac{1}{2}$, with weak Lipschitz continuity of the data-to-solution map with respect to the $(H^1 \times L^2)_x$-topology.*

Remark. The equation (QNLW) is invariant under the scaling symmetry

$$\phi(x^\mu) \mapsto \phi(\tfrac{x^\mu}{\lambda})$$

which also preserves the homogeneous Sobolev norm $(\dot{H}^{s_{\text{crit}}} \times \dot{H}^{s_{\text{crit}}-1})_x(\mathbb{R}^n)$, where $s_{\text{crit}} := \frac{n}{2}$. It is natural to ask whether one can prove well-posedness up to the critical regularity. Unfortunately, Theorem 1.1 is sharp for generic (QNLW) in dimensions $n = 2, 3$ due to counterexamples of Lindblad [Lin93, Lin96]. On the other hand, it is not difficult to show that the *Nirenberg example*

$$\Box\phi = \partial^\alpha\phi\partial_\alpha\phi,$$

is locally well-posed for $s > s_{\text{crit}}$, thanks to the *null structure* of the non-linearity.

[2] To be more precise, one can introduce the notion of the *maximal lifespan* $T \equiv T(\phi[0])$ of a solution, and require it to be lower semi-continuous as a function of initial data $\phi[0] \in (H^s_x \times H^{s-1}_x)(\mathbb{R}^n)$.

Remark. The proof contained in [ST05] breaks down in higher dimensions $n \geq 6$ due to a technical failure in the orthogonality argument for the wave packet decomposition; see Section 5.

The basic starting point is the energy estimate, and local well-posedness result for sufficiently smooth initial data,

Theorem 1.2 (Local well-posedness for (QNLW) with smooth data [HKM77])**.** *The initial data problem for the quasi-linear wave equation (QNLW) is (unconditionally) locally well-posed in $(H^s_x \times H^{s-1}_x)(\mathbb{R}^n)$ for $s > \frac{n}{2}+1$. Furthermore, for all $s \geq 0$, there exists $C \gg 1$ such that any smooth solution ϕ obeys the a priori estimate*

$$\|\partial\phi\|_{L^\infty_t H^{s-1}_x} \lesssim \exp\left(C\int_0^T \|\partial\phi\|_{L^\infty_x}\,dt\right)\|\partial\phi(0)\|_{H^{s-1}_x}. \tag{1.1}$$

Proof. See [IT22] for a modern treatment. □

Remark. The dimension of the $L^p_t L^\infty_x$-norm of $\partial\phi$ under the scaling symmetry reads

$$\|\partial\phi\|_{L^p_t L^\infty_x} \approx [t]^{\frac{1}{p}}[x]^{-1} \approx [\partial]^{1-\frac{1}{p}}.$$

Thus, the continuation criterion $L^1_t L^\infty_x$ is scale-invariant, controlling $L^\infty_{t,x}$ via Sobolev embedding incurs a full derivative difference from scaling $1-\frac{1}{\infty}=1$, while control of $L^2_t L^\infty_x$ in $n \geq 3$ via Strichartz leads to half-derivative from scaling $1-\frac{1}{2}=\frac{1}{2}$, and similarly $L^4_t L^\infty_x$ in $n=2$ leads to three-quarters $1-\frac{1}{4}=\frac{1}{2}+\frac{1}{4}$.

To prove Theorem 1.1, one essentially needs to close the energy estimate. This is accomplished by proving Strichartz estimates for the linearised equation[3]

$$\begin{aligned} \Box_{\mathbf{g}(\phi)}\psi &= 0, \\ (\psi,\partial_t\psi)_{|t=0} &= (\psi_0,\psi_1), \end{aligned} \tag{LW}$$

and its paradifferential counterpart

$$\begin{aligned} \Box_{\mathbf{g}(\phi)_{<\lambda}}\mathsf{P}_\lambda\psi &= 0, \\ (\psi,\partial_t\psi)_{|t=0} &= (\mathsf{P}_\lambda\psi_0,\mathsf{P}_\lambda\psi_1), \end{aligned} \tag{PLW}$$

where ϕ is a solution to (QNLW). For flat backgrounds $\mathbf{g}(\phi) \equiv \mathbf{m}$, i.e. when (LW) reduces to the classical linear wave equation, one has the classical Strichartz estimates

$$\|\partial\phi\|_{L^\infty_t H^{s-1}_x} + \|\partial\phi\|_{L^2_t L^\infty_x} \lesssim \|\phi[0]\|_{H^s_x\times H^{s-1}_x}, \qquad s > \tfrac{n}{2}+\tfrac{1}{2},$$

with the $L^2_t L^\infty_x$-control allowing us to close the argument. If one only assumes $\partial\mathbf{g} \in L^2_t L^\infty_x$ in (LW), then the best one can do is to obtain the above Strichartz estimate but with a $\frac{1}{6}$-derivative loss [Tat01a, Tat01b]. The major breakthrough of Smith-Tataru was that actually there is no derivative-loss in Strichartz for (LW) when one uses that ϕ is a solution to (QNLW), which in turn roughly implies $\Box_{\mathbf{g}(\phi)}\mathbf{g}(\phi) \approx 0$, see Table 1 for a brief summary of this historical progression.

Theorem 1.3 (Loss-less Strichartz estimates for (QNLW) [ST05])**.** *Let $\phi[t] \in C^0_t(H^s_x \times H^{s-1}_x)([0,T]\times\mathbb{R}^n)$ be a solution to (QNLW) as in Theorem 1.1. Then*

(a) Strichartz bounds for non-linear evolution*: the solution satisfies*

$$\|\partial\phi\|_{L^\infty_t H^{s-1}_x} + \|\partial\phi\|_{L^2_t L^\infty_x} \lesssim \|\phi[0]\|_{H^s_x\times H^{s-1}_x}.$$

(b) Strichartz bounds for linear evolution*: for $1 \leq \sigma \leq s+1$ and each $t_0 \in [0,T]$, the initial data problem for the linear equation (LW) is well-posed in $(H^\sigma_x \times H^{\sigma-1}_x)(\mathbb{R}^n)$. Furthermore, the solution satisfies*

$$\|\partial\psi\|_{L^\infty_t H^{\sigma-1}_x} + \|\langle\nabla_x\rangle^{\sigma-\frac{n}{2}-\frac{1}{2}-1-}\partial\psi\|_{L^2_t L^\infty_x} \lesssim \|\psi[t_0]\|_{H^\sigma_x\times H^{\sigma-1}_x}. \tag{1.2}$$

[3] Strictly speaking, this is not quite the linearisation of (QNLW), as there is an extra order-zero term $(\partial_\phi\mathbf{g})^{\mu\nu}(\phi)\partial_\mu\partial_\nu\phi\,\psi$, however this can be easily treated as perturbative on short time-scales.

	g in (LW)	Strichartz	Well-posedness
Hughes-Kato-Marsden [HKM77]	$\partial \mathbf{g} \in L^\infty_{t,x}$	N/A	$s > \frac{n}{2} + 1$
Bahouri-Chemin [BC99]	$\partial \mathbf{g} \in L^2_t L^\infty_x$	$\frac{1}{4}$ loss	$s > \frac{n}{2} + \frac{1}{2} + \frac{1}{4}$
Tataru [Tat01a, Tat01b]	$\partial \mathbf{g} \in L^2_t L^\infty_x$	$\frac{1}{6}$ loss	$s > \frac{n}{2} + \frac{1}{2} + \frac{1}{6}$
Klainerman-Rodnianski [KR03]	$\square_{\mathbf{g}} \mathbf{g} \approx 0$	$\frac{2-\sqrt{3}}{2}$ loss	$s > \frac{n}{2} + \frac{1}{2} + \frac{2-\sqrt{3}}{2}$
Smith-Tataru [ST05]	$\square_{\mathbf{g}} \mathbf{g} \approx 0$	lossless	$s > \frac{n}{2} + \frac{1}{2}$

TABLE 1. A non-exhaustive historical overview of the local well-posedness of quasi-linear wave equations for $n \geq 3$, though one can find results concerning $n = 2$ among the references, and the result of Klainerman-Rodnianski [KR03] works only with $n = 3$. Non-French speakers may find the Bahouri-Chemin result in the textbook [BCD11, Chapter 9].

2. THE BOOTSTRAP ARGUMENT

To simplify presentation, we shall only work in dimensions $n = 3, 4, 5$ for the remainder of this article. We shall also recast the original bootstrap argument of Smith-Tataru in more modern language, following the expository article of Ifrim-Tataru [IT22] (see also [AIT24]).

Let $s > \frac{n}{2} + \frac{1}{2}$, then by standard scaling and finite speed of propagation arguments, it is enough to consider initial data $\phi[0] \in (H^s \times H^{s-1})_x(\mathbb{R}^n)$ which is sufficiently small, i.e.

$$\|\phi[0]\|_{(H^s \times H^{s-1})_x} \leq \varepsilon_3, \tag{2.1}$$

and prove local well-posedness for this class of data on unit time-scales $[0, 1]$. For each $\lambda > 0$, define

$$\phi^{(\lambda)} := \text{solution to (QNLW) with } \phi^{(\lambda)}[0] := \mathsf{P}_{\leq \lambda}\phi[0],$$
$$\psi^{(\lambda)} := \text{solution to (QNLW) with } \psi^{(\lambda)}[0] := \mathsf{P}_{\lambda}\phi[0].$$

Here it is convenient to bootstrap on λ, and as such we use a continuous Littlewood-Paley decomposition rather than a discrete one. For the remainder of the article though, we will follow the Smith-Tataru convention of using the later, though all these ideas can probably be consistently put together without too much difficulty. By persistence of regularity, these are smooth solutions. Furthermore, the initial data satisfy the bounds

$$\begin{aligned} \|\phi^{(\lambda)}[0]\|_{(H^s \times H^{s-1})_x} &\lesssim \varepsilon, \\ \|\psi^{(\lambda)}[0]\|_{(H^1 \times L^2)_x} &\lesssim \lambda^{1-s}\varepsilon. \end{aligned} \tag{2.2}$$

Theorem 2.1 (Estimates for smooth solutions from regularised data)**.** *The solutions $\phi^{(\lambda)}$ to (QNLW) satisfy the following properties:*

(a) Uniform bounds: *the solutions exist up to unit time-scale $[-1, 1]$, and satisfy the uniform bounds*

$$\|\phi^{(\lambda)}[t]\|_{L^\infty_t (H^s \times H^{s-1})_x} + \|\partial\phi^{(\lambda)}\|_{L^2_t C^{0,0+}_x} \lesssim \varepsilon, \tag{2.3}$$

(b) Difference bounds: *the linearised equations (LW) around $\phi^{(\lambda)}$ are well-posed in $(H^1 \times L^2)_x(\mathbb{R}^n)$, and satisfy the Strichartz estimates*

$$\|\partial\psi\|_{L^\infty_t L^2_x} + \|\langle\nabla_x\rangle^{-\frac{n-1}{2}-}\partial\psi\|_{L^2_t L^\infty_x} \lesssim \|\partial\psi[0]\|_{L^2_x}, \tag{2.4}$$

uniformly in λ.

It is easy to see that the result holds for $\lambda = 1$, where all norms, rough and smooth, are effectively the same and thus one can use Hughes-Kato-Marsden freely. To extend the range to all scales λ, we make the following bootstrap assumptions,

This needs some work and should not be taken literally.

(a) Uniform $(H^s \times H^{s-1})_x$-bounds,

$$\|\phi^{(\lambda)}[t]\|_{L_t^\infty (H^s\times H^{s-1})_x} \leq 2\varepsilon_2, \tag{2.5}$$

(b) Strichartz holds for the linear equation. probably one can make this more precise and substitute this with difference bounds, i.e. those on $\psi^{(\lambda)}$.

(c) Null geometry bounds,

$$\sup_{\theta,r} \|\phi^{(\lambda)}\|_{L_t^2 H^s_{x'_\theta}(\Sigma_{\theta,r})} + \|\partial_t \phi^{(\lambda)}\|_{L_t^2 H^{s-1}_{x'_\theta}(\Sigma_{\theta,r})} \leq 2\varepsilon_1. \tag{2.6}$$

It is not difficult to see that, once one improves on the bootstrap, perturbation theory allows one to push λ further and conclude the theorem. Then the scheme of Ifrim-Tataru [IT22] would allow one to conclude the main local well-posedness theorem for (QNLW). We leave these details to the interested reader.

3. Geometry and regularity of null hypersurfaces

Let u_θ be the solution the eikonal equation initialised at $t=-2$,

$$\begin{aligned} \mathbf{g}^{\mu\nu}\partial_\mu u \partial_\nu u &= 0, \\ u_{|t=-2} &= \theta\cdot x + 2. \end{aligned} \tag{3.1}$$

We denote the level sets $u_\theta = r$, which are null hypersurfaces with respect to the metric $\mathbf{g}$, by $\Sigma_{\theta,r}$. Write $x_\theta := x\cdot\theta$ and let x'_θ be coordinates on the orthogonal complement $\theta^\perp$, so that (x'_θ, x_θ) form an orthonormal coordinate system on $\mathbb{R}^n$. By the implicit function theorem, we can write these level sets as the graphs of functions $\tau_{\theta,r} \equiv \tau_{\theta,r}(t, x'_\theta)$. Thus we may parametrise $\Sigma_{\theta,r}$ by $(t, x'_\theta) \in [-2,2]\times\mathbb{R}^{n-1}$, and we can write

$$\Sigma_{\theta,r} := \{(t,x) \in [-2,2]\times\mathbb{R}^n : x_\theta - \tau_{\theta,r}(t,x'_\theta) = 0\}.$$

Example. In the flat case $\mathbf{g} = \mathbf{m}$, we have the explicit solution

$$u_\theta(t,x) = x\cdot\theta - t,$$

so that the null hypersurfaces are hyperplanes given by

$$\Sigma_{\theta,r} = \{(t,x) : x_\theta - t = r\},$$

where $r\in\mathbb{R}$ is a fixed constant (not radius! confusingly enough), and $\tau := t + r$.

3.1. **Regularity of null hypersurfaces.** It is convenient to define the Sobolev-type norms on the null hypersurfaces $\Sigma_{\theta,r}$ by

$$\|f\|_{\mathcal{H}^s_{t,x'_\theta}(\Sigma_{\theta,r})} := \sup_{\theta,r} \|f\|_{L_t^2 H^s_{x'_\theta}(\Sigma_{\theta,r})} + \|\partial_t f\|_{L_t^2 H^{s-1}_{x'_\theta}(\Sigma_{\theta,r})}.$$

Proposition 3.1 (Characteristic energy estimates)**.**

$$\|\mathbf{g}-\mathbf{m}\|_{\mathcal{H}^s(\Sigma_{\theta,r})} + \|\partial \mathbf{g}_{<\lambda}\|_{(\Sigma_{\theta,r})} + \lambda^{-1}\|\partial_x\partial\mathbf{g}_{<\lambda}\|_{\mathcal{H}^s(\Sigma_{\theta,r})} \lesssim \varepsilon_2. \tag{3.2}$$

Proposition 3.2 (Regularity of null hypersurfaces)**.**

$$\|d\tau_{\theta,r}(t) - dt\|_{C^{1,0+}_{x'_\theta}(\mathbb{R}^{n-1})} \lesssim \varepsilon_2 + \|\partial\mathbf{g}(t)\|_{C^{0,0+}_x(\mathbb{R}^n)}. \tag{3.3}$$

It will be convenient to introduce some notation on Σ. Let $\mathbf{V}$ be the vector field obtained by raising the indices of du,

$$\mathbf{V}^\alpha := \mathbf{g}^{\alpha\beta}\partial_\beta u.$$

Set

$$\sigma := \langle dt, \mathbf{V}\rangle = \mathbf{V}^0,$$

and

$$\mathbf{L} := \sigma^{-1}\mathbf{V}.$$

Thus $\mathbf{L}$ is the $\mathbf{g}$-normal field to Σ normalised so that $\mathbf{L}^0 = 1$. Set

$$\underline{\mathbf{L}}^\alpha := \mathbf{L}^\alpha + 2\mathbf{g}^{\alpha 0}\partial_0.$$

Then $\{\mathbf{L}, \underline{\mathbf{L}}\}$ form a null frame.

3.2. Geometry of light cones.

Proposition 3.3 (Angle of null generators). *Let $\theta, \omega \in \mathbb{S}^{n-1}$, then*

$$\mathbf{L}_\theta - \mathbf{L}_\omega = (\theta - \omega) + o(|\theta - \omega|). \tag{3.4}$$

Also,

$$\langle \mathbf{L}_\theta, \mathbf{L}_\omega \rangle_{\mathbf{g}} = -\frac{1}{2}|\theta - \omega|^2 + o(|\theta - \omega|^2). \tag{3.5}$$

Proposition 3.4 (Separation of null geodesics). *Let $\theta, \omega \in \mathbb{S}^{n-1}$, and fix $(t_1, x_1) \in [-2, 2] \times \mathbb{R}^n$. Denote γ_θ and γ_ω the null geodesics with data*

$$\gamma_\theta(t_1) = \gamma_\omega(t_1) = x_1, \qquad \dot{\gamma}_\theta(t_1) \parallel \theta, \qquad \dot{\gamma}_\omega(t_1) \parallel \omega.$$

Then

$$\gamma_\theta(t) - \gamma_\omega(t) = (t - t_1)(\theta - \omega) + o(|t - t_1| \cdot |\theta - \omega|). \tag{3.6}$$

4. Paradifferential decomposition

Decomposing the left-hand side via the Littlewood-Paley trichotomy, we schematically write the low-high, high-low, and high-high interactions of $\Box_{\mathbf{g}(\phi)}$ as

$$\Box_{\mathbf{g}(\phi)}\phi \approx \Box_{\mathbf{g}(\phi)_{<\lambda}}\phi_\lambda + \mathbf{g}(\phi)_\lambda \cdot \partial\partial\phi_{<\lambda} + \sum_{\mu \geq \lambda} \mathbf{g}(\phi)_\mu \cdot \partial\partial\phi_\mu.$$

Placing the latter two on the right-hand side, we obtain the paradiagonal formulation of the equation (QNLW)

$$\Box_{\mathbf{g}(\phi)_{<\lambda}}\phi_\lambda \approx \mathbf{g}(\phi)_\lambda \cdot \partial\partial\phi_{<\lambda} + \mathbf{g}(\phi)_\lambda \cdot \partial\partial\phi_{<\lambda} + \sum_{\mu \geq \lambda} \mathbf{g}(\phi)_\mu \cdot \partial\partial\phi_\mu + \mathcal{N}(\phi)(\partial\phi, \partial\phi)_\lambda. \tag{4.1}$$

A similar decomposition can be made for the linear wave counterpart,

$$\begin{aligned} \Box_{\mathbf{g}(\phi)}\psi &= 0, \\ (\psi, \partial_t\psi)_{|t=0} &= (\psi_0, \psi_1). \end{aligned} \tag{LW}$$

Thus, the main non-perturbative part of the argument resides in analysing the linear paradifferential equation,

$$\begin{aligned} \Box_{\mathbf{g}(\phi)_{<\lambda}} \mathsf{P}_\lambda\psi &= 0, \\ (\psi, \partial_t\psi)_{|t=0} &= (\mathsf{P}_\lambda\psi_0, \mathsf{P}_\lambda\psi_1). \end{aligned} \tag{PLW}$$

Our strategy will consist of constructing a suitable approximate solution to this initial data problem, and proving dispersive estimates hold for the approximation.

Proposition 4.1 (Existence of a parametrix). *Assuming the bootstrap, for each $\phi[0] \in (H^1 \times L^2)_x(\mathbb{R}^n)$, there exists a family of smooth functions $\psi^\lambda \in C^\infty([-2, 2] \times \mathbb{R}^n)$ which satisfy*

(a) frequency localisation,

$$\operatorname{supp} \widehat{\psi^\lambda(t)} \subseteq \{\xi \in \mathbb{R}^n : |\xi| \sim \lambda\}, \tag{4.2}$$

(b) initial data matching,

$$\psi^\lambda[-2] = \mathsf{P}_\lambda\phi[0], \tag{4.3}$$

(c) small error,

$$\|\Box_{\mathbf{g}_{<\lambda}}\psi^\lambda\|_{L^1_t L^2_x} \lesssim \varepsilon_0 \|\phi[0]\|_{H^1_x \times L^2_x}, \tag{4.4}$$

(d) Strichartz-type estimates for $\sigma > \frac{n-1}{2}$,

$$\|\psi^\lambda\|_{L^2_t L^\infty_x} \lesssim \varepsilon_0^{-\frac{1}{2}} \lambda^{\sigma - 1} \|\phi[0]\|_{H^1_x \times L^2_x}, \tag{4.5}$$

With the approximate solution at hand, we can analytically set-up the iteration scheme for solving (LW) as follows,

- approximate solution to the homogeneous problem (PLW),

$$\mathcal{S}^{\text{appr}}(t)\phi[0] := \sum_\lambda \psi^\lambda$$

- approximate solution to the inhomogeneous problem with zero initial data,

$$(\Box_{\mathbf{g}}^{-1})^{\mathrm{appr}} F := \int_0^t \mathcal{S}^{\mathrm{appr}}(t) \begin{pmatrix} 0 \\ F(s) \end{pmatrix} ds,$$

It is a standard argument to replace the $L^1_t L^2_x$-error bound to an $L^2_{t,x}$-error bound. This will be convenient for estimating terms $\partial \mathbf{g} \partial \psi$ in $L^1_t L^2_x$, placing the metric in Strichartz and the solution in energy norm. Then the parametrix error bounds implies that the linear map

$$\begin{aligned} L^2_{t,x}([0,1] \times \mathbb{R}^n) &\longrightarrow L^2_{t,x}([0,1] \times \mathbb{R}^n), \\ F &\longmapsto \left(\Box_{\mathbf{g}} \, (\Box_{\mathbf{g}}^{-1})^{\mathrm{appr}} - \mathrm{Id} \right) F \end{aligned} \tag{4.6}$$

is a contraction. In particular, each time we feed in an approximate solution to (LW), and solve away the error on the right, we gain. Thus, we can write the exact solution to (LW) as

$$\phi = \mathcal{S}^{\mathrm{appr}}(t)\phi[0] + (\Box_{\mathbf{g}}^{-1})^{\mathrm{appr}} F, \tag{4.7}$$

where F is some power series expansion which kills the error completely, and can easily be checked to satisfy

$$\|F\|_{L^2_{t,x}} \lesssim \varepsilon_0 \|\phi[0]\|_{H^1_x \times L^2_x}.$$

In this form, it is easy to pass the Strichartz bounds on the parametrix to the exact solution, with energy norm on the right-hand side. After suitable commutations, one can do the same for higher regularity norms.

5. Wave packet parametrix

To prove Strichartz estimates for the linear wave equation, we construct a *wave packet parametrix*, that is, a superposition of wave packets which form an approximate solution the initial data problem. Given a null geodesic $\gamma(t)$ contained in a null hypersurface $\Sigma_{\theta,r} = \{x_\theta - \tau_{\theta,r} = 0\}$, we define a *wave packet* $\mathfrak{w}$ localised around γ at scale $\lambda \in 2^{\mathbb{N}}$ to be a function of the form

$$\mathfrak{w} := (\varepsilon_0 \lambda)^{\frac{1}{2}\frac{n-1}{2}} \lambda^{\frac{1}{2}-1-1} \mathsf{T}_{<\lambda} \left(w \, \delta(u_{\theta,r}) \right),$$

where δ is the Dirac mass at zero, and thus $\delta(u_{\theta,r})$ is a measure[4] supported on the null hypersurface $\Sigma_{\theta,r}$ which takes the form

$$\langle \delta(u_\theta), \varphi \rangle := \int_{[-2,2]} \int_{\mathbb{R}^{n-1}} \varphi(t, x'_\theta, \tau_{\theta,r}) \, dx'_\theta \, dt \qquad \text{for } \varphi \in C^\infty_c((-2,2) \times \mathbb{R}^n),$$

and w is a smooth bump function on $\Sigma_{\theta,r}$ localised at scale $(\varepsilon_0\lambda)^{-\frac{1}{2}}$ about the null geodesic γ, i.e.

$$w(t, x'_\theta) = w_0\big((\varepsilon_0\lambda)^{\frac{1}{2}}(x'_\theta - \gamma'_\theta(t))\big), \qquad w_0 \in C^\infty_c(|x'| \leq 1),$$

and finally $\mathsf{T}_{<\lambda}$ is a mollification to spatial scales $\triangle x \sim \frac{1}{\lambda}$, taking it with kernel $\chi_{<\lambda}(x) := \lambda^n \chi(\lambda x)$ for some compactly-supported cut-off $\chi \in C^\infty_c(|x| \leq \frac{1}{2000})$; morally, one should think of this as a localisation to frequency $|\xi| \lesssim \lambda$.

The parameter $\varepsilon_0 \ll 1$ shall be chosen such that the error as an approximate solution $\Box_{\mathbf{g}}\mathfrak{w}$ is small in $L^1_t L^2_x$, while the choice of amplitude will normalise the wave packet to be of size $O(1)$ in $(H^1 \times L^2)_x$; see Lemmas 5.4 and 5.2 respectively.

Given a function restricted to frequency λ for all t, we want to construct a wave packet resolution, i.e. show that it arises as a superposition of wave packets. To that end, decompose $\mathbb{R}^n$ into a parallel tiling of

[4] Alternatively, one can think of this as the pushforward of the Lebesgue measure on $[-2,2] \times \mathbb{R}^{n-1}$ under the embedding $(t, x') \mapsto (t, x', \tau)$. This convention is common in harmonic analysis in the context of the Fourier extension and restriction problems. Strictly speaking, this is not the same as *surface measure* as it is referred in [ST05].

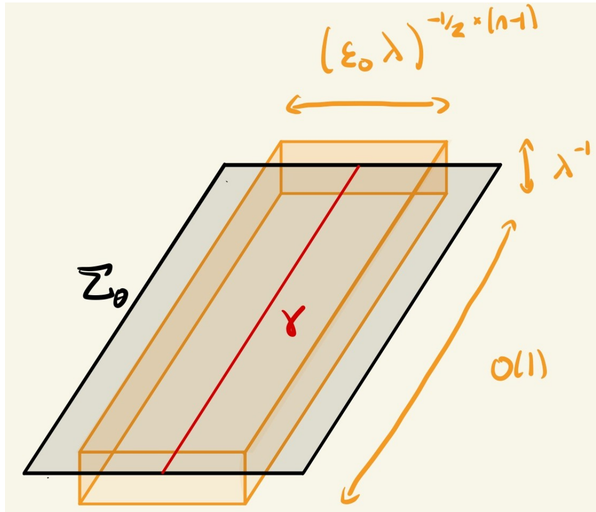

FIGURE 1. Given a null geodesic γ residing in a null hypersurface $\Sigma_{\theta,u}$, the corresponding wave packet is supported in a $1 \times \lambda^{-1} \times (\varepsilon_0\lambda)^{-\frac{1}{2}\times(n-1)}$-slab.

rectangles of dimensions $(4\varepsilon_0\lambda)^{-\frac{1}{2}\times(n-1)} \times (8\lambda)^{-1}$ in (x'_θ, x_θ)-coordinates. Then define

$$\begin{aligned}
R_{\theta,j} &:= \text{doubles of these rectangles at } t = -2, \\
\Sigma_{\theta,j} &:= \text{null hypersurface centered on } R_{\theta,j}, \\
\gamma_{\theta,j} &:= \text{null geodesic in } \Sigma_{\theta,j} \text{ through center of } R_{\theta,j}, \\
\mathtt{T}_{\theta,j} &:= (32\lambda)^{-1}\text{-neighborhood of } \Sigma_{\theta,j} \cap \{|x'_\theta - \gamma_{\theta,j}(t)| \leq (\varepsilon_0\lambda)^{-\frac{1}{2}}\}.
\end{aligned}$$

We shall refer to the space-time regions $\mathtt{T}_{\theta,j}$ as slabs; by construction, wave packets are supported in slabs. These slabs satisfy a finite-overlap condition; indeed, those associated to different null hypersurfaces are disjoint, while those associated to the same null hypersurface have finite overlap in the x'_θ-variable. Furthermore, we consider slabs with angles θ taken from

$$\Omega := \text{maximal collection of } \left(\tfrac{\lambda}{\varepsilon_0}\right)^{\frac{n-1}{2}}\text{-many unit vectors } \theta \in \mathbb{S}^{n-1} \text{ separated by at least } \left(\tfrac{\varepsilon_0}{\lambda}\right)^{\frac{1}{2}}.$$

Proposition 5.1 (Existence of wave packet parametrix)**.** *Let* $(\phi_0, \phi_1) \in (H^1 \times L^2)_x(\mathbb{R}^n)$ *be initial data. Then, in dimensions* $n = 2, 3, 4, 5$, *there exists a superposition of wave packets*

$$\phi := \sum_{\theta,j} a_{\theta,j}\, \mathfrak{w}^{\theta,j}$$

which is an approximate solution to the paralinearised initial data problem (PLW) in the sense that

(a) it matches the initial data at $t = -2$,

$$\mathsf{P}_\lambda\phi[-2] = (\mathsf{P}_\lambda\phi_0, \mathsf{P}_\lambda\phi_1) \tag{5.1}$$

(b) the size of the coefficients is comparable to the size of the initial data,

$$\Big(\sum_{\theta,j} |a_{\theta,j}|^2\Big)^{\frac{1}{2}} \lesssim \|\phi[0]\|_{(H^1\times L^2)_x}. \tag{5.2}$$

(c) the energy estimate holds,

$$\|\partial\mathsf{P}_\lambda\phi\|_{L^\infty_t L^2_x} \lesssim \Big(\sum_{\theta,j} |a_{\theta,j}|^2\Big)^{\frac{1}{2}} \tag{5.3}$$

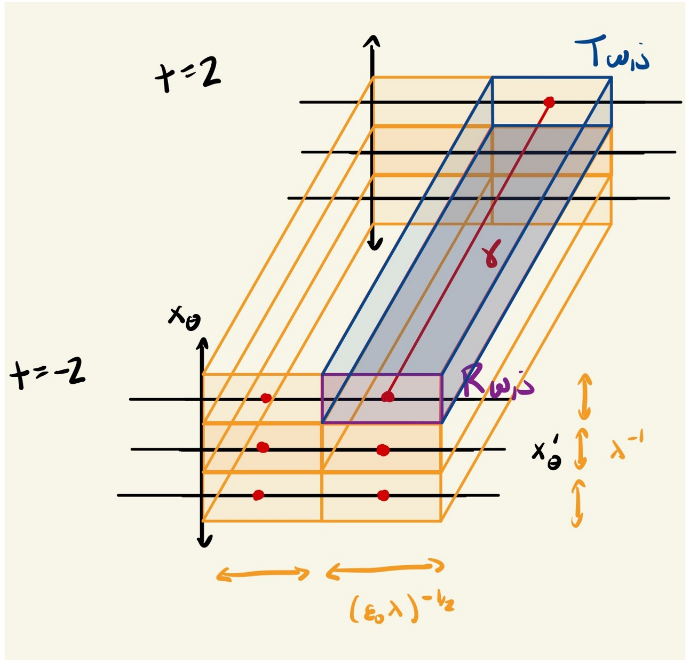

FIGURE 2. Given a fixed angle θ, we cover $[-2,2]\times\mathbb{R}^n$ by slabs localised along null geodesics $\gamma_{\theta,j}$ with dimensions $(\varepsilon_0\lambda)^{-\frac{1}{2}\times(n-1)}\times\lambda^{-1}$.

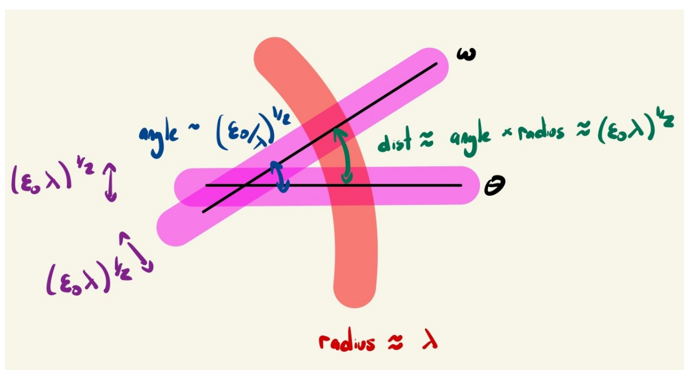

FIGURE 3. In frequency space, the wave packets at time $t=-2$ are effectively localised to $(\varepsilon_0\lambda)^{\frac{1}{2}}$-neighborhoods of the angles θ. These angles are separated by at least $(\frac{\varepsilon_0}{\lambda})^{\frac{1}{2}}$, so when restricted to the dyadic shell $|\xi|\sim\lambda$, there is $O(1)$-overlap between wave packets.

(d) *the error on the right-hand side is small,*

$$\|\Box_{\mathbf{g}_{<\lambda}}\phi_\lambda\|_{L^1_tL^2_x}\lesssim\varepsilon_0\left(\sum_{\theta,j}|a_{\theta,j}|^2\right)^{\frac{1}{2}}. \tag{5.4}$$

5.1. **Properties of wave packets.** As a first step, we prove that a singular wave packet is an approximate solution, in the sense that it satisfies the energy bound and error estimate.

Lemma 5.2 (Energy estimate for $\mathfrak{w}$). *Wave packets have $O(1)$-energy,*

$$\|\partial P_\lambda\mathfrak{w}\|_{L^\infty_tL^2_x}\lesssim 1. \tag{5.5}$$

Proof for $\partial=\partial_x$. The estimate for the full space-time gradient will follow from the usual energy estimate once one has the error estimate (5.9), though it is instructive to see how to read off the result for the spatial derivatives from the construction. Roughly speaking[5],

$$\text{support of } \mathsf{T}_{<\lambda}(w\delta(u_{\theta,r})) \approx 1 \times \lambda^{-1} \times (\varepsilon_0\lambda)^{-\frac{1}{2}\times(n-1)},$$
$$\text{amplitude of } \mathsf{T}_{<\lambda}(w\delta(u_{\theta,r})) \approx \lambda.$$

The former is fairly clear; to say a word about the latter, $\delta(x_\theta - \tau)$ is a unit point mass, while the mollification spreads it out in the θ-direction to scale λ^{-1}, so dimensional analysis tells us the resulting amplitude is λ. Using the usual heuristic that $\partial_x \approx \lambda$ at frequencies comparable to λ, we arrive at

$$\|\partial_x \mathsf{P}_\lambda \mathfrak{w}\|_{L^\infty_t L^2_x} \lessapprox \lambda \cdot (\varepsilon_0\lambda)^{\frac{1}{2}\frac{n-1}{2}} \lambda^{\frac{1}{2}-1-1} |\text{amplitude}| \cdot |\text{support}|^{\frac{1}{2}} \lessapprox 1.$$

This gives (5.5), modulo time-derivative control. □

Lemma 5.3 (Wave packet error decomposition)**.** *Let $\mathfrak{w}$ be a wave packet, then the error decomposes as*

$$\Box_{\mathbf{g}_{<\lambda}} \mathsf{P}_\lambda \mathfrak{w} = \mathcal{L}(\partial \mathbf{g}, \partial \widetilde{\mathsf{P}_\lambda} \widetilde{\mathfrak{w}}) + (\varepsilon_0\lambda)^{\frac{1}{2}\frac{n-1}{2}} \lambda^{\frac{1}{2}-1-1} \mathsf{P}_\lambda \mathsf{T}_{<\lambda} \sum_{j=0,1,2} \psi_j \delta^{(j)}(x_\theta - \tau_{\theta,r}), \tag{5.6}$$

where $\mathcal{L}$ is a bilinear form which obeys Hölder's inequality, $\widetilde{\mathsf{P}_\lambda}$ is a fattened frequency projection, $\widetilde{\mathfrak{w}}$ is another wave packet, $\delta^{(j)}$ are derivatives of the Dirac mass, and $\psi_j(t, x')$ obey

$$\|\psi_j((\varepsilon_0\lambda)^{-\frac{1}{2}} x'_\theta)\|_{L^2_t H^{s-1}_{x'_\theta}} \lesssim \varepsilon_0 \lambda^{1-j}. \tag{5.7}$$

Corollary 5.4 (Error estimate for $\mathfrak{w}$)**.** *Each wave packet has small error,*

$$\|\Box_{\mathbf{g}_{<\lambda}} \mathsf{P}_\lambda \mathfrak{w}\|_{L^1_t L^2_x} \lesssim \varepsilon_0, \tag{5.8}$$
$$\|\Box_{\mathbf{g}_{<\lambda}} \mathsf{P}_\lambda \mathfrak{w}\|_{L^2_{t,x}} \lesssim \varepsilon_0. \tag{5.9}$$

Proof. Obviously (5.9) is stronger than (5.8), so we focus on proving an $L^2_{t,x}$-error estimate. For the first term on the right-hand side of (5.6), we place $\partial\mathbf{g}$ in $L^2_t L^\infty_x$, gaining smallness from our bootstrap assumption, and $\partial\widetilde{\mathfrak{w}}$ in $L^\infty_t L^2_x$, in which it is unit size by[6] (5.5), yielding

$$\|\mathcal{L}(\partial \mathbf{g}, \partial \widetilde{\mathsf{P}_\lambda} \widetilde{\mathfrak{w}})\|_{L^2_{t,x}} \lesssim \|\partial \mathbf{g}\|_{L^2_t L^\infty_x} \|\partial \widetilde{\mathfrak{w}}\|_{L^\infty_t L^2_x} \lesssim \varepsilon_2.$$

Taking $\varepsilon_2 \ll \varepsilon_0$ is an acceptable contribution towards (5.9).

For the second term on the right-hand side of (5.6), dimensional analysis yields

$$\text{support of } \mathsf{P}_\lambda \mathsf{T}_{<\lambda}(\psi_j \delta^{(j)}) \approx 1 \times \lambda^{-1} \times (\varepsilon_0\lambda)^{-\frac{1}{2}\times(n-1)},$$
$$\|\text{amplitude of } \mathsf{P}_\lambda \mathsf{T}_{<\lambda}(\psi_j \delta^{(j)})\|_{L^2_t} \lessapprox \|\text{amplitude of } \psi_j\|_{L^2_t} \cdot \lambda^{1+j} \lessapprox \varepsilon_0 \lambda^2.$$

Roughly speaking, $s - 1 > \frac{n-1}{2}$ under the assumptions of Theorem 1.1, we can use Sobolev embedding on $\mathbb{R}^{n-1}$ to read off the amplitude bounds using the scaled Sobolev estimate (5.7), while derivatives of the Dirac mass contribute amplitude λ. By dimensional analysis, we arrive at (5.9). □

Proof of error decomposition (5.6). Suppressing subscripts for clarity $u \equiv u_{\theta,r}$ and $\tau \equiv \tau_{\theta,r}$, we compute

$$\begin{aligned}
\Box_{\mathbf{g}_{<\lambda}} \mathsf{P}_\lambda \mathfrak{w} &= (\varepsilon_0\lambda)^{\frac{1}{2}\frac{n-1}{2}} \lambda^{-\frac{1}{2}+1} \Big([\Box_{\mathbf{g}_{<\lambda}}, \mathsf{P}_\lambda \mathsf{T}_{<\lambda}] + \mathsf{P}_\lambda \mathsf{T}_{<\lambda} \Box_{\mathbf{g}_{<\lambda}} \Big) w\, \delta(u) \\
&= (\varepsilon_0\lambda)^{\frac{1}{2}\frac{n-1}{2}} \lambda^{-\frac{1}{2}+1} [\Box_{\mathbf{g}_{<\lambda}}, \mathsf{P}_\lambda \mathsf{T}_{<\lambda}] w\, \delta(u) \\
&\quad + (\varepsilon_0\lambda)^{\frac{1}{2}\frac{n-1}{2}} \lambda^{-\frac{1}{2}+1} \mathsf{P}_\lambda \mathsf{T}_{<\lambda} \Big(\Box_{\mathbf{g}_{<\lambda}} w \cdot \delta(u) + 2\overline{\mathbf{g}}^{\alpha\beta}_{<\lambda} \partial_\alpha w \cdot \partial_\beta \delta(u) + w \cdot \Box_{\mathbf{g}_{<\lambda}} \delta(u) \Big) \\
&=: \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV},
\end{aligned}$$

commuting $\Box$ with the frequency projections and then applying the product rule. Here we have denoted $\overline{\mathbf{g}}^{\alpha\beta}_{<\lambda} = \frac{1}{2}(\mathbf{g}^{\alpha\beta}_{<\lambda} + \mathbf{g}^{\beta\alpha}_{<\lambda})$.

---

[5] Here, we write $\lessapprox$ and $\approx$ to mean that we aren't making any rigorous claims.

[6] To be perfectly rigorous, the energy estimate and the error estimate should be proved in conjunction.

We analyse each term. The main game one has the play is that derivatives of the metric $\partial \mathbf{g}_{<\lambda}$ are small thanks to the bootstrap assumption, while the low frequencies of the metric itself $\mathbf{g}_{<\lambda}$ is *a priori* $O(1)$, but we can replace it by the high frequencies $\mathbf{g}_{>\lambda} \lessapprox \frac{1}{\lambda}\partial \mathbf{g}_{>\lambda}$ plus the metric itself $\mathbf{g}$, which will be contracted by terms such that there is favourable cancellation.

Term I. Since the metric is cut-off to frequencies much lower than λ, the commutator clearly projects to frequencies $|\xi| \sim \lambda$. Thus, one can harmlessly insert fattened projections $\widetilde{\mathsf{P}}_\lambda \widetilde{\mathsf{T}}_{<\lambda}$ in front of the commutator. Furthermore, while two derivatives fall on the wave packet, standard commutator arguments[7] allow us to move one derivative onto the metric. In total, we can rewrite

$$\mathrm{I} = [\mathbf{g}^{\alpha\beta}_{<\lambda}, \mathsf{P}_\lambda \mathsf{T}_{<\lambda}]\partial_\alpha\partial_\beta \widetilde{\mathfrak{w}} = \mathcal{L}(\partial \mathbf{g}, \partial \widetilde{\mathfrak{w}})$$

for another wave packet $\widetilde{\mathfrak{w}}$ and some translation-invariant bilinear operator $\mathcal{L}(-,-)$.

Term II. We compute two derivatives of the bump function on $\mathbb{R}^{n-1}$ localised to the null geodesic γ,

$$\partial_\alpha\partial_\beta w = \begin{cases} O(\varepsilon_0\lambda) & \text{if two spatial derivatives,} \\ O((\varepsilon_0\lambda)^{\frac{1}{2}}\ddot{\gamma}) & \text{if two time derivatives,} \\ O(\varepsilon_0\lambda\dot{\gamma}) & \end{cases}$$

Since $\|\ddot{\gamma}\|_{L^2_t} \lesssim \varepsilon_1$, this is acceptable for (5.7) when placed into ψ_0.

Term III. We compute, schematically,

$$\begin{aligned} \overline{\mathbf{g}}_{<\lambda} \cdot \partial w \cdot \partial \delta(u) &= \overline{\mathbf{g}_{<\lambda}} \cdot \partial w \cdot \partial u \cdot \delta^{(1)}(u) \\ &= (\overline{\mathbf{g}}_{<\lambda})_{|\Sigma} \cdot \partial w \cdot \partial u \cdot \delta^{(1)}(u) + (\partial \overline{\mathbf{g}}_{<\lambda})_{|\Sigma} \cdot \partial w \cdot \partial u \cdot \delta(u). \end{aligned}$$

Putting the coefficients then in their respective boxes,

$$\begin{aligned} \psi_0 &:= \partial u \cdot \partial \overline{\mathbf{g}}_{<\lambda} \cdot \partial w, \\ \psi_1 &:= \partial u \cdot \overline{\mathbf{g}}_{<\lambda} \cdot \partial w. \end{aligned}$$

It is easy to see that the first coefficient is acceptable for (5.7). To see that the second is also acceptable, we decompose

$$\begin{aligned} \psi_1 = \overline{\mathbf{g}}^{\alpha\beta}_{<\lambda}\partial_\alpha u \partial_\beta w &= \overline{\mathbf{g}}^{\alpha\beta}\partial_\alpha u \partial_\beta w - \overline{\mathbf{g}}^{\alpha\beta}_{>\lambda}\partial_\alpha u \partial_\beta w \\ &= \left(\overline{\mathbf{g}}^{\alpha\beta} - \overline{\mathbf{g}}^{\alpha\beta}_{|x'=\gamma(t)}\right)\partial_\alpha u \partial_\beta w - \overline{\mathbf{g}}^{\alpha\beta}_{>\lambda}\partial_\alpha u \partial_\beta w. \end{aligned}$$

The first line is obvious, for the second line, we use the fact that $\mathbf{g} \cdot \partial u \cdot \partial w = 0$ when restricted to the null geodesic γ, since $(1, \dot{\gamma}) \propto \mathbf{g}^{\alpha\beta}\partial_\alpha u$. These terms can then easily be estimated by $\partial \mathbf{g}$.

Term IV. We compute

$$\Box_{\mathbf{g}_{<\lambda}} \delta(u) = \mathbf{g}^{\alpha\beta}_{<\lambda}\left(\partial_\alpha u\, \partial_\beta u\, \delta^{(2)}(u) + \partial_\alpha\partial_\beta \tau\, \delta^{(1)}(u)\right).$$

Applying the distributional product rule, we can rewrite the terms above schematically as

$$\begin{aligned} \mathbf{g}_{<\lambda} \cdot \partial\partial\tau \cdot \delta^{(1)}(x_\theta - \tau) &= (\partial \mathbf{g})_{|\Sigma} \cdot \partial\partial\tau \cdot \delta(x_\theta - \tau) \\ \mathbf{g}_{<\lambda} \cdot (\partial u)^2 \cdot \delta^{(2)}(x_\theta - \tau) &= (\partial\partial \mathbf{g}_{<\lambda})_{|\Sigma} \cdot (\partial u)^2 \cdot \delta(x_\theta - \tau) \\ &\quad + 2(\partial \mathbf{g}_{<\lambda})_{|\Sigma} \cdot (\partial u)^2 \cdot \delta^{(1)}(x_\theta - \tau) \\ &\quad + (\mathbf{g}_{>\lambda})_{|\Sigma}(\partial u)^2 \cdot \delta^{(2)}(x_\theta - \tau) \end{aligned}$$

Here, the derivatives on the metric are in the θ-direction; observe that ∂u depends only on t and x'_θ so no x^θ-derivatives fall on these terms. In the last line, we have freely replaced $\mathbf{g}_{<\lambda}(\partial u)^2$ by $\mathbf{g}_{>\lambda}(\partial u)^2$, since u is optical, i.e. $\mathbf{g}^{\alpha\beta}\partial_\alpha u \partial_\beta u = 0$.

[7] In a word, the principal symbol of the commutator is given by the Poisson bracket, so one can, to leading order, write $[\mathbf{g}(x), \chi(\nabla/\lambda)] \approx \{\mathbf{g}(x), \chi(\xi/\lambda)\} \approx \partial_x \mathbf{g} \cdot \partial_\xi \chi(\xi/\lambda) \approx \frac{1}{\lambda}\partial_x \mathbf{g}$.

Putting the coefficients then in their respective boxes,

$$\psi_0 := w \cdot \left(\partial\partial \mathbf{g}_{<\lambda} \cdot (\partial u)^2 + \partial \mathbf{g}_{<\lambda} \cdot \partial\partial\tau \right),$$
$$\psi_1 := w \cdot \left(2\partial \mathbf{g}_{<\lambda} \cdot (\partial u)^2 - \mathbf{g}_{<\lambda} \cdot \partial\partial\tau \right),$$
$$\psi_2 := w \cdot (\mathbf{g}_{>\lambda})_{|\Sigma} \cdot (\partial u)^2,$$

the error bound (5.7) follow from characteristic energy estimates.

□

Remark. It is instructive to compare the computation above against the flat case. Writing $u := x_\theta - t$ and $\underline{u}_\theta := x_\theta + t$, we write in coordinates $\Box = -4\partial\underline{\partial} + \Delta_{x'_\theta}$, so

$$\Box\delta(u) = \partial_\alpha u \partial^\alpha u\, \delta^{(2)}(u) + \Box u\, \delta^{(1)}(u) = 0.$$

Thus, the measures $\delta(u)$ on null hypersurfaces are exact solutions to the linear wave equation on flat backgrounds.

5.2. **Orthogonality of wave packets.** We want to show that, given any superposition of wave packets, the energy estimate (5.3) and the error estimate (5.4) hold. In view of the corresponding estimate for a single wave packet, we want to show that the interactions between wave packets are negligible, i.e. they are almost orthogonal. In place of the $L^\infty_t L^2_x$-bound (5.3), it is convenient to prove the weaker $L^2_{t,x}$-bound,

$$\|\partial \mathsf{P}_\lambda \phi\|_{L^2_{t,x}} \lesssim \Big(\sum_{\theta,j} |a_{\theta,j}|^2 \Big)^{\frac{1}{2}}. \tag{5.10}$$

Then (5.10) and (5.4) imply (5.3) by the energy estimate.

We prove (5.10) by showing a fixed-time $t =$ const orthogonality estimate. This will be a consequence of the following two "properties" of wave packets which are "morally" true,

- Parallel wave packets, i.e. fixing θ, have finitely-overlapping supports $\mathtt{T}^t_{\theta,j}$ in physical space, see Figure 2
- Non-parallel wave packets, i.e. distinct $\theta, \omega \in \Omega$, at scale λ are effectively localised in frequency space to the dual rectangles $(\mathtt{T}^t_{\theta,j})^*$ and $(\mathtt{T}^t_{\omega,k})^*$, which are contained in $(\varepsilon_0\lambda)^{\frac{1}{2}}$-neighborhoods of θ and ω respectively. Since we have chosen angles separated by at least $(\frac{\varepsilon_0}{\lambda})^{\frac{1}{2}}$, the supports are effectively disjoint on the dyadic shell $|\xi| \sim \lambda$, see Figure 3.

Evidently these hold in the flat case, we want to show these persist for the variable background. The picture one should keep in mind is as follows:

Lemma 5.5 (Dual tube frequency localisation)**.** *Let $n \leq 5$ and suppose $t \in [-2, 2]$ is a time such that $\|\partial \mathbf{g}(t)\|_{C^{0,0+}_x} \leq \varepsilon_0$, then there exists coordinates (y_θ, y'_θ) such that*

$$\|\mathsf{T}_{<\lambda}(\psi^{\theta,j} \cdot \delta(x_\theta - \tau_\theta))\|_{L^2 H^{\frac{n-1}{2}+}_{a,y'_\theta}(\mathbb{R}^n)} \lesssim \lambda^{\frac{1}{2}} \|\psi^{\theta,j}\|_{H^{\frac{n-1}{2}+}_{a,x'_\theta}(\mathbb{R}^{n-1})}. \tag{5.11}$$

Here the subscript $a = (\varepsilon_0\lambda)^{-\frac{1}{2}}$ denotes the scaled-Sobolev space

$$\|f(y)\|_{H^s_{a,y}(\mathbb{R}^{n-1})} := \|f(ay)\|_{H^s_y(\mathbb{R}^{n-1})}.$$

Proof. In the flat case, this is true by direct application of Bernstein's inequality in the θ-direction,

$$\left\| \mathsf{T}_{<\lambda}\left(\psi(x'_\theta)\, \delta(x_\theta - t) \right) \right\|_{L^2_{x_\theta} H^{\frac{n-1}{2}+}_{x'_\theta}(\mathbb{R}^n)} \lesssim \lambda^{\frac{1}{2}} \|\psi\|_{H^{\frac{n-1}{2}+}(\mathbb{R}^{n-1})}.$$

In the variable-coefficient case, since the derivative of the metric is $O(1)$, we know that the level hypersurfaces are C^2-regular, and, roughly speaking, thanks to (3.3), $x_\theta \approx \tau$, so the time t-sections of the tubes are contained in a rectangle of dimensions $(\varepsilon_0\lambda)^{-\frac{1}{2}\times(n-1)} \times \lambda^{-1}$. In particular,

$$\operatorname{supp} \mathcal{F}_x \mathsf{T}_{<\lambda}\left(\psi(x'_\theta)\, \delta(x_\theta - \tau) \right) \subseteq (\varepsilon_0\lambda)^{\frac{1}{2}\times(n-1)} \times \lambda \text{ dual rectangle.}$$

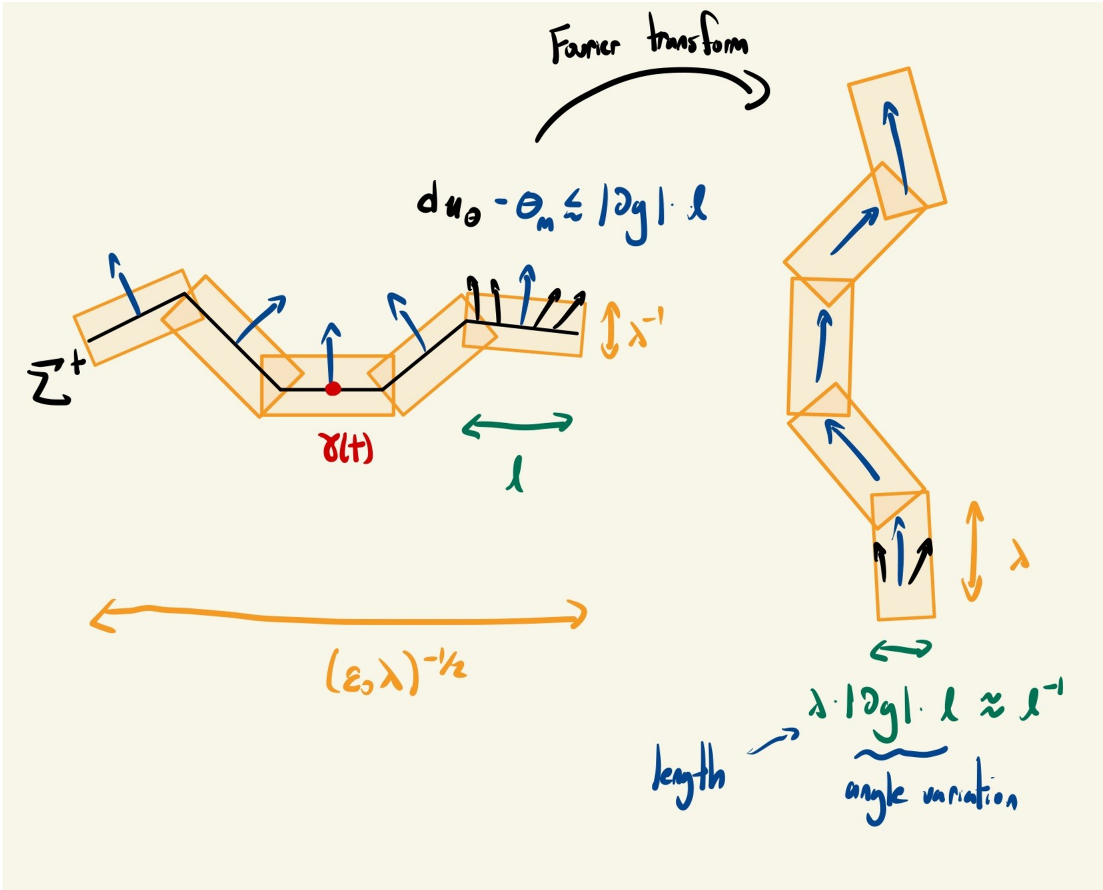

FIGURE 4. On a "bad" time slice, the conormal direction du_θ to a single wave packet can vary by a large amount on the total support, so to ensure that we are looking at appropriately localised Fourier supports, we partition the support in physical space into smaller cubes of length ℓ where the conormal direction does not vary much.

We can choose coordinates adapted to this rectangle, and take the Fourier transform in the $\lambda \mapsto \lambda^{-1}$-direction. By Plancharel's theorem and suitable rescaling, it will suffice to prove

$$\|\psi e^{i\eta\tau}\|_{H^{\frac{n-1}{2}+}(\mathbb{R}^{n-1})} \lesssim (1+|\eta|)^{100000}\|\psi\|_{H^{\frac{n-1}{2}+}(\mathbb{R}^{n-1})},$$

where, denoting η the dual Fourier variable to x_θ, we have used the standard identity

$$e^{i\eta\tau} = \mathcal{F}_{x_\theta}\left(\delta(x_\theta - \tau)\right).$$

When both derivatives fall on ψ, we are happy, however when both derivatives fall on the oscillation, we need to control

$$\|e^{i\eta\tau}\|_{C^{2,0+}} \lesssim \|\tau\|_{C^{2,0+}} \lesssim (1+|\eta|)^{2+}.$$

This is acceptable since we have $C^{1,0+}$-control of the metric. □

Remark. Here we used the famous inequality

$$\frac{n-1}{2} \leq 2 \qquad \text{if and only if } n \leq 5.$$

Lemma 5.6 (Orthogonality at "good" t). *Set*

$$\Phi := (\varepsilon_0\lambda)^{\frac{1}{2}\frac{n-1}{2}}\lambda^{\frac{1}{2}-1}\mathsf{P}_\lambda \mathsf{T}_{<\lambda}\sum_{\theta,j}\psi^{\theta,j}\,\delta(u_{\theta,j}).$$

Let $n \le 5$, and suppose $t \in [-2, 2]$ is a time such that $\|\partial \mathbf{g}(t)\|_{C_x^{0,0+}} \le \varepsilon_0$, then

$$\|\Phi\|_{L_x^2}^2 \lesssim \sum_{\theta,j} \left\| \psi^{\theta,j}((\varepsilon_0\lambda)^{-\frac{1}{2}} x) \right\|_{H^{\frac{n-1}{2}+}(\mathbb{R}^{n-1})}^2 \tag{5.12}$$

Proof. Decompose into packets with the same angle,

$$\begin{aligned} \Phi &:= \sum_{\theta\in\Omega} \Phi_\theta, \\ \Phi_\theta &:= \mathsf{P}_\lambda \sum_j \mathfrak{v}^{\theta,j}, \\ \mathfrak{v}^{\theta,j} &:= (\varepsilon_0\lambda)^{\frac{1}{2}\frac{n-1}{2}} \lambda^{\frac{1}{2}-1} \mathsf{T}_{<\lambda}\big(\psi^{\theta,j}\delta(u_{\theta,j})\big). \end{aligned}$$

Thanks to the finite-overlap property in physical space and the estimate (5.11), we have that the packets with the same angle are essentially orthogonal,

$$\|\Phi_\theta\|^2_{L^2_{y_\theta} H^{\frac{n-1}{2}+}_{y'_\theta}(\mathbb{R}^n)} \lesssim (\varepsilon_0\lambda)^{\frac{n-1}{2}} \sum_j \|\psi^{\theta,j}((\varepsilon_0\lambda)^{-\frac{1}{2}} x'_\theta)\|^2_{H^{\frac{n-1}{2}+}(\mathbb{R}^{n-1})}.$$

after rescaling. By Plancherel's theorem and frequency localisation, super-positions of packets with different angles Φ_θ are also effectively orthogonal, so we can conclude the result. □

Lemma 5.7 (Orthogonality at "bad" t)**.** *Set*

$$\Phi := (\varepsilon_0\lambda)^{\frac{1}{2}\frac{n-1}{2}} \lambda^{\frac{1}{2}-1} \mathsf{P}_\lambda \mathsf{T}_{<\lambda} \sum_{\theta,j} \psi^{\theta,j}\, \delta(u_{\theta,j}).$$

Suppose $t \in [-2, 2]$ is a time such that $\|\partial \mathbf{g}(t)\|_{C_x^{0,0+}} \ge \varepsilon_0$. Then

$$\|\Phi(t)\|_{L_x^2}^2 \lesssim \left(\frac{1}{\varepsilon_0} \|\partial \mathbf{g}(t)\|_{C_x^{0,0+}}\right)^{\frac{n-1}{2}} \sum_{\theta,j} \|\psi^{\theta,j}\|^2_{H^{\frac{n-1}{2}+}(\mathbb{R}^{n-1})}. \tag{5.13}$$

Proof. Suppose $\|\partial \mathbf{g}(t)\|_{C_x^{0,0+}} \ge \lambda$, then the bound follows from Cauchy-Schwartz in θ. Indeed, since there are $O((\frac{\lambda}{\varepsilon_0})^{\frac{n-1}{2}})$-many angles,

$$|\Phi|^2 \lesssim (\varepsilon_0\lambda)^{\frac{n-1}{2}} \lambda^{-1} \cdot \sum_\theta \left| \lambda^{\frac{n-1}{2}-\frac{1}{2}} \mathsf{P}_\lambda \sum_j \mathsf{T}_{<\lambda}\big(\psi^{\theta,j}\, \delta(u_{\theta,j})\big) \right|^2,$$

then integrating, using physical space separation, and applying Bernstein gives the result. Morally this should be the easy case, since the equation we are effectively studying is with the truncated metric $\mathbf{g}_{<\lambda}$, so contributions when the metric is larger than λ should be negligible.

Suppose then $\|\partial \mathbf{g}(t)\|_{C_x^{0,0+}} \le \lambda$. At the end of the day, we want to use the fact that the metric at fixed t can be "bad", but it is only bad on average. The issue with large metric is that the conormal direction du_θ can vary a lot on the total support of each wave packet. By the null regularity estimate (3.3), we know that the derivative of the conormal vector satisfies $\|\partial_x du\|_{C^{0+}} \lesssim \|\partial \mathbf{g}\|_{C^{0+}}$, so this Lipschitz control implies

$$\text{variation of conormal on region diamater } \ell \lesssim \|\partial \mathbf{g}\|_{C^{0+}} \cdot \ell.$$

Moving in ℓ-units in physical space corresponds to ℓ^{-1}-units in frequency space. On the other hand, the angle change can lead to frequencies curving in bad directions as one moves along the long direction in frequency space, see Figure 4. To ensure consistency, we set

$$\begin{aligned} \text{short length} \approx \text{long length} \cdot \text{angle} &\approx \lambda \cdot \|\partial \mathbf{g}\|_{C^{0+}} \cdot \ell \\ &\approx \ell^{-1}. \end{aligned}$$

This gives

$$\ell \approx \left(\lambda \|\partial \mathbf{g}\|_{C^{0+}}\right)^{-\frac{1}{2}}$$

as the optimal scale to partition the support of the wave packet in physical space such that the angle does not change as much. Accordingly, we form a partition of unity $\{\chi_m\}_m$ of $\Sigma^t_{\theta,j}$ adapted to cubes of length ℓ, and decompose

$$\psi^{\theta,j} = \sum_m \chi_m \psi^{\theta,j}.$$

Then the uncertainty principle analogous to (5.11) yields

$$\|\mathsf{T}_{<\lambda}\big(\chi_m \psi^{\theta,j}\cdot\delta(x_\theta-\tau)\big)\|_{L^2 H^{\frac{n-1}{2}}_{a,y'_\theta}{}^{+}(\mathbb{R}^n)} \lesssim \lambda^{\frac12}\|\psi^{\theta,j}\|_{H^{\frac{n-1}{2}+}_{a,x'_\theta}(\mathbb{R}^{n-1})}, \tag{5.14}$$

where, by construction, θ is such that

$$|\theta - du_\theta| \lesssim \|\partial\mathbf{g}\|_{C^{0+}}\cdot\ell \lesssim (\lambda\ell)^{-1} \qquad \text{uniformly on } \chi_m.$$

From here the proof proceeds similarly to the good time case, replacing ε_0 with $\|\partial\mathbf{g}(t)\|_{C^{0,0+}_x}$ in appropriate places. We leave this to the reader. $\square$

Proof of Proposition 5.1 (c) (energy bounds). It immediately follows from the fixed-time orthogonality estimates (5.12)-(5.13)

$$\|\partial_x \mathsf{P}_\lambda\phi(t)\|^2_{L^2_x} \lesssim \Big(1+\Big(\frac{1}{\varepsilon_0}\|\partial\mathbf{g}(t)\|_{C^{0,0+}_x}\Big)^{\frac{n-1}{2}}\Big)\sum_{\theta,j}|a_{\theta,j}|^2.$$

Integrating in time, we can control the right-hand side by the bootstrap $\|\partial\mathbf{g}\|_{L^2_tC^{0,0+}_x} \lesssim \varepsilon_0$ in dimensions $n=3,4,5$, which yields the energy bound (5.3) for the spatial derivatives. We leave the time derivatives as an exercise. $\square$

Proof of Proposition 5.1 (d) (error bounds). Summing the error decompositions (5.6) of each wave packet,

$$\Box_{\mathbf{g}_{<\lambda}}\mathsf{P}_\lambda\phi = \mathcal{L}\big(\partial\mathbf{g},\partial\widetilde{\mathsf{P}_\lambda\phi}\big) + (\varepsilon_0\lambda)^{\frac12\frac{n-1}{2}}\lambda^{\frac12-1-1}\mathsf{P}_\lambda\mathsf{T}_{<\lambda}\sum_{\theta,j}a_{\theta,j}\sum_{k=0,1,2}\psi^{\theta,j}_k\delta^{(k)}(x_\theta-\tau_{\theta,j}),$$

where $\widetilde{\phi}$ is another superposition of wave packets with the same coefficients. Hölder's inequality and the energy bound (5.3) tell us that the first term on the right is harmless,

$$\|\mathcal{L}(\partial\mathbf{g},\partial\widetilde{\mathsf{P}_\lambda\phi})\|_{L^1_tL^2_x} \lesssim \|\partial\mathbf{g}\|_{L^2_tL^\infty_x}\|\partial\widetilde{\phi}\|_{L^\infty_tL^2_x} \lesssim \varepsilon_0\Big(\sum_{\theta,j}|a_{\theta,j}|^2\Big)^{\frac12}.$$

Applying the fixed-time orthogonality estimates (5.12)-(5.13) to the remaining terms, setting

$$f(t) := (\varepsilon_0\lambda)^{\frac12\frac{n-1}{2}}\lambda^{\frac12-1-1}\mathsf{P}_\lambda\mathsf{T}_{<\lambda}\sum_{\theta,j}a_{\theta,j}\sum_{k=0,1,2}\psi^{\theta,j}_k\delta^{(k)}(x_\theta-\tau_{\theta,j})$$

then

$$\Big\|f(t)\Big\|^2_{L^2_x} \lesssim \Big(1+\Big(\frac{1}{\varepsilon_0}\|\partial\mathbf{g}(t)\|_{C^{0,0+}_x}\Big)^{\frac{n-1}{2}}\Big)\sum_{\theta,j}|a_{\theta,j}|^2\sum_{k=0,1,2}\lambda^{k-1}\|\psi^{\theta,j}_k((\varepsilon_0\lambda)^{-\frac12}x'_\theta)\|^2_{H^{\frac{n-1}{2}+}(\mathbb{R}^{n-1})}.$$

Applying Cauchy-Schwartz in time,

$$\begin{aligned}\|f\|^2_{L^1_tL^2_x} &\lesssim \Big(\int_{-2}^2 1+\Big(\frac{1}{\varepsilon_0}\|\partial\mathbf{g}(t)\|_{C^{0,0+}_x}\Big)^{\frac{n-1}{2}}dt\Big)\\ &\quad\times\left(\int_{-2}^2\sum_{\theta,j}|a_{\theta,j}|^2\sum_{k=0,1,2}\lambda^{2(k-1)}\|\psi^{\theta,j}_k((\varepsilon_0\lambda)^{-\frac12}x'_\theta)\|^2_{H^{\frac{n-1}{2}+}(\mathbb{R}^{n-1})}dt\right)\\ &\lesssim \varepsilon_0\sum_{\theta,j}|a_{\theta,j}|^2,\end{aligned}$$

again using the $L^2_tC^{0,0+}$-smallness of the metric and the error bounds (5.7). $\square$

Remark. Again we needed to use

$$\frac{n-1}{2} \leq 2 \qquad \text{if and only if } n \leq 5,$$

to ensure that the "bad" times are good on average.

5.3. **Matching wave packets to initial data.** It remains to show Proposition 5.1 (a)-(b); put loosely, any initial data $(\phi_0, \phi_1) \in (H^1 \times L^2)_x(\mathbb{R}^n)$ can be matched at time $t = -2$ to a superposition of wave packets.

Maximal collection of θ. We decompose

$$\phi[0] = \sum_{\theta \in \Omega} \phi^\theta[0],$$

where

$$\phi^\theta := \frac{1}{2}\left(\phi_0^\theta(x + t\theta) + u_0\right)$$

Fourier transform trick, u_0^ω compact support in frequency, take Fourier transform in x_θ, then extend periodically the Fourier transform with period $\lambda\theta$,

$$\widehat{\phi^\theta} = \sum_{k \in \mathbb{Z}}$$

Remark. Here we did not need to use the restriction on dimenions $n \leq 5$.

6. Dispersive estimates

The analysis in the previous sections tell us that the geometry of slabs is approximately that of Minkowski space. Thus, one can expect that the same harmonic analysis counting arguments used to prove Strichartz estimate hold in our setting.

Proposition 6.1 (Strichartz estimate for wave packet parametrix)**.** *Let $\mathcal{T}$ be the collection of λ-scale wave packets, then*

$$\Big\| \sum_{\mathtt{T} \in \mathcal{T}} a_{\mathtt{T}} \mathfrak{w}^{\mathtt{T}} \Big\|_{L_t^2 L_x^\infty} \lessapprox \lambda^{\frac{n-1}{2}} \lambda^{-1} \|a_{\mathtt{T}}\|_{\ell^2_{\mathtt{T}}}. \tag{6.1}$$

Here we use the standard harmonic analysis asymptotic notation $A \lessapprox B$, i.e. the inequality holds up to logarithmic losses in λ. For our purposes, it will suffice $A \lesssim_\varepsilon \lambda^\varepsilon B$ for any $\varepsilon > 0$. Our proof will proceed in two steps. First, we count the number of slabs containing any pair of points. This can be thought of as a dispersive decay-type estimate, considering data localised around a singular point, and estimating pointwise how the ensuing wave packets spread. Second, we discretise the Strichartz-type estimate and reduce it to counting.

6.1. **Dispersive decay.** Fix two points $\mathtt{P}_1 = (t_1, x_1)$ and $\mathtt{P}_2 = (t_2, x_2)$ in $[-2, 2] \times \mathbb{R}^n$. Our goal is to estimate

$$\#_\lambda(\mathtt{P}_1, \mathtt{P}_2) := \# \text{ of slabs at scale } \lambda \text{ containing } \mathtt{P}_1 \text{ and } \mathtt{P}_2.$$

Proposition 6.2 (Dispersive decay)**.** *The number of slabs at scale λ containing a pair of points $\mathtt{P}_1$ and $\mathtt{P}_2$ is bounded by*

$$\#_\lambda(\mathtt{P}_1, \mathtt{P}_2) \lesssim \left(\frac{\lambda}{\varepsilon_0}\right)^{\frac{n-1}{2}} \left(\lambda|t_1 - t_2|\right)^{-1}. \tag{6.2}$$

Remark. One can think of this as a dispersive estimate in the sense that, given a high frequency λ bump localised in a λ^{-1}-ball at the origin, the corresponding approximate solution to the wave equation consists of wave packets spreading out along the light cone. The dispersive decay then is captured by how many of these packets overlap at any given point in space-time.

To this end, we introduce $C_{\mathtt{P}_1} \subseteq [-2, 2] \times \mathbb{R}^n$ the forward light cone starting from $\mathtt{P}_1$, and let γ_θ be the null geodesic contained in Σ_θ starting from $\gamma_\theta(t_1) = x_1$. Denote $q_\theta := \gamma_\theta(t_2)$. Define

$$\text{dist}(x_2, C_{\mathtt{P}_1}^{t_2}) = \inf_{\theta \in \mathbb{S}^{n-1}} |x_2 - \gamma_\theta(t_2)|$$

which is the distance of x_2 to the t_2-slice of the cone, and set

$$\delta u(\mathtt{P}_1, \mathtt{P}_2) := \sup_{\theta \in \mathbb{S}^{n-1}} |u_\theta(\mathtt{P}_2) - u_\theta(\mathtt{P}_1)|.$$

These two are related as follows,

Lemma 6.3 (δu is signed distance to cone)**.** *The parameter δu is negative if $\mathtt{P}_2$ is inside the cone, positive in the exterior. Furthermore, $\delta u \approx \mathrm{dist}(x_2, C_1^{t_2})$.*

Proof. See [ST05, Lemma 9.1]. □

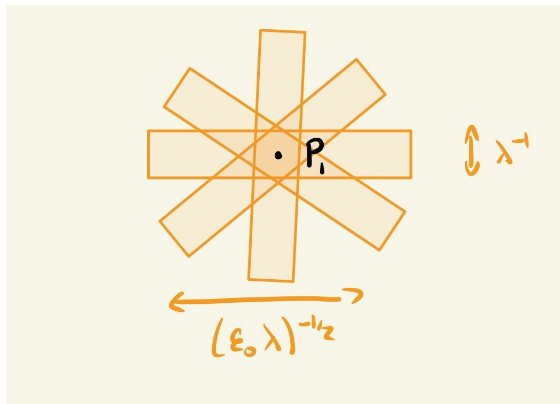

FIGURE 5. At time $t = 0$, we consider the wave packets of all possible angles θ containing $\mathtt{P}_1$. One should imagine that, as time evolves, these wave packets should spread outwards along the direction of their angles, and in space-time should cover the light cone $\mathtt{C}_{\mathtt{P}_1}$.

Lemma 6.4 (Characterisation of slabs containing two points)**.** *Define the set of angles*

$$A_\lambda := \left\{ \theta \in \mathbb{S}^{n-1} : |u_\theta(\mathtt{P}_2) - u_\theta(\mathtt{P}_1)| \leq \lambda^{-1} \text{ and } |\gamma_\theta(t_2) - x_2| \leq (\varepsilon_0 \lambda)^{-\frac{1}{2}} \right\},$$

Then

$$\#_\lambda(\mathtt{P}_1, \mathtt{P}_2) \lesssim \#\text{ of } (\tfrac{\varepsilon_0}{\lambda})^{\frac{1}{2}}\text{-balls covering } A_\lambda. \tag{6.3}$$

Proof. Recall that if a slab at scale λ in the direction θ contains both points $\mathtt{P}_1$ and $\mathtt{P}_2$, then the thicker slab centered on γ_θ of scale $\lambda/4$ must also contain $\mathtt{P}_1$ and $\mathtt{P}_2$. The result follows from the $(\frac{\varepsilon_0}{\lambda})^{\frac{1}{2}}$-angular separation of slabs at scale λ. □

We will need the following lemma, which essentially formalises the geometric intuition illustrated in Figures 7-8. The game one plays here is that if one slab of angle θ contains both points $\mathtt{P}_1$ and $\mathtt{P}_2$, then we want to study whether slabs of nearby angles ω also contain both points.

Lemma 6.5 (Making Figures 7-8 rigorous)**.** *Let $\theta \in A_\lambda$, then we can find a point (t_2, Q) on the hyperplane $u_\theta = u_\theta(\mathtt{P}_1)$, i.e. $u_\theta(t_2, Q) = u_\theta(t_1, x_1)$, such that $|Q - x_2| \leq 2\lambda^{-1}$. Then*

$$A_\lambda \subseteq \left\{ \omega \in \mathbb{S}^{n-1} : |u_\omega(t_2, Q) - u_\omega(\mathtt{P}_1)| \leq 3\lambda^{-1} \text{ and } |\gamma_\omega(t_2) - Q_2| \leq 3(\varepsilon_0\lambda)^{-\frac{1}{2}} \right\},$$

and

$$u_\omega(t_2, Q) - u_\omega(t_1, x_1) = -\frac{1}{2}|t_1 - t_2| \cdot |\omega - \theta|^2 + (\omega - \theta) \cdot \overrightarrow{\gamma_\theta(t_2)Q} + \text{error}, \tag{6.4}$$

where $\overrightarrow{\gamma_\theta(t_2)Q}$ is the projection of the straight-line from $\gamma_\theta(t_2)$ to Q onto the surface $\Sigma^t_{\theta, u_\theta(\mathtt{P}_1)}$.

Proof. The existence of such a Q is immediate. Let $s \mapsto \mu_\theta(s)$ be the path in Σ_θ^t from $\gamma_\theta(t_2)$ to Q obtained by projecting the straight line. Then

$$\dot{\mu}_\theta - \overrightarrow{\gamma_\theta(t_2)Q} = o(|\gamma_\theta(t_2) - Q|), \qquad \theta \cdot \overrightarrow{\gamma_\theta(t_2)Q} = o(|\gamma_\theta(t_2) - Q|).$$

To measure then the distance between $u_\omega(t_2, Q)$ and $u_\omega(\mathtt{P}_1)$ for different angle ω, we use the fundamental theorem of calculus along both the null geodesic γ_θ and the path μ_θ,

$$u_\theta(t_2, Q) - u_\theta(\mathtt{P}_1) = \int_{t_1}^{t_2} \sigma_\omega(s, \gamma_\omega(s)) \langle \mathbf{L}_\omega, \mathbf{L}_\theta \rangle_{\mathbf{g}} \, ds + \int_0^1 \sigma_\omega(t_2, \mu_\theta(s)) \langle \mathbf{L}_\omega, \dot{\mu}_\theta \rangle_{\mathbf{g}} \, ds.$$

Recall $\langle \mathbf{L}_\theta, \dot{\mu}_\theta \rangle = 0$ since $\dot{\mu}_\theta$ is tangent to Σ_θ. Using our estimates on the null generators (3.4)-(3.5), we conclude the result. $\square$

We now turn to the proof of Proposition 6.2.

Short-time case $|t_1 - t_2| < 2\lambda^{-1}$, Figure 6. In this case, the wave packets emanating from $\mathtt{P}_1$ have not had time to spread, so we are free to use the trivial bound

$$\#_\lambda(\mathtt{P}_1, \mathtt{P}_2) \lesssim \left(\frac{\lambda}{\varepsilon_0}\right)^{\frac{n-1}{2}} \lesssim \left(\frac{\lambda}{\varepsilon_0}\right)^{\frac{n-1}{2}} (\lambda|t_1 - t_2|)^{-1}.$$

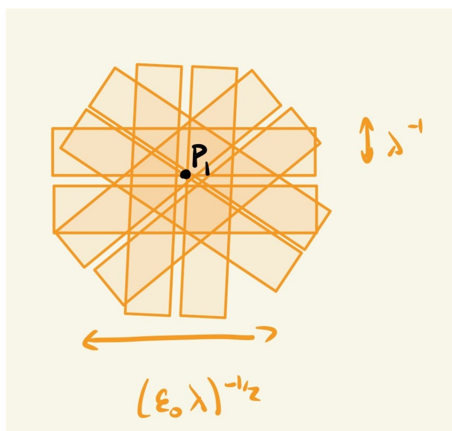

FIGURE 6. On time scale $|t_1 - t_2| < 2\lambda^{-1}$, the wave packets are still approximately maximally-overlapping, so we use the trivial bound.

Long-time $|t_1 - t_2| > 2\lambda^{-1}$ and close to cone case $|\delta u| \leq 4\lambda^{-1}$, Figure 7. In this case, if $\mathtt{P}_2$ is contained in one slab, then it is not difficult to rotate the slab along its long edge and continue to cover the point. The angle of aperture for such a cap of directions is approximately, as one learns in trigonometry,

$$\text{angle} \lesssim \frac{\text{opposite}}{\text{hypotenuse}} \lesssim \frac{(\varepsilon_0\lambda)^{-\frac{1}{2}}}{|t_1 - t_2|} \lesssim (\lambda|t_1 - t_2|)^{-\frac{1}{2}}.$$

Thus

$$A_\lambda \subseteq \{\omega \in \mathbb{S}^{n-1} : |\omega - \omega_0| \lesssim (\lambda|t_1 - t_2|)^{-\frac{1}{2}}\},$$

which is a radius $(\lambda|t_1 - t_2|)^{-\frac{1}{2}}$-cap on the unit sphere, so it can be covered by $O((\varepsilon_0|t_1 - t_2|)^{-\frac{n-1}{2}})$-many balls of radius $(\frac{\varepsilon_0}{\lambda})^{\frac{1}{2}}$. Thus, invoking (6.3), we have

$$\#_\lambda(\mathtt{P}_1, \mathtt{P}_2) \lesssim (\varepsilon_0|t_1 - t_2|)^{-\frac{n-1}{2}},$$

which for $n \geq 3$ is stronger than the desired bound.

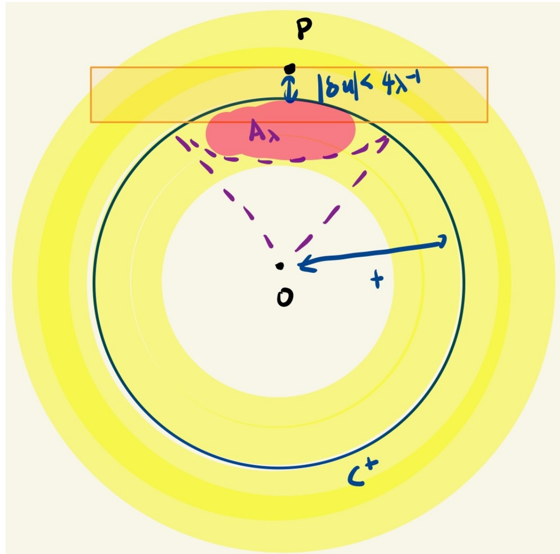

FIGURE 7. On time scale $|t_1 - t_2| > 2\lambda^{-1}$ and close to the cone $|\delta u| \leq 4\lambda^{-1}$, it is easy to rotate a tube containing the point to obtain another tube containing the point. This should form a set of angles which is like a cap of the sphere.

Long-time $|t_1 - t_2| > 2\lambda^{-1}$ *and far from cone case* $|\delta u| > 4\lambda^{-1}$. In this case, the wave packets have had some time to decohere. If $\mathtt{P}_2$ is in the interior, i.e. $\delta u < -4\lambda^{-1}$, then there can be no slabs containing both points by Huygen's principle, see Figure 8,

$$\#_\lambda(\mathtt{P}_1, \mathtt{P}_2) = 0.$$

We now turn to the interesting case $\delta u \geq 4\lambda^{-1}$, where $\mathtt{P}_2$ is in the exterior of the cone. For simplicity, let us assume $\mathtt{P}_1 = (0,0)$, and $\mathtt{P}_2 = (t, P)$, and the angle of P is θ. Using the flat case as intuition, the angle ω needs to be uniformly far away from $\theta = 0$ so that the wave packet may focus at P. There is some freedom to slide the wave packets, so A_λ should look like an annulus of angles.

To make this rigorous, we have by (6.4)

$$\delta u \lesssim |\gamma_\theta - Q|^2 |t|^{-1} \lesssim (\varepsilon_0 \lambda |t|)^{-1},$$

where we assume there is some slab intersection, so $|\gamma_\theta - Q| \lesssim 2(\varepsilon_0\lambda)^{-\frac{1}{2}}$. Since the max occurs at some ω_0, it follows that, for all $\theta \in A_\lambda$,

$$|\theta - \omega_0|^2 \approx |Q - \gamma_\theta(t)| \cdot |t|^{-1} \approx (\delta u)^{\frac{1}{2}} |t|^{-\frac{1}{2}}.$$

This allows us to cover A_λ by $O(1)$-many balls of radius $(\delta u)^{\frac{1}{2}}|t|^{-\frac{1}{2}}$. Thus we need to only consider A_λ intersected with such a ball. Using (6.4) with some facts about null geodesics and angles,

$$u_\omega(t, Q) - u_\theta(0,0) = (\omega - \theta') \cdot \overrightarrow{\gamma_\theta(t)Q} + \text{error}.$$

Since $\gamma_\theta(t)$ and P are in the same slab, $\gamma_\theta(t)$ is $(\varepsilon_0\lambda)^{-\frac{1}{2}}$-close to Q, so the thickness of the annulus is controlled by $(\frac{\varepsilon_0}{\lambda})^{\frac{1}{2}}$. After being quite lossy, we cover A_λ by

$$\#_\lambda(\mathtt{P}_1, \mathtt{P}_2) \lesssim \left(\frac{\lambda}{\varepsilon_0}\right)^{\frac{n-1}{2}} (\lambda|t_1 - t_2|)^{-1}$$

many balls of radius $(\frac{\varepsilon_0}{\lambda})^{\frac{1}{2}}$.

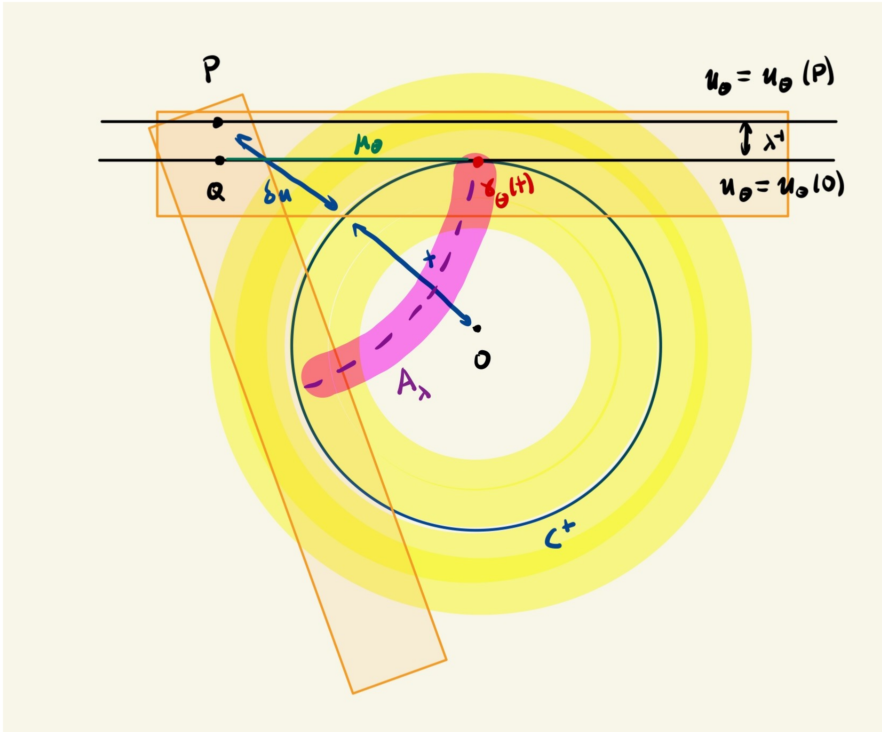

FIGURE 8. On time scale $|t_1 - t_2| > 2\lambda^{-1}$ and far outside the cone $\delta u \geq 4\lambda^{-1}$, for two tubes to intersect a point, the angles need to sufficiently far away so that the long sides have room to focus. The angles which do this should form an annulus. Inside the cone, the tubes do not focus at all, so the case $\delta u < -4\lambda^{-1}$ is trivial.

6.2. **Strichartz estimates.** Our proof of the dispersive estimate for the parametrix relies only on pointwise bounds on the wave packets, not their oscillation. That is, it will suffice to bound a superposition of wave packets pointwise by

$$\Big| \sum_{\mathtt{T}\in\mathcal{T}} a_{\mathtt{T}} \mathfrak{w}^{\mathtt{T}} \Big| \lesssim (\varepsilon_0\lambda)^{\frac{1}{2}\frac{n-1}{2}} \lambda^{-\frac{1}{2}} \sum_{\mathtt{T}\in\mathcal{T}} |a_{\mathtt{T}}| \cdot \mathbb{1}_{\mathtt{T}},$$

which holds since wave packets have amplitude $O((\varepsilon_0\lambda)^{\frac{1}{2}\frac{n-1}{2}}\lambda^{-\frac{1}{2}})$. Thus, to prove Strichartz for the wave packet parametrix (6.1), it suffices to prove that, for functions of the form

$$\phi := \sum_{\mathtt{T}\in\mathcal{T}} a_{\mathtt{T}} \mathbb{1}_{\mathtt{T}}$$

we have the Strichartz-type bound

$$\|\phi\|_{L^2_t L^\infty_x} \lessapprox \left(\frac{\lambda}{\varepsilon_0}\right)^{\frac{1}{2}\frac{n-1}{2}} \lambda^{-\frac{1}{2}} \|a_{\mathtt{T}}\|_{\ell^2_{\mathtt{T}}} \tag{6.5}$$

By scaling, we may assume $\|a_{\mathtt{T}}\| \leq 1$.

Step 1: discretising the problem. Dividing $[0,1]$ into $O(\lambda)$-many sub-intervals I_j of length $2\lambda^{-1}$, we can find points $\mathtt{P}_j = (t_j, x_j)$ nearly maximising $|\phi|$ on each space-time region $I_j \times \mathbb{R}^n$. It follows that

$$\|\phi\|_{L^2_t L^\infty_x} \lesssim \Big(\sum_j \int_{I_j} \|\phi(t)\|^2_{L^\infty_x} \, dt\Big)^{\frac{1}{2}} \lesssim \Big(\sum_j \lambda^{-1} |\phi(t_j, x_j)|^2\Big)^{\frac{1}{2}}.$$

After passing to an $O(\lambda)$ subset of points, we can choose t_j to be λ^{-1}-separated and such that the inequality above continues to hold. Next, we dyadically decompose the sum over slabs $\mathtt{T}$ with respect to the size $N^{-\frac{1}{2}}$ of the coefficients a_T, and similarly the sum over points $\mathtt{P}_j$ with respect to L the number of slabs containing them; we denote the number of such points by $m(L)$. It follows that

$$\begin{aligned}
\|\phi\|_{L^2_t L^\infty_x} &\lesssim \lambda^{-\frac{1}{2}} \sum_{\substack{L\in 2^{\mathbb{N}} \\ L\lesssim (\frac{\lambda}{\varepsilon_0})^{\frac{n-1}{2}}}} \Big(\sum_{\substack{j \\ \#\{\mathtt{T}\in\mathcal{T} : \mathtt{P}_j\in\mathtt{T}\}\sim L}} |\phi(t_j,x_j)|^2 \Big)^{\frac{1}{2}} \\
&\lesssim \lambda^{-\frac{1}{2}} \sum_{\substack{L\in 2^{\mathbb{N}} \\ L\lesssim (\frac{\lambda}{\varepsilon_0})^{\frac{n-1}{2}}}} \Big(\sum_{\substack{j \\ \#\{\mathtt{T}\in\mathcal{T} : \mathtt{P}_j\in\mathtt{T}\}\sim L}} \Big| \sum_{\substack{N\in 2^{\mathbb{N}} \\ N\lesssim (\frac{\lambda}{\varepsilon_0})^{\frac{n-1}{2}}}} \sum_{\substack{\mathtt{T}\in\mathcal{T} \\ |a_{\mathtt{T}}|\sim N^{-\frac{1}{2}}}} a_{\mathtt{T}} \mathbb{1}_{\mathtt{T}}(t_j,x_j) \Big|^2 \Big)^{\frac{1}{2}} \\
&\lesssim \lambda^{-\frac{1}{2}} \sum_{\substack{L\in 2^{\mathbb{N}} \\ L\lesssim (\frac{\lambda}{\varepsilon_0})^{\frac{n-1}{2}}}} \Big(\sum_{\substack{j \\ \#\{\mathtt{T}\in\mathcal{T} : \mathtt{P}_j\in\mathtt{T}\}\sim L}} \Big| \sum_{\substack{N\in 2^{\mathbb{N}} \\ N\lesssim (\frac{\lambda}{\varepsilon_0})^{\frac{n-1}{2}}}} L N^{-\frac{1}{2}} \Big|^2 \Big)^{\frac{1}{2}} \\
&\lesssim \lambda^{-\frac{1}{2}} \sum_{\substack{L\in 2^{\mathbb{N}} \\ L\lesssim (\frac{\lambda}{\varepsilon_0})^{\frac{n-1}{2}}}} \sum_{\substack{N\in 2^{\mathbb{N}} \\ N\lesssim (\frac{\lambda}{\varepsilon_0})^{\frac{n-1}{2}}}} \Big(m(L) L^2 N^{-1} \Big)^{\frac{1}{2}}
\end{aligned}$$

In the above, we needed only to sum over a finite range of scales, since each point lies in $O((\frac{\lambda}{\varepsilon_0})^{\frac{n-1}{2}})$-many slabs. This restricts us to a finite range of scales for L, and allows us to regard the contribution of slabs with small coefficients $|a_T| \lesssim (\frac{\varepsilon_0}{\lambda})^{\frac{n-1}{2}}$ as $O(1)$. The dyadic sums are thus harmless, as they only contribute a logarithm. Therefore the Strichartz-type estimate (6.5) is reduced to

$$m(L) N^{-1} L^2 \lessapprox \Big(\frac{\lambda}{\varepsilon_0}\Big)^{\frac{n-1}{2}}, \tag{6.6}$$

fixing N and L for the remainder of the argument.

Step 2: counting argument (or Fubini's theorem). We conclude with a counting argument. To summarise notation and introduce new ones,

$$\begin{aligned}
N^{-\frac{1}{2}} &:= \text{size of coefficient}, \\
L &:= \#\text{ of slabs containing a point}, \\
m &:= \#\text{ of points intersecting } L\text{-many slabs}, \\
k &:= \#\text{ of pairs } (i,j) \text{ for which } \mathtt{P}_i, \mathtt{P}_j \text{ lie in a common slab w/ multiplicity}, \\
n(\mathtt{T}) &:= \#\text{ of points in slab } \mathtt{T}, \\
\mathcal{T}_N &:= \text{set of slabs with coefficients of size } a_{\mathtt{T}} \sim N^{-\frac{1}{2}}.
\end{aligned}$$

We count k in two different ways[8]: we could either fix a slab $\mathtt{T}$ with at least two points, and count the approximately $\binom{n(\mathtt{T})}{2} \approx |n(\mathtt{T})|^2$ pairs; or we could fix a pair of points $(\mathtt{P}_i, \mathtt{P}_j)$ and count the number of slabs containing both points, i.e. $\#_\lambda(\mathtt{P}_i, \mathtt{P}_j)$,

$$\begin{aligned}
k &\approx \sum_{\substack{\mathtt{T}\in\mathcal{T}_N \\ n(\mathtt{T})\geq 2}} |n(\mathtt{T})|^2 \\
&\approx \sum_{\substack{1\leq i,j\leq m \\ i\neq j}} \#_\lambda(\mathtt{P}_i, \mathtt{P}_j).
\end{aligned} \tag{6.7}$$

[8]Otherwise known as Fubini's theorem.

On the other hand, by counting in two different ways, namely fixing a slab and counting the points, or fixing a point and counting the slabs,

$$\sum_{\mathtt{T}\in\mathcal{T}_N} n(\mathtt{T}) \sim m \cdot L. \tag{6.8}$$

Step 3: establishing a dichotomy. Partitioning the collection of slabs based on the size of their coefficients, $\sum_{\mathtt{T}} |a_{\mathtt{T}}|^2 \sim \sum_N \sum_{\mathtt{T}\in\mathcal{T}_N} N^{-1}$, we see from the pigeonhole principle that the number of slabs with coefficients of size $N^{-\frac{1}{2}}$ is of size

$$|\mathcal{T}_N| \lesssim N. \tag{6.9}$$

It follows from Cauchy-Schwartz in the slabs $\mathtt{T}$ and counting pairs (6.7) that

$$\sum_{\substack{\mathtt{T}\in\mathcal{T}_N\\ n(\mathtt{T})\geq 2}} n(\mathtt{T}) \lesssim N^{\frac{1}{2}}\Big(\sum_{\substack{\mathtt{T}\in\mathcal{T}_N\\ n(\mathtt{T})\geq 2}} |n(\mathtt{T})|^2\Big)^{\frac{1}{2}} \approx N^{\frac{1}{2}}k^{\frac{1}{2}}. \tag{6.10}$$

Our dichotomy will follow by considering two cases in the sum (6.8): one where the contribution from slabs with only a single point dominates, which allows us to estimate using the naive estimate on the number of slabs (6.9), or when the contribution from slabs with at least two points dominates, in which Cauchy-Schwartz (6.10) is more judicious,

(a) Either

$$\sum_{\substack{\mathtt{T}\in\mathcal{T}_N\\ n(\mathtt{T})\geq 2}} n(\mathtt{T}) \leq \sum_{\substack{\mathtt{T}\in\mathcal{T}_N\\ n(\mathtt{T})=1}} n(\mathtt{T})$$

which, combined with (6.8) and counting the number of slabs (6.9), implies

$$\begin{aligned} m \cdot L &\approx \sum_{\mathtt{T}\in\mathcal{T}_N} n(\mathtt{T}) \\ &\approx \sum_{\substack{\mathtt{T}\in\mathcal{T}_N\\ n(\mathtt{T})=1}} n(\mathtt{T}) \lesssim \sum_{\substack{\mathtt{T}\in\mathcal{T}_N\\ n(\mathtt{T})=1}} 1 \lesssim |\mathcal{T}_N| \lesssim N. \end{aligned}$$

(b) Or

$$\sum_{\substack{\mathtt{T}\in\mathcal{T}_N\\ n(\mathtt{T})\geq 2}} n(\mathtt{T}) \geq \sum_{\substack{\mathtt{T}\in\mathcal{T}_N\\ n(\mathtt{T})=1}} n(\mathtt{T}),$$

so by (6.8), Cauchy-Schwartz (6.10), we have

$$\begin{aligned} m \cdot L &\approx \sum_{\mathtt{T}\in\mathcal{T}_N} n(\mathtt{T}) \\ &\approx \sum_{\substack{\mathtt{T}\in\mathcal{T}_N\\ n(\mathtt{T})\geq 2}} n(\mathtt{T}) \lesssim N^{\frac{1}{2}}k^{\frac{1}{2}}. \end{aligned}$$

Rearranging gives

$$k \gtrsim m^2 N^{-1} L^2.$$

Step 4: concluding the argument. Proceeding from the conclusions of the dichotomy,

(a) $N \gtrsim mL$; in this case,

$$mN^{-1}L^2 \lesssim L,$$

so (6.6) follows from the trivial bound $O((\frac{\lambda}{\varepsilon_0})^{\frac{n-1}{2}})$ on the number of slabs.

(b) $k \gtrsim m^2 N^{-1} L^2$; then, counting k by the number of slabs intersecting two points (6.7), using the "dispersive decay" estimate (6.2), and also the λ^{-1}-separation of t_j, we have

$$\begin{aligned} mN^{-1}L^2 &\lesssim m^{-1}k \\ &\lesssim m^{-1} \sum_{\substack{1\le i,j\le m \\ i\neq j}} \#_\lambda(\mathtt{P}_i, \mathtt{P}_j) \\ &\lesssim \left(\frac{\lambda}{\varepsilon_0}\right)^{\frac{n-1}{2}} m^{-1} \sum_{\substack{1\le i,j\le m \\ i\neq j}} (\lambda|t_i - t_j|)^{-1} \lessapprox \left(\frac{\lambda}{\varepsilon_0}\right)^{\frac{n-1}{2}}, \end{aligned}$$

which concludes (6.5).

References

[AIT24] Albert Ai, Mihaela Ifrim, and Daniel Tataru. The time-like minimal surface equation in Minkowski space: Low regularity solutions. *Inventiones mathematicae*, 235(3):745–891, March 2024.

[BC99] Hajer Bahouri and Jean-Yves Chemin. équations d'ondes quasilinéaires et estimations de Strichartz. *American Journal of Mathematics*, 121(6):1337–1377, 1999.

[BCD11] Hajer Bahouri, Jean-Yves Chemin, and Raphaël Danchin. *Fourier Analysis and Nonlinear Partial Differential Equations*, volume 343 of *Grundlehren Der Mathematischen Wissenschaften*. Springer, Berlin, Heidelberg, 2011.

[Fou52] Y. Fourès-Bruhat. Théorème d'existence pour certains systèmes d'équations aux dérivées partielles non linéaires. *Acta Mathematica*, 88(none):141–225, January 1952.

[HKM77] Thomas J. R. Hughes, Tosio Kato, and Jerrold E. Marsden. Well-posed quasi-linear second-order hyperbolic systems with applications to nonlinear elastodynamics and general relativity. *Archive for Rational Mechanics and Analysis*, 63(3):273–294, September 1977.

[IT22] Mihaela Ifrim and Daniel Tataru. Local well-posedness for quasi-linear problems: A primer. *Bulletin of the American Mathematical Society*, 60(2):167–194, July 2022.

[KR03] S. Klainerman and I. Rodnianski. Improved local well-posedness for quasilinear wave equations in dimension three. *Duke Mathematical Journal*, 117(1):1–124, March 2003.

[Lin93] Hans Lindblad. A sharp counterexample to the local existence of low-regularity solutions to nonlinear wave equations. *Duke Mathematical Journal*, 72(2):503–539, November 1993.

[Lin96] Hans Lindblad. Counterexamples to local existence for semi-linear wave equations. *American Journal of Mathematics*, 118(1):1–16, 1996.

[Luk] Jonathan Luk. Introduction to nonlinear wave equations.

[ST05] Hart Smith and Daniel Tataru. Sharp local well-posedness results for the nonlinear wave equation. *Annals of Mathematics*, 162(1):291–366, July 2005.

[Tat01a] Daniel Tataru. Strichartz estimates for second order hyperbolic operators with nonsmooth coefficients, II. *American Journal of Mathematics*, 123(3):385–423, 2001.

[Tat01b] Daniel Tataru. Strichartz estimates for second order hyperbolic operators with nonsmooth coefficients III. *Journal of the American Mathematical Society*, 15(2):419–442, December 2001.