ANALYSIS & PDE

Volume 12 No. 4 2019

msp

Analysis & PDE

msp.org/apde

See inside back cover or msp.org/apde for submission instructions.

The subscription price for 2019 is US $310/year for the electronic version, and $520/year (+$60, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscriber address should be sent to MSP.

Analysis & PDE (ISSN 1948-206X electronic, 2157-5045 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

APDE peer review and production are managed by EditFlow® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

ANALYSIS AND PDE
Vol. 12, No. 4, 2019
dx.doi.org/10.2140/apde.2019.12.867

QUANTUM DYNAMICAL BOUNDS FOR ERGODIC POTENTIALS WITH UNDERLYING DYNAMICS OF ZERO TOPOLOGICAL ENTROPY

RUI HAN AND SVETLANA JITOMIRSKAYA

We show that positive Lyapunov exponents imply upper quantum dynamical bounds for Schrödinger operators $H_{f,\theta}u(n) = u(n+1) + u(n-1) + \phi(f^n\theta)u(n)$, where $\phi : \mathcal{M} \to \mathbb{R}$ is a piecewise Hölder function on a compact Riemannian manifold $\mathcal{M}$, and $f : \mathcal{M} \to \mathcal{M}$ is a uniquely ergodic volume-preserving map with zero topological entropy. As corollaries we also obtain localization-type statements for shifts and skew-shifts on higher-dimensional tori with arithmetic conditions on the parameters. These are the first localization-type results with precise arithmetic conditions for multifrequency quasiperiodic and skew-shift potentials.

1. Introduction

Positive Lyapunov exponents are generally viewed as a signature of localization. While it is known that they can coexist even with almost ballistic transport [Last 1996; del Rio et al. 1996], vanishing of certain dynamical exponents has been identified as a reasonable expected consequence of hyperbolicity of the corresponding transfer-matrix cocycle. Results in this direction were obtained in [Damanik and Tcheremchantsev 2007; 2008] for one-frequency trigonometric polynomials, and recently in [Jitomirskaya and Mavi 2017] for one-frequency quasiperiodic potentials under very mild assumptions on regularity of the sampling function. In this paper we identify a general property responsible for positive Lyapunov exponents implying vanishing of the dynamical quantities in the rather general case of underlying dynamics defined by volume-preserving maps of Riemannian manifolds with zero topological entropy, and under very minimal regularity assumptions. This work presents the first localization-type results that hold in such generality. We expect that positive topological entropy should also lead to vanishing of the dynamical quantities for a.e. (but not every!) phase, but this should be approached by completely different methods and will be explored in a future work.

Our general results allow us, in particular, to obtain localization-type statements for potentials defined by shifts and skew-shifts of higher-dimensional tori. Pure point spectrum with exponentially decaying eigenfunctions has been obtained for almost all multifrequency shifts in the regime of positive Lyapunov exponents in [Bourgain and Goldstein 2000] and for the skew-shift on $\mathbb{T}^2$ with a perturbative condition in [Bourgain et al. 2001], both very delicate results. While bounds on transport exponents are certainly weaker than dynamical localization that often (albeit not always [Jitomirskaya et al. 2003]) accompanies pure point spectrum [Bourgain and Jitomirskaya 2000], we note that pure point spectrum can be destroyed

MSC2010: 47B36, 81Q10.
Keywords: transport exponent, multifrequency quasiperiodic, skew-shift.

by generic rank-1 perturbations [Del Rio et al. 1994], while vanishing of the transport exponents is robust in this respect [Damanik and Tcheremchantsev 2007]. Finally, our results are the first ones for both of these families that hold under purely arithmetic conditions and the first nonperturbative ones for the skew-shift.

Let $(\mathcal{M}, g)$ be a d-dimensional compact (smooth) Riemannian manifold with a metric g. Let Vol_g be its Riemannian volume density; see (2-1). Let f be a uniquely ergodic volume-preserving map on $\mathcal{M}$, which means Vol_g is its unique invariant probability measure. We will study the dynamical properties of the Schrödinger operator acting on $l^2(\mathbb{Z})$,

$$H_{f,\theta}u(n) = u(n+1) + u(n-1) + \phi(f^n\theta)u(n), \tag{1-1}$$

where $\theta \in \mathcal{M}$ is the phase.

The time-dependent Schrödinger equation

$$i\,\partial_t u = H_\theta u,$$

leads to a unitary dynamical evolution

$$u(t) = e^{-itH_\theta}u(0).$$

Under the time evolution, the wavepacket will in general spread out with time. For operators with absolutely continuous spectrum, scattering theory leads to a good understanding of the quantum dynamics. In this paper we will study the spreading of the wavepacket under the assumption of positive Lyapunov exponent, which automatically implies the absence of absolutely continuous spectrum.

Let $e^{-itH_\theta}\delta_0$ be the time evolution with the localized initial state δ_0. Let

$$a_\theta(n,t) = |\langle e^{-itH_\theta}\delta_0, \delta_n\rangle|^2;$$

$a_\theta(n,t)$ describes the probability of finding the wavepacket at site n at time t. We denote the p-th moment of $a_\theta(n,t)$ by

$$\langle |X|_\theta^p(t)\rangle = \sum_n (1+|n|)^p a_\theta(n,t).$$

Dynamical localization is defined as boundedness of $\langle |X|_\theta^p(t)\rangle$ in time t. This implies purely point spectrum; therefore for general operators with positive Lyapunov exponent such a strong control of the wavepacket is not possible. Thus we need to define proper transport exponents which describe the rate of the spreading of the wavepacket. For $p > 0$ define the upper and lower transport exponents

$$\beta_\theta^+(p) = \limsup_{t\to\infty} \frac{\ln\langle |X|_\theta^p(t)\rangle}{p\ln t}, \quad \beta_\theta^-(p) = \liminf_{t\to\infty} \frac{\ln\langle |X|_\theta^p(t)\rangle}{p\ln t}.$$

Obtaining upper bounds for the two transport exponents above implies a power-law control of the spreading rate of the entire wavepacket.

It is also interesting to consider a portion of the wavepacket. For a nonnegative function $A(t)$ of time, let

$$\langle A(t)\rangle_T = \frac{2}{T}\int_0^\infty e^{-2t/T}A(t)\,dt$$

be its time average. Set

$$P_{\theta,T}(L) = \sum_{|n|\le L} \langle a_\theta(n,t)\rangle_T.$$

Roughly speaking, $P_{\theta,T}(T^a) > \tau$ means that, in average, over time T, a portion of the wavepacket stays inside a box of size T^a. Let us consider two other scaling exponents

$$\bar{\xi}_\theta = \lim_{\tau\to 0}\limsup_{T\to\infty} \frac{\ln\inf\{L : P_{\theta,T}(L) + P_{f\theta,T}(L) > \tau\}}{\ln T},$$
$$\underline{\xi}_\theta = \lim_{\tau\to 0}\liminf_{T\to\infty} \frac{\ln\inf\{L : P_{\theta,T}(L) + P_{f\theta,T}(L) > \tau\}}{\ln T}$$

introduced, in the half-line case, in [Killip et al. 2003].

The vanishing of $\beta^\pm$ and $\bar{\xi}$, $\underline{\xi}$ can be viewed as localization-type statements. If $\mathcal{M} = \mathbb{T}$ is the 1-dimensional torus and $f : \theta \to \theta + \alpha$ is the irrational rotation, the Lebesgue measure m is the unique invariant probability measure of f. It was first proved in [Damanik and Tcheremchantsev 2007; 2008] that in this case, for ϕ a trigonometric polynomial, under the assumption of positive Lyapunov exponent, $\beta_\theta^+(p) = 0$ for all $p > 0$, all θ and Diophantine α, and $\beta_\theta^-(p) = 0$ for all $p > 0$, all θ and all α. It was recently proved in [Jitomirskaya and Mavi 2017] that under very mild restrictions on the regularity of the potential, under the assumption of positivity and continuity of the Lyapunov exponent, $\beta_\theta^+(p) = 0$ for all $p > 0$, all θ and Diophantine α, and $\beta_\theta^-(p) = 0$ for all $p > 0$, all θ and all α. It was also proved in that paper that for piecewise Hölder functions, under the assumption of positive Lyapunov exponent, $\bar{\xi}_\theta = 0$ for a.e. θ and Diophantine α, and $\underline{\xi}_\theta = 0$ for a.e. θ and all α.

Remark 1.1. The two Diophantine sets of α are different between [Damanik and Tcheremchantsev 2007; 2008] and [Jitomirskaya and Mavi 2017]. They are both full-measure sets, but [Jitomirskaya and Mavi 2017] covers a slightly thinner set of frequencies because of the need to handle potentials with weaker regularity.

In this paper we consider a d-dimensional compact Riemannian manifold $\mathcal{M}$ and a uniquely ergodic volume-preserving map f. We consider maps with the following volume-scaling property. For $1 \le l \le d$, for a smooth map $\sigma : Q^l \to \mathcal{M}$, where $Q^l = [0,1]^l$, let

$$\mathrm{Vol}_{g,l}(\sigma) := \int_{Q^l} \overline{\mathrm{Vol}}_{g,l}(d\sigma),$$

where $\overline{\mathrm{Vol}}_{g,l}(d\sigma)$ is the volume form on Q^l induced by σ from the given Riemannian metric g on $\mathcal{M}$. Let $\Sigma(l)$ be the set of all C^∞ mappings $\sigma : Q^l \to \mathcal{M}$. For $n = 1, 2, \ldots$ and $1 \le l \le d$, let

$$V_l(f) := \sup_{\sigma\in\Sigma(l)} \limsup_{n\to\infty} \frac{1}{n}\log \mathrm{Vol}_{g,l}(f^n\sigma) \quad\text{and}\quad V(f) := \max_l V_l(f). \tag{1-2}$$

A volume-preserving f always satisfies $V_d(f) = V_d(f^{-1}) = 0$. Here we need to make an extra assumption that $V(f) = V(f^{-1}) = 0$. It is known that for a smooth invertible map f, $V(f) = V(f^{-1})$

is equal to the *topological entropy* of f [Yomdin 1987]; thus our class of maps includes all smooth maps with zero topological entropy. In particular, it includes both the irrational rotation and the skew-shift.

For such maps we will assume that f has a bounded discrepancy.

Let $J_N(\theta) = J(\theta, f\theta, \dots, f^{N-1}\theta)$, see (2-16), be the isotropic discrepancy function of the sequence $\{f^n\theta\}_{n=0}^{N-1}$. For $\delta > 0$, we will say f has *strongly δ-bounded isotropic discrepancy* if $J_N(\theta) \le |N|^{-\delta}$ uniformly in θ for $|N| > N_0$; f has *weakly δ-bounded isotropic discrepancy* if there exists a sequence $\{N_j\}$ such that $J_{N_j}(\theta) \le |N_j|^{-\delta}$ uniformly in θ. It turns out many concrete dynamical systems feature these properties. We will show in Lemmas 3.7–3.9 that the following hold:

- A shift of higher-dimensional tori, $f : \theta \to \theta + \alpha$, has strongly bounded isotropic discrepancy for Diophantine α.
- A skew-shift, $f : (y_1, y_2, \dots, y_d) \to (y_1 + \alpha, y_2 + y_1, \dots, y_d + y_{d-1})$, has strongly bounded isotropic discrepancy for Diophantine α, and weakly bounded isotropic discrepancy for Liouvillean α.

Under the assumption of boundedness of discrepancy and a scaling property of f, we are ready to formulate the following two abstract results.

Let μ_θ be the spectral measure of H_θ corresponding to δ_0. Let $N = \int_{\mathcal{M}} \mu_\theta \, \mathrm{d}\,\mathrm{Vol}_g$ be the integrated density of states. Let $L(E)$ be the Lyapunov exponent; see (2-6).

Theorem 1. *Let ϕ be a piecewise Hölder function. Suppose $L(E)$ is positive on a Borel subset U with $N(U) > 0$. Suppose f is a uniquely ergodic volume-preserving map satisfying $V(f) = V(f^{-1}) = 0$. We have*:

- *If, for some $\delta > 0$, f has weakly δ-bounded isotropic discrepancy, then $\underline{\xi}_\theta = 0$, for Vol_g-a.e. $\theta \in \mathcal{M}$;*
- *If, for some $\delta > 0$, f has strongly δ-bounded isotropic discrepancy, then $\bar{\xi}_\theta = 0$ for Vol_g-a.e. $\theta \in \mathcal{M}$.*

Remark 1.2. The full-measure set of θ appearing in Theorem 1 is precisely the set $\{\theta : \mu_\theta + \mu_{f\theta}(U) > 0\}$.

Theorem 2. *Under the assumptions of Theorem 1, assume also $L(E)$ is continuous in E and $L(E) > 0$ for every $E \in \mathbb{R}$. We have*:

- *If, for some $\delta > 0$, f has weakly δ-bounded isotropic discrepancy, then $\beta_\theta^-(p) = 0$ for all $\theta \in \mathcal{M}$ and $p > 0$;*
- *If, for some $\delta > 0$, f has strongly δ-bounded isotropic discrepancy, then $\beta_\theta^+(p) = 0$ for all $\theta \in \mathcal{M}$ and $p > 0$.*

Remark 1.3. Strongly δ-bounded isotropic discrepancy is essential for vanishing of $\bar{\xi}$ and $\beta_\theta^+(p)$, see Remarks 1.6 and 1.9. However, it is not yet clear whether weakly δ-bounded isotropic discrepancy (or any condition at all other than mere positivity of the Lyapunov exponent) is essential for vanishing of the $\underline{\xi}$ or of β_θ^-.

Theorems 1 and 2 extend the results of [Damanik and Tcheremchantsev 2007; 2008; Jitomirskaya and Mavi 2017] from irrational rotations of the circle to general uniquely ergodic maps of compact Riemannian manifolds with zero topological entropy and bounded discrepancy. One key to achieving such generality is a new argument that does not rely on harmonic analysis/approximation by trigonometric polynomials.

By [Damanik and Tcheremchantsev 2003], $\beta_\theta^-(p) \geq p \dim_H(\mu_\theta)$, where $\dim_H(\mu)$ is the Hausdorff dimension of μ. Thus as a consequence of $\beta_\theta^-(p) = 0$ we have the following.

Corollary 1.4. *Under the assumptions of Theorem 2*, $\dim_H(\mu_\theta) = 0$ *for all* $\theta \in \mathcal{M}$.

Remark 1.5. The point here is that we obtain zero Hausdorff dimension of the spectral measure for *all* rather than a.e. $\theta \in \mathcal{M}$ (the latter is known for general ergodic potentials [Simon 2007]). The statement for all θ has only been known for irrational rotations of $\mathbb{T}^1$ (proved for trigonometric polynomials in [Jitomirskaya and Last 2000], and follows easily for piecewise functions from the results of [Jitomirskaya and Mavi 2017]).

The following Theorems 3–6 are all corollaries of our abstract results. Theorems 7 and 8 depend on a somewhat different technique (bypassing the discrepancy considerations), which allows us to cover more frequencies in the case of the shift of $\mathbb{T}^2$. To our knowledge, Theorems 3–8 are the first arithmetic localization-type results.

Let us introduce the Diophantine condition (DC) and the weak Diophantine condition (WDC) on $\mathbb{T}^d$:

$$\mathrm{DC}(\tau) = \bigcup_{c>0} \mathrm{DC}(c,\tau) = \bigcup_{c>0} \left\{ (\alpha_1, \ldots, \alpha_d) : \|\langle \vec{h}, \alpha\rangle\|_{\mathbb{R}/\mathbb{Z}} \geq \frac{c}{r(\vec{h})^\tau} \text{ for any } \vec{h} \neq \vec{0} \right\},$$

where $r(\vec{h}) = \prod_{i=1}^d \max(|h_i|, 1)$ (it is well known that when $\tau > 1$, $\mathrm{DC}(\tau)$ is a full-measure set), and

$$\mathrm{WDC}(\tau) = \bigcup_{c>0} \mathrm{WDC}(c,\tau) = \bigcup_{c>0} \left\{ (\alpha_1, \ldots, \alpha_d) : \max\{\|h\alpha_i\|_{\mathbb{R}/\mathbb{Z}}\} \geq \frac{c}{|h|^\tau} \text{ for any } h \neq 0 \right\}, \quad h \in \mathbb{Z}$$

(it is well known that when $\tau > 1/d$, $\mathrm{WDC}(\tau)$ is a full-measure set).

Theorem 1 reduces vanishing of (upper or lower) ξ_θ to bounds on the isotropic discrepancy. As corollaries, we obtain:

Theorem 3. *Let f be an irrational shift on $\mathbb{T}^d$. For piecewise Hölder ϕ, suppose $L(E)$ is positive on a Borel subset U with $N(U) > 0$. Then if $\alpha \in \mathrm{DC}(\tau) \subset \mathbb{T}^d$, $\tau > 1$, we have $\bar{\xi}_\theta = 0$ for a.e. $\theta \in \mathbb{T}^d$.*

Remark 1.6. The Diophantine condition is essential for the vanishing of $\bar{\xi}$ [Jitomirskaya and Zhang 2015].

Theorem 4. *Let f be a skew-shift. For piecewise Hölder ϕ, suppose $L(E)$ is positive on a Borel subset U with $N(U) > 0$. Then*:

- *For all irrational α, we have $\xi_{\vec{y}} = 0$ for a.e. $\vec{y} \in \mathbb{T}^d$.*
- *If $\alpha \in \mathrm{DC}(\tau)$ for some $\tau > 1$, then $\bar{\xi}_{\vec{y}} = 0$ for a.e. $\vec{y} \in \mathbb{T}^d$.*

Remark 1.7. The full-measure set appearing in Theorems 3 and 4 is precisely the set $\{\theta : \mu_\theta + \mu_{f\theta}(U) > 0\}$.

Similarly, for systems with continuous Lyapunov exponent, Theorem 2 reduces vanishing of $\beta_\theta^\pm(p)$ to the same discrepancy bounds, and we obtain:

Theorem 5. *Under the assumptions of Theorem 3, assume in addition that $L(E)$ is continuous in E and $L(E) > 0$ for every $E \in \mathbb{R}$. Then if $\alpha \in \mathrm{DC}(\tau) \subset \mathbb{T}^d$, we have $\beta_\theta^+(p) = 0$ for all $\theta \in \mathbb{T}^d$, $p > 0$.*

Corollary 1.8. *Under the assumptions of Theorem 5, if* $\alpha \in \mathrm{DC}(\tau)$, *then* $\dim_H(\mu_\theta) = 0$ *for all* $\theta \in \mathbb{T}^d$.

Remark 1.9. The Diophantine condition is essential for $\beta^+ = 0$ [Jitomirskaya and Zhang 2015].

Theorem 6. *Under the assumptions of Theorem 4, assume in addition that* $L(E)$ *is continuous in* E *and* $L(E) > 0$ *for every* $E \in \mathbb{R}$. *Then*:

- *For all irrational* α, *we have* $\beta^-_{\vec{y}}(p) = 0$ *for all* $\vec{y} \in \mathbb{T}^d$, $p > 0$.
- *If* $\alpha \in \mathrm{DC}(\tau)$ *for some* $\tau > 1$, *then* $\beta^+_{\vec{y}}(p) = 0$ *for all* $\vec{y} \in \mathbb{T}^d$, $p > 0$.

Corollary 1.10. *Under the assumptions of Theorem 6, for all irrational* α, *we have* $\dim_H(\mu_{\vec{y}}) = 0$ *for all* $\vec{y} \in \mathbb{T}^d$.

Finally, for the case of an irrational shift on $\mathbb{T}^2$ we can make two more delicate statements, using a different technique to obtain arithmetic estimates.

Theorem 7. *Let* f *be an irrational shift on* $\mathbb{T}^2$. *For piecewise Hölder* ϕ, *suppose* $L(E)$ *is positive on a Borel subset* U *with* $N(U) > 0$. *Then if* $\alpha = (\alpha_1, \alpha_2) \in \bigcup_{\tau>1} \mathrm{WDC}(\tau)$, *we have* $\underline{\xi}_\theta = 0$ *for a.e.* $\theta \in \mathbb{T}^2$.

Remark 1.11. The full-measure set appearing in Theorem 7 is precisely the set $\{\theta : \mu_\theta + \mu_{f\theta}(U) > 0\}$.

Theorem 8. *Under the assumptions of Theorem 7, assume in addition that* $L(E)$ *is continuous in* E *and* $L(E) > 0$ *for every* $E \in \mathbb{R}$. *Then if* $\alpha = (\alpha_1, \alpha_2) \in \bigcup_{\tau>1} \mathrm{WDC}(\tau)$, *we have* $\beta^-_\theta(p) = 0$ *for all* $\theta \in \mathbb{T}^2$, $p > 0$.

Corollary 1.12. *Under the assumptions of Theorem 8, if* $\alpha \in \bigcup_{\tau>1} \mathrm{WDC}(\tau)$, *we have* $\dim_H(\mu_\theta) = 0$ *for all* $\theta \in \mathbb{T}^2$.

The most technically complex part of the paper consists in obtaining arithmetic estimates on the covering of the torus by the trajectory of a small ball in a polynomial (in the inverse radius) time, which we obtain by estimating the discrepancy in Theorems 3–6, and by the bounded remainder set technique in Theorems 7 and 8. The discrepancy estimates are standard for the Diophantine shifts and are ideologically similar to the known results on the equidistribution of $n^k\alpha$ for the case of higher-dimensional Diophantine skew-shifts. We still develop the proof for the Diophantine skew-shift case in full detail because we did not find it in the literature and also because it serves as a good preparation for the Liouville higher-dimensional skew-shift, for which, to the best of our knowledge, our estimates are new. We note that for the *Diophantine* skew-shift of $\mathbb{T}^2$ and shifts of $\mathbb{T}^d$ the results on the covering of the torus by the trajectory of a ball are shown in [Avila et al. 2014] by a completely different technique. The authors therein considered smoothed-out indicator functions of small disks, and converted the covering problem to solving cohomological equations. It is unclear to us if that technique is extendable to the Liouville or weakly Diophantine case.

We organize this paper as follows: In Section 2 we introduce some basic definitions. Some of them have been mentioned in the Introduction but not in detail. In Section 3 we will present some key lemmas and prove Theorems 1–8. In Sections 4–7 we prove the key lemmas that are listed in Section 3.

2. Preparation

Riemannian manifolds. Let $\mathcal{M}$ be a d-dimensional compact Riemannian manifold with a Riemannian metric g.

Let K be a compact set in some coordinate patch $(U, x^1, \dots, x^d)$. We define the volume of K to be

$$\mathrm{Vol}_g(K) := \int_{x(K)} \sqrt{|G \circ x^{-1}|}\, dx^1 \cdots dx^d,$$

where $G = \det g_{ij}$, $g_{ij} = g(\partial/\partial x_i, \partial/\partial x_j)$ and $dx^1 \cdots dx^d$ is the Lebesgue measure on $\mathbb{R}^d$. This definition is free of the choice of coordinate. If K is not contained in a single coordinate patch, one could apply a partition of unity to define $\mathrm{Vol}_g(K)$. More precisely, we pick an atlas $(U_\alpha, x_\alpha^1, \dots, x_\alpha^d)$ of $\mathcal{M}$ and a partition of unity $\{\rho_\alpha\}$ subordinate to this atlas. Now we can set

$$\mathrm{Vol}_g(K) = \sum_\alpha \int_{x^\alpha(K \cap U_\alpha)} (\rho_\alpha \sqrt{|G^\alpha|}) \circ (x^\alpha)^{-1}\, dx_\alpha^1 \cdots dx_\alpha^d.$$

The *Riemannian volume density*, see, e.g., [Nicolaescu 2007, Section 3.4], on $(\mathcal{M}, g)$ is

$$d\,\mathrm{Vol}_g = \sum_\alpha (\rho_\alpha \sqrt{|G^\alpha|}) \circ (x^\alpha)^{-1}\, dx_\alpha^1 \cdots dx_\alpha^d. \tag{2-1}$$

With a rescaling, we could always assume $d\,\mathrm{Vol}_g$ is a probability measure on $\mathcal{M}$. In the above definition, we do not assume $\mathcal{M}$ to be oriented. If $\mathcal{M}$ is oriented, then the volume density is actually a positive n-form, called the volume form.

If $\varrho : [a, b] \to \mathcal{M}$ is a continuously differentiable curve in the Riemannian manifold $\mathcal{M}$, then we define its length $l(\varrho)$ by

$$l(\varrho) = \int_a^b \sqrt{g_{\varrho(t)}(\dot\varrho(t), \dot\varrho(t))}\, dt,$$

where $g_{\varrho(t)}$ is the inner product g at the point $\varrho(t)$. One could define the distance between any two points $x, y \in \mathcal{M}$ as

$$\mathrm{dist}(x,y) = \inf\{l(\varrho) : \varrho \text{ is a continuous, piecewise continuously differentiable curve connecting } x \text{ and } y\}.$$

With the definition of distance, *geodesics* in a Riemannian manifold are then the locally distance-minimizing paths.

Let $v \in \mathrm{T}_x\mathcal{M}$ be a tangent vector to the manifold $\mathcal{M}$ at x. Then there is a unique geodesic ϱ_v satisfying $\varrho_v(0) = x$ with initial tangent vector $\dot\varrho_v(0) = v$. The corresponding *exponential map* is defined by $\exp_x(v) = \varrho_v(1)$.

Let $B_r(x) = \{y \in \mathcal{M} : \mathrm{dist}(x, y) < r\}$ be a *geodesic ball* centered at $x \in \mathcal{M}$ with radius r. It is known that $B_r(x) = \exp_x(B(0, r))$, where $B(0, r) = \{v \in \mathrm{T}_x\mathcal{M} : g_x(v, v) < r\}$.

Proposition 2.1. *There exists $r_g > 0$ such that, for all $r < r_g$, there exist positive constants C_g and c_g which are independent of $x \in \mathcal{M}$ so that*

$$c_g r^d \le \mathrm{Vol}_g(B_r(x)) \le C_g r^d \quad \textit{for any } x \in \mathcal{M}. \tag{2-2}$$

Proof. We will discuss the proof briefly. We can identify the tangent space $\mathrm{T}_x\mathcal{M}$ isometrically with $\mathbb{R}^d$. Now $\exp_x : \mathbb{R}^d \to \mathcal{M}$ is a diffeomorphism on some small ball $B_{\mathbb{R}^d}(0,r)$. On this ball, straight lines are mapped to length-minimizing geodesics [do Carmo 1992, Proposition 3.6], and thus Euclidean balls are mapped to geodesic balls of the same radius. Taking r smaller if necessary, we can assume the Jacobian of $\exp_x$ is bounded away from 0 and ∞ on $B_{\mathbb{R}^d}(0,r)$; thus for $r < r_x$ we have that $c_{g_x} r^d \le \mathrm{Vol}_g(B_r(x)) \le C_{g_x} r^d$. Since $\mathcal{M}$ is a compact manifold, we can take r_x, c_{g_x}, C_{g_x} independent of $x \in \mathcal{M}$. □

A subset C of $\mathcal{M}$ is said to be a *geodesically convex set* if, given any two points in C, there is a minimizing geodesic contained within C that joins those two points.

The *convexity radius at a point* $x \in \mathcal{M}$ is the supremum (which may be $+\infty$) of $r_x \in \mathbb{R}$ such that for all $r < r_x$ the geodesic ball $B_{r_x}(x)$ is geodesically convex. The *convexity radius of* $(\mathcal{M}, g)$ is the infimum over the points $x \in \mathcal{M}$ of the convexity radii at these points.

Proposition 2.2 [Berger 2003]. *For a compact manifold $\mathcal{M}$, the convexity radius r'_g of $(\mathcal{M}, g)$ is positive.*

This clearly implies that for any $x \in \mathcal{M}$ and any $r < r'_g$, $B_r(x)$ is geodesically convex.

Piecewise Hölder functions. Let $L_\gamma(\mathcal{M})$ be the space of γ-Lipschitz functions on $\mathcal{M}$. For $\phi \in L_\gamma(\mathcal{M})$ define

$$\|\phi\|_{L_\gamma} = \|\phi\|_\infty + \sup_{\theta_1,\theta_2 \in M} \frac{|\phi(\theta_1) - \phi(\theta_2)|}{\mathrm{dist}\,(\theta_1,\theta_2)^\gamma}. \tag{2-3}$$

We say ϕ is piecewise Hölder if there exists $\gamma > 0$, a positive integer K and $\{\phi_j\}_{j=1}^K \subset L_\gamma(\mathcal{M})$ such that

$$\phi(\theta) = \sum_{j=1}^K \chi_{S_j}(\theta)\phi_j(\theta),$$

where $\{S_j\}_{j=1}^M$ are sets with "good boundary", namely $\{\partial S_j\}_{j=1}^K$ are $(d-1)$-dimensional smooth submanifolds of $\mathcal{M}$. Clearly the discontinuity set J_ϕ of ϕ is $\bigcup_{j=1}^K \partial S_j$, and

$$\mathrm{Vol}_{g,d-1}(J_\phi) \le \sum_{j=1}^K \mathrm{Vol}_{g,d-1}(\partial S_j) < \infty. \tag{2-4}$$

Clearly for any two points θ_1, θ_2 such that $\mathrm{dist}(\theta_i, J_\phi) \ge r$, if $\mathrm{dist}(\theta_1,\theta_2) < r$ then we have

$$|\phi(\theta_1) - \phi(\theta_2)| \le \mathrm{dist}(\theta_1,\theta_2)^\gamma \sum_{j=1}^K \|\phi_j\|_{L_\gamma}. \tag{2-5}$$

Cocycles and Lyapunov exponent. We now introduce the Lyapunov exponent. For a given $z \in \mathbb{C}$, a formal solution u of $Hu = zu$ can be reconstructed using the transfer matrix

$$A(\theta, z) = \begin{pmatrix} z - \phi(\theta) & -1 \\ 1 & 0 \end{pmatrix}$$

via the equation

$$\begin{pmatrix} u(n+1) \\ u(n) \end{pmatrix} = A(f^n\theta, z) \begin{pmatrix} u(n) \\ u(n-1) \end{pmatrix}.$$

Indeed, let $A_k(\theta, z)$ be the product of consecutive transfer matrices:

$$A_k(\theta, z) = \begin{cases} A(f^{k-1}\theta, z) \cdots A(f\theta, z)A(\theta, z) & \text{if } k > 0, \\ I & \text{if } k = 0, \\ (A_{-k}(f^k\theta, z))^{-1} & \text{if } k < 0. \end{cases}$$

Then for any $k \in \mathbb{Z}$ we have the relation

$$\begin{pmatrix} u(k) \\ u(k-1) \end{pmatrix} = A_k(\theta, z) \begin{pmatrix} u(0) \\ u(-1) \end{pmatrix}.$$

We define the Lyapunov exponent

$$L(z) = \lim_k \frac{1}{k} \int_{\mathcal{M}} \ln \|A_k(\theta, z)\| \, \mathrm{d}\,\mathrm{Vol}_g(\theta) = \inf_k \frac{1}{k} \int_{\mathcal{M}} \ln \|A_k(\theta, z)\| \, \mathrm{d}\,\mathrm{Vol}_g(\theta). \tag{2-6}$$

Furthermore, $L(z) = \lim_k (1/k) \ln \|A_k(\theta, z)\|$ for Vol_g-a.e. $\theta \in \mathcal{M}$.

Spectral measure and integrated density of states. Let μ_θ be the spectral measure of H_θ corresponding to δ_0 defined by

$$\langle (H_\theta - z)^{-1}\delta_0, \delta_0 \rangle = \int_{\mathbb{R}} \frac{d\mu_\theta(x)}{x - z}.$$

Then clearly $\mu_{f\theta}$ is the spectral measure of H_θ corresponding to δ_1. Let $N = \int_{\mathcal{M}} \mu_\theta \, \mathrm{d}\,\mathrm{Vol}_g(\theta)$ be the integrated density of states. Then $N = \int_{\mathcal{M}} \frac{1}{2}(\mu_\theta + \mu_{f\theta}) \, \mathrm{d}\,\mathrm{Vol}_g(\theta)$, so $N(U) > 0$ for some set U implies $\frac{1}{2}(\mu_\theta + \mu_{f\theta})(U) > 0$ for Vol_g-a.e. $\theta \in \mathcal{M}$.

Rational approximation.

Single frequency. Let α be an irrational number and let $\{p_n/q_n\}$ be its continued fraction approximants. We have the following properties; see, e.g., [Khinchin 1964]:

$$\frac{1}{2q_{n+1}} \le \|q_n\alpha\|_{\mathbb{T}} \le \frac{1}{q_{n+1}}, \tag{2-7}$$

$$\|k\alpha\| > \|q_n\alpha\| \quad \text{for } q_n < k < q_{n+1}. \tag{2-8}$$

(1) If $\alpha \in \mathrm{DC}(c, \tau)$ for some $c > 0$, we have

$$\|k\alpha\|_{\mathbb{T}} \ge \frac{c}{|k|^\tau} \quad \text{for any } k \ne 0. \tag{2-9}$$

In particular, combining (2-7) with (2-9) we have

$$cq_{n+1} \le q_n^\tau. \tag{2-10}$$

(2) If $\alpha \notin \mathrm{DC}(\tau)$, there exists a subsequence of the continued fraction approximants $\{p_{n_k}/q_{n_k}\}$ such that

$$q_{n_k+1} > q_{n_k}^\tau. \tag{2-11}$$

Multiple frequencies. Let $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_d)$ be a set of irrational frequencies. Let $\{\vec{p}_n/q_n\}$ be its best simultaneous approximation with respect to the Euclidean norm on $\mathbb{T}^d$, namely,

$$\sum_{j=1}^{d} \|q_n\alpha_j\|_{\mathbb{T}}^2 < \sum_{j=1}^{d} \|k\alpha_j\|_{\mathbb{T}}^2 \quad \text{for any } |k| < q_n.$$

Clearly by the pigeonhole principle, we have

$$\sqrt{\sum_{j=1}^{d} \|q_n\alpha_j\|_{\mathbb{T}}{}^2} \le \frac{2\Gamma\big(\frac{1}{2}d+1\big)^{1/d}}{\sqrt{\pi}\, q_{n+1}^{1/d}}. \tag{2-12}$$

We say that

(1) $\alpha \in \mathrm{DC}(c, \tau)$ if

$$\|\langle \vec{k}, \alpha\rangle\|_{\mathbb{T}} \ge \frac{c}{r(\vec{k})^{\tau}} \quad \text{for any } \vec{k} \in \mathbb{Z}^d \backslash \{\vec{0}\}, \tag{2-13}$$

(2) $\alpha \in \mathrm{WDC}(c, \tau)$ if

$$\max_{1\le j\le d} \|k\alpha_j\|_{\mathbb{T}} \ge \frac{c}{|k|^{\tau}} \quad \text{for any } k \in \mathbb{Z}\backslash\{\vec{0}\}. \tag{2-14}$$

Discrepancy. Let $\vec{x}_1, \dots, \vec{x}_N \in \mathcal{M}$. For a subset C of $\mathcal{M}$, let $A(C; \{\vec{x}_n\})$ be the counting function

$$A(C; \{\vec{x}_n\}_{n=1}^N) = \sum_{n=1}^{N} \chi_C(\vec{x}_n). \tag{2-15}$$

The *isotropic discrepancy* $J_N(\{\vec{x}_n\}_{n=1}^N)$ is defined as

$$J_N(\{\vec{x}_n\}_{n=1}^N) = \sup_{C\in\mathscr{C}} \left| \frac{A(C; \{\vec{x}_n\}_{n=1}^N)}{N} - \mathrm{Vol}_g(C)\right|, \tag{2-16}$$

where $\mathscr{C}$ is the family of all geodesically convex subsets of $\mathcal{M}$.

For a point $\theta \in \mathcal{M}$, let $J_N(\theta) = J(\{f^n\theta\}_{n=0}^{N-1})$. We say a map $f : \mathcal{M} \to \mathcal{M}$ has *strongly δ-bounded isotropic discrepancy* if, for some $N > N_0$, we have $J_N(\theta) \le N^{-\delta}$ uniformly in $\theta \in \mathcal{M}$. We say f has *weakly δ-bounded isotropic discrepancy* if there is a subsequence $\{N_j\}$ such that $J_{N_j}(\theta) \le N_j^{-\delta}$ uniformly in $\theta \in \mathcal{M}$.

If $\mathcal{M} = \mathbb{T}^d$ is the d-dimensional torus, we define the *discrepancy* $D_N(\{\vec{x}_n\}_{n=1}^N)$ as

$$D(\{\vec{x}_n\}_{n=1}^N) = \sup_{C\in\mathscr{J}} \left| \frac{A(C; \{\vec{x}\}_{n=1}^N)}{N} - m(C)\right|, \tag{2-17}$$

where $\mathscr{J}$ is the family of boxes C of the form $C = \{(\theta_1, \dots, \theta_d) \in \mathbb{T}^d : \beta_i \le \theta_i < \kappa_i \text{ for } 1 \le i \le d\}$.

For a point $\theta \in \mathbb{T}^d$, let $D_N(\theta) = D(\{f^n\theta\}_{n=0}^{N-1})$. We say a map $f : \mathbb{T}^d \to \mathbb{T}^d$ has *strongly δ-bounded discrepancy* if for some N_0 and all $N > N_0$, we have $D_N(\theta) \le N^{-\delta}$ uniformly in $\theta \in \mathbb{T}^d$. We say f has *weakly δ-bounded discrepancy* if there is a subsequence $\{N_j\}$ such that $D_{N_j}(\theta) \le N_j^{-\delta}$ uniformly in $\theta \in \mathbb{T}^d$.

When $\mathcal{M} = \mathbb{T}^d$, the isotropic discrepancy and discrepancy can be tightly controlled by each other:

Lemma 2.3 [Kuipers and Niederreiter 1974, Theorem 1.6 in Chapter 2]. *For any sequence $\{\vec{x}_n\}_{n=1}^N$ in $\mathbb{T}^d$, we have*

$$D_N(\{\vec{x}_n\}_{n=1}^N) \le J_N(\{\vec{x}_n\}_{n=1}^N) \le (4d\sqrt{d}+1)D_N(\{\vec{x}_n\}_{n=1}^N)^{1/d}. \tag{2-18}$$

Therefore, by (2-18), when $\mathcal{M}=\mathbb{T}^d$:

Proposition 2.4. *A map f has strongly (weakly) δ-bounded isotropic discrepancy for some $\delta>0$ if and only if f has strongly (weakly) $\tilde{\delta}$-bounded discrepancy for some $\tilde{\delta}>0$.*

In Section 5 and the Appendix we are going to apply the following two inequalities to estimate the discrepancy from above. Recall that $r(\vec{h})=\prod_{i=1}^d \max(|h_i|,1)$.

Lemma 2.5 (Erdős–Turán–Koksma [Koksma 1950]). *For any positive integer H_0, we have*

$$D(\{\vec{x}_n\}_{n=1}^N) \le C_d\left(\frac{1}{H_0}+\sum_{0<|\vec{h}|\le H_0}\frac{1}{r(\vec{h})}\left|\frac{1}{N}\sum_{n=1}^N e^{2\pi i\langle\vec{h},\vec{x}_n\rangle}\right|\right), \tag{2-19}$$

where $|\vec{h}|=\max_{j=1}^d |h_j|$.

Lemma 2.6 (Van der Corput's fundamental inequality; see, e.g., [Kuipers and Niederreiter 1974], Lemma 3.1 in Chapter 1). *For any integer $1\le H\le N$, we have*

$$\left|\frac{1}{N}\sum_{n=1}^N u_n\right|^2 \le \frac{N+H-1}{N^2H}\sum_{n=1}^N |u_n|^2+\frac{2(N+H-1)}{N^2H^2}\sum_{k=1}^{H-1}(H-k)\,\mathrm{Re}\sum_{n=1}^{N-k} u_n\bar{u}_{n+k}. \tag{2-20}$$

3. Key lemmas and proofs of Theorems 1–8

Covering $\mathcal{M}$ with the orbit of a geodesic ball and proofs of Theorems 1, 7, 2 and 8.

Lemma 3.1. *Let ϕ be a piecewise Hölder function with $1\ge\gamma>0$. Suppose $L(E)$ is positive on a Borel subset U with $N(U)>0$:*

(1) *If there exists a sequence $r_k\to 0$ such that any geodesic ball in $\mathcal{M}$ with radius r_k covers the whole $\mathcal{M}$ in r_k^{-M} steps, then $\underline{\xi}_\theta=0$ for Vol_g-a.e. $\theta\in\mathcal{M}$.*

(2) *If, for any small $r>0$, any geodesic ball with radius r covers the whole $\mathcal{M}$ in r^{-M} steps, then $\bar{\xi}_\theta=0$ for Vol_g-a.e. $\theta\in\mathcal{M}$.*

Lemma 3.2. *Let ϕ be a piecewise Hölder function with $1\ge\gamma>0$. Suppose $L(E)$ is continuous in E and $L(E)>0$ for every $E\in\mathbb{R}$:*

(1) *If there exists a sequence $r_k\to 0$ such that any geodesic ball in $\mathcal{M}$ with radius r_k covers the whole $\mathcal{M}$ in r_k^{-M} steps, then $\beta_\theta^-(p)=0$ for all $\theta\in\mathcal{M}$ and $p>0$.*

(2) *If, for any small $r>0$, any geodesic ball with radius r covers the whole $\mathcal{M}$ in r^{-M} steps, then $\beta_\theta^+(p)=0$ for all $\theta\in\mathcal{M}$ and $p>0$.*

Lemmas 3.1 and 3.2 are key to our abstract argument. They are proved in Section 4. The connection to bounded discrepancy comes in the following:

Let r_g be as in Proposition 2.1 and r'_g as in Proposition 2.2.

Lemma 3.3. *If f has weakly δ-bounded isotropic discrepancy, then there exists $r_k \to 0$ as $k \to \infty$ such that any geodesic ball in $\mathcal{M}$ with radius r_k will cover the whole $\mathcal{M}$ in $r_k^{-2d/\delta}$ steps.*

Proof. There exists a sequence $\{N_k\}$ and $k_0 > 0$ such that for any $k > k_0$ we have $J_{N_k}(\{f^n\theta\}_{n=0}^{N-1}) \le N_k^{-\delta}$. This means for any geodesically convex set $C \subset \mathcal{M}$,

$$\frac{\sum_{n=0}^{N_k-1} \chi_C(f^n\theta)}{N_k} - \mathrm{Vol}_g(C) \ge -N_k^{-\delta}$$

holds for all $\theta \in \mathcal{M}$. Thus if we take $r_k = N_k^{-\delta/(2d)} < \min(r_g, r'_g)$, then by Proposition 2.2, we know $B_{r_k}(\theta)$ is geodesically convex. By Proposition 2.1, $\mathrm{Vol}_g(B_{r_k}(\theta)) \ge c_g r_k^d = c_g N_k^{-\delta/2} > N_k^{-\delta}$. Thus

$$\sum_{n=0}^{r_k^{-2d/\delta}-1} \chi_{B_{r_k}(\theta)}(f^n\theta) > 0$$

for any $\theta \in \mathcal{M}$. □

Lemma 3.4. *If f has strongly δ-bounded isotropic discrepancy, then for any $0 < r < \min(r_g, r'_g)$, any geodesic ball in $\mathcal{M}$ with radius r will cover the whole $\mathcal{M}$ in $r^{-2d/\delta}$ steps.*

Proof. There exists N_0 such that for any $N > N_0$ we have $J_N(\{f^n\theta\}_{n=0}^{N-1}) \le N^{-\delta}$ for all $\theta \in \mathcal{M}$. This means for any $0 < r < \min(r_g, r'_g)$, any geodesic ball $B_r(\theta)$ (it is geodesically convex by Proposition 2.2) and $N = r^{-2d/\delta}$ we have

$$\frac{\sum_{n=0}^{r^{-2d/\delta}-1} \chi_{B_r(\theta)}(f^n\theta)}{r^{-2d/\delta}} - \mathrm{Vol}_g(B_r(\theta)) \ge -r^{2d}.$$

Since by Proposition 2.1, $\mathrm{Vol}_g(B_r(\theta)) \ge c_g r^d > r^{2d}$, we have

$$\sum_{n=0}^{r^{-2d/\delta}-1} \chi_{B_r(\theta)}(f^n\theta) > 0$$

for any $\theta \in \mathcal{M}$. □

In the case of 2-dimensional irrational rotation, we also have:

Lemma 3.5. *For any $(\alpha_1, \alpha_2) \in \bigcup_{\tau>1} \mathrm{WDC}(\tau)$, there exists $r_k(\alpha_1, \alpha_2, \tau) \to 0$ as $k \to \infty$ such that any Euclidean ball with radius r_k covers the whole $\mathbb{T}^2$ in $r_k^{-800\tau^4}$ steps.*

Remark 3.6. This lemma will be proved in Section 7.

We are now ready to complete the proofs of the main theorems.

Proof of Theorem 1. Combine Lemmas 3.3 and 3.4 with Lemma 3.1. □

Proof of Theorem 7. Combine Lemma 3.5 with Lemma 3.1. □

Proof of Theorem 2. Combine Lemmas 3.3 and 3.4 with Lemma 3.2. □

Proof of Theorem 8. Combine Lemma 3.5 with Lemma 3.2. □

Estimation of discrepancy and proofs of Theorems 3, 5, 4 and 6. We have the following control of the discrepancies of irrational rotation and skew-shift.

Lemma 3.7. *If $\alpha \in \mathrm{DC}(\tau)$, then for some constant $\delta > 0$ we have $D_N(\{\theta + n\alpha\}_{n=0}^{N-1}) \le N^{-\delta}$ uniformly in $\theta \in \mathbb{T}^d$.*

Let

$$\vec{Y}_n = \left(y_1 + \binom{n}{1}\alpha,\ y_2 + \binom{n}{1}y_1 + \binom{n}{2}\alpha,\ \dots,\ y_d + \binom{n}{1}y_{d-1} + \cdots + \binom{n}{d}\alpha\right) = f^n(y_1, \cdots, y_d),$$

where f is the skew-shift.

Lemma 3.8. *If $\alpha \in \mathrm{DC}(\tau)$, then for some constant $\delta > 0$ we have $D_N(\{\vec{Y}_n\}_{n=1}^{N}) \le N^{-\delta}$ uniformly in $(y_1, \dots, y_d) \in \mathbb{T}^d$.*

Lemma 3.9. *If $\alpha \notin \mathrm{DC}(d)$, then for some constant $\delta > 0$ there exists a sequence $\{N_j\}$ such that $D_{N_j}(\{\vec{Y}_n\}_{n=1}^{N_j}) \le N_j^{-\delta}$ uniformly in $(y_1, \dots, y_d) \in \mathbb{T}^d$.*

Remark 3.10. Lemma 3.7 is standard. Its proof will be given in the Appendix. Proofs of Lemmas 3.8 and 3.9 will be given in Section 5.

Proof of Theorems 3, 5. These follow from Lemma 3.7 and Theorems 1 and 2. □

Proof of Theorems 4, 6. These follow from Lemmas 3.8 and 3.9 and Theorems 1 and 2. □

4. Proofs of Lemmas 3.1 and 3.2

Upper and lower bounds on transfer matrices. The following lemma on the uniform upper bound of the transfer matrix is essentially from [Jitomirskaya and Mavi 2017]. We have adapted it into the following form for convenience.

Lemma 4.1 [Jitomirskaya and Mavi 2017, Theorem 3.1]. *Let ϕ be a function whose discontinuity set has measure 0 and f be a uniquely ergodic map on $\mathcal{M}$. Then:*

4.1.1. *Let $L(E)$ be positive on a Borel set U and μ be a measure with $\mu(U) > 0$. Then for any $\zeta > 0$ there exists a number $D_\zeta > 0$, and for any $\epsilon > 0$ there exists a set $B_{\zeta,\epsilon}$ with $0 < \mu(B_{\zeta,\epsilon}) < \zeta$, and an integer $N_{\zeta,\epsilon}$ such that for any $E \in U \setminus B_{\zeta,\epsilon}$*

(1) $L(E) \ge D_\zeta$,

(2) *for $n > N_{\zeta,\epsilon}$, $|z - E| < e^{-4\epsilon n}$ and $\theta \in \mathcal{M}$, we have $(1/n) \ln \|A_n(\theta, z)\| < L(E) + \epsilon$.*

4.1.2. *Furthermore, if $L(E)$ is continuous in E and U is a compact set, there exists $D > 0$ and for any $\epsilon > 0$ there exists an integer N_ϵ such that for any $E \in U$*

(1) $L(E) \ge D$,

(2) *for $n > N_\epsilon$, $|z - E| < e^{-4\epsilon n}$ and $\theta \in \mathcal{M}$, we have $(1/n) \ln \|A_n(\theta, z)\| < L(E) + \epsilon$.*

We are also able to formulate the following lower bound for the norm of transfer matrices.

Lemma 4.2. *Let ϕ be a piecewise Hölder function with $1 \geq \gamma > 0$ and f be a uniquely ergodic volume-preserving map on $\mathcal{M}$ with $V(f) = V(f^{-1}) = 0$. Then*:

4.2.1. *Let $L(E)$ be positive on a Borel set U and μ be a measure with $\mu(U) > 0$. Then for any $\zeta, \epsilon > 0$, let D_ζ, $B_{\zeta,\epsilon}$ and $N_{\zeta,\epsilon}$ be defined as in Lemma 4.1.1*:

(1) *If there exists a sequence $r_k \to 0$ such that any geodesic ball in $\mathcal{M}$ with radius r_k covers the whole $\mathcal{M}$ in r_k^{-M} steps, then there exists a sequence $\{n_k(\epsilon)\}$ such that for $k > k_{\zeta,\epsilon}$, any $E \in U \setminus B_{\zeta,\epsilon}$, $|z - E| < e^{-4\epsilon n_k}$ and $\theta \in \mathcal{M}$ we have*

$$\min_{\iota \in \{-1,1\}} \max_{\iota j = 0,\ldots,e^{(5M\epsilon/\gamma)n_k}} \|A_{n_k}(f^j\theta, z)\| \geq e^{n_k(L(E)-3\epsilon)}.$$

(2) *If, for any small $r > 0$, any geodesic ball with radius r covers the whole $\mathcal{M}$ in r^{-M} steps, then for $n > N'_{\zeta,\epsilon}$, any $E \in U \setminus B_{\zeta,\epsilon}$, $|z - E| < e^{-4\epsilon n}$ and $\theta \in \mathcal{M}$ we have*

$$\min_{\iota \in \{-1,1\}} \max_{\iota j = 0,\ldots,e^{(5M\epsilon/\gamma)n}} \|A_n(f^j\theta, z)\| \geq e^{n(L(E)-3\epsilon)}.$$

4.2.2. *Furthermore, if $L(E)$ is continuous in E and U is a compact set, let D be defined as in Lemma 4.1.2 and for any $\epsilon > 0$ let N_ϵ be defined as in Lemma 4.1.2. Then for any $E \in U$ we have $L(E) \geq D$ and for any $|z - E| < e^{-4\epsilon n}$ we have*:

(1) *If there exists a sequence $r_k \to 0$ such that any geodesic ball in $\mathcal{M}$ with radius r_k covers the whole $\mathcal{M}$ in r_k^{-M} steps, then there exists a sequence $\{n_k(\epsilon)\}$ such that for $k > k_\epsilon$ and any $\theta \in \mathcal{M}$,*

$$\min_{\iota \in \{-1,1\}} \max_{\iota j = 0,\ldots,e^{(5M\epsilon/\gamma)n_k}} \|A_{n_k}(f^j\theta, z)\| \geq e^{n_k(L(E)-3\epsilon)}.$$

(2) *If, for any small $r > 0$, any geodesic ball with radius r covers the whole $\mathcal{M}$ in r^{-M} steps, then for $n > N'_\epsilon$ and any $\theta \in \mathcal{M}$,*

$$\min_{\iota \in \{-1,1\}} \max_{\iota j = 0,\ldots,e^{(5M\epsilon/\gamma)n}} \|A_n(f^j\theta, z)\| \geq e^{n(L(E)-3\epsilon)}.$$

Proof of Lemma 4.2. We will focus on the proof of part (1) of Lemma 4.2.1. The other three proofs will be discussed briefly at the end of this section.

For any $E \in U \setminus B_{\zeta,\epsilon}$ and $n > N_{\zeta,\epsilon}$, by Lemma 4.1.1 we have $(1/n)\|A_n(\theta, E)\| < L(E) + \epsilon$. Since $\int_{\mathcal{M}} (1/n) \ln \|A_n(\theta, E)\| \, \mathrm{d}\,\mathrm{Vol}_g(\theta) \geq L(E)$, we have

$$\mathrm{Vol}_g(M_{n,E,L(E),\epsilon}) := \mathrm{Vol}_g\left(\left\{\theta \in \mathcal{M} : \frac{1}{n} \ln \|A_n(\theta, E)\| > L(E) - \epsilon\right\}\right) > \frac{1}{2}. \tag{4-1}$$

Now we take any $\theta \in M_{n,E,L(E),\epsilon}$ and $|z - E| < e^{-4\epsilon n}$. When $n > 2N_{\zeta,\epsilon} + 3$, by standard telescoping we have

$$\begin{aligned}
\|A_n(\theta, z)\| &\geq \|A_n(\theta, E)\| - \|A_n(\theta, z) - A_n(\theta, E)\| \\
&\geq e^{n(L(E)-\epsilon)} - (n + 2(N_{\zeta,\epsilon} + 1)\|A\|_\infty^{N_{\zeta,\epsilon}})e^{n(L(E)-3\epsilon)} \\
&> e^{n(L(E)-2\epsilon)}
\end{aligned}$$

for large enough $n > N'_{\zeta,\epsilon}$. This means

$$M_{n,E,L(E),\epsilon} \subset M_{n,z,L(E),2\epsilon}. \tag{4-2}$$

We know the discontinuity set of $(1/n)\ln\|A_n(\theta,z)\|$ is $J_n = \bigcup_{l=0}^{n-1} f^{-l}(J_\phi)$, where $J_\phi = \bigcup_{j=1}^{K}\partial S_j$ is defined in the subsection on piecewise Hölder functions on page 874. By our assumption (2-4) and the fact the $V_{d-1}(f^{-1}) = 0$ (by the definition (1-2) of $V(f^{-1})$), for n large enough, we have

$$\mathrm{Vol}_{g,d-1}(J_n) \le e^{n\epsilon}\,\mathrm{Vol}_{g,d-1}(J_\phi); \tag{4-3}$$

note that the largeness depends only on f. Define

$$\widetilde{M}_{n,z,L(E),2\epsilon} = M_{n,z,L(E),2\epsilon} \setminus \overline{F_{2e^{-5\epsilon n/\gamma}}(J_n)},$$

where a neighborhood is defined as

$$F_r(A) = \{\theta \in \mathcal{M} : \mathrm{dist}(\theta, A) < r\}.$$

Then by (4-3),

$$\begin{aligned}\mathrm{Vol}_g(\widetilde{M}_{n,z,L(E),2\epsilon}) &\ge \mathrm{Vol}_g(M_{n,z,L(E),2\epsilon}) - 4e^{-5\epsilon n/\gamma}\,\mathrm{Vol}_{g,d-1}(J_n)\\ &\ge \mathrm{Vol}_g(M_{n,z,L(E),2\epsilon}) - 4e^{-n(5\epsilon/\gamma-\epsilon)}\,\mathrm{Vol}_{g,d-1}(J_\phi) > \tfrac{2}{5}.\end{aligned}$$

In particular, it is a nonempty set. Now we take any $\tilde{\theta} \in \widetilde{M}_{n,z,L(E),2\epsilon}$ and $\theta \in B_{e^{-5\epsilon n/\gamma}}(\tilde{\theta})$. We have, by telescoping, (2-5) and the fact that $V_1(f) = 0$ (by the definition (1-2) of $V(f)$),

$$\begin{aligned}&\|A_n(\theta,z)\|\\ &\quad\ge \|A_n(\tilde{\theta},z)\| - \|A_n(\theta,z) - A_n(\tilde{\theta},z)\|\\ &\quad\ge e^{n(L(E)-2\epsilon)} - \left(\sum_{l=1}^{K}\|\phi_l\|_{L^\gamma}\right)(n+2(N_{\zeta,\epsilon}+1)\|A\|_\infty^{N_{\zeta,\epsilon}})e^{n(L(E)+\epsilon)} \max_{j=0,\dots,n-1}(\mathrm{dist}(f^j\theta, f^j\tilde{\theta}))^\gamma\\ &\quad\ge e^{n(L(E)-2\epsilon)} - \left(\sum_{l=1}^{K}\|\phi_l\|_{L^\gamma}\right)(\mathrm{dist}(\theta,\tilde{\theta}))^\gamma(n+2(N_{\zeta,\epsilon}+1)\|A\|_\infty^{N_{\zeta,\epsilon}})e^{n(L(E)+\epsilon+\gamma\epsilon)}\\ &\quad> e^{n(L(E)-3\epsilon)}\end{aligned}$$

for $n > N''_{\zeta,\epsilon}$. This means

$$F_{e^{-5\epsilon n/\gamma}}(\widetilde{M}_{n,z,L(E),2\epsilon}) \subset M_{n,z,L(E),3\epsilon}.$$

Hence for $E \in U \setminus B_{\zeta,\epsilon}$, $n > N''_{\zeta,\epsilon}$ and $|z-E| < e^{-4\epsilon n}$, we know $M_{n,z,L(E),3\epsilon}$ contains a geodesic ball with radius $e^{-(5\epsilon/\gamma)n}$. Then there exists a sequence $\{n_k(\epsilon)\}$ such that a geodesic ball with radius $e^{-(5\epsilon/\gamma)n_k} \sim r_k$ covers the whole $\mathcal{M}$ in at most $e^{(5M\epsilon/\gamma)n_k}$ steps. Thus for $E \in U \setminus B_{\zeta,\epsilon}$, $k > k_{\zeta,\epsilon}$ such that $n_k(\epsilon) > N''_{\zeta,\epsilon}$, any $|z-E| < e^{-4\epsilon n_k}$ and any $\theta \in \mathbb{T}^d$ we have

$$\min_{\iota\in\{-1,1\}}\ \max_{\iota j=0,\dots,e^{(5M\epsilon/\gamma)n_k}} \|A_{n_k}(f^j\theta,z)\| > e^{n_k(L(E)-3\epsilon)}.$$

Remark 4.3. Notice that part (2) of Lemma 4.2.1 follows without taking a subsequence $\{n_k(\epsilon)\}$. Also, Lemma 4.2.2 follows without excluding the set $B_{\zeta,\epsilon}$. □

Dynamical bounds on ξ_θ. The key to estimating ξ_θ is to apply the following lemma by Killip, Kiselev and Last.

Following [Jitomirskaya and Last 1999], for $f:\mathbb{Z}\to H$ where H is a Banach space, the truncated l^2 norms in the positive and negative directions are defined by

$$\|f\|_L^2=\sum_{n=1}^{\lfloor L\rfloor}|f(n)|^2+(L-\lfloor L\rfloor)|f(\lfloor L\rfloor+1)|^2 \qquad \text{for } L>0,$$

$$\|f\|_L^2=\sum_{n=0}^{\lfloor L\rfloor+1}|f(n)|^2+(\lfloor L\rfloor+1-L)|f(\lfloor L\rfloor)|^2 \quad \text{for } L<0.$$

The truncated l^2 norm in both directions is defined by

$$\|f\|_{L_1,L_2}^2=\sum_{n=-\lfloor L_1\rfloor}^{\lfloor L_2\rfloor}|f(n)|^2+(L_1-\lfloor L_1\rfloor)|f(-\lfloor L_1\rfloor-1)|^2+(L_2-\lfloor L_2\rfloor)|f(\lfloor L_2\rfloor+1)|^2 \quad \text{for } L_1,L_2\geq 1.$$

With $A_\bullet(\theta,z)$ being a function on $\mathbb{Z}$, define $\tilde{L}_\epsilon^+(\theta,z)\in\mathbb{R}^+$ and $\tilde{L}_\epsilon^-(\theta,z)\in\mathbb{R}^-$ by requiring

$$\|A_\bullet(\theta,z)\|_{\tilde{L}_\epsilon^\pm(\theta,z)}=2\|A(\theta,z)\|\epsilon^{-1}.$$

Lemma 4.4 [Killip et al. 2003, Theorem 1.5]. *Let H_θ be a Schrödinger operator and μ_θ be the spectral measure of H_θ and δ_0. Let $T>0$ and $L_1,L_2>2$. Then*

$$\left\langle \tfrac{1}{2}(\|e^{-itH_\theta}\delta_0\|_{L_1,L_2}^2+\|e^{-itH_\theta}\delta_1\|_{L_1,L_2}^2)\right\rangle_T > C\tfrac{1}{2}(\mu_\theta+\mu_{f\theta})(\{E:|\tilde{L}_{T^{-1}}^-|\leq L_1,\ \tilde{L}_{T^{-1}}^+\leq L_2\}), \quad (4\text{-}4)$$

where C is an universal constant.[1]

This lemma directly implies

$$P_{\theta,T}(L)+P_{f\theta,T}(L)>C\tfrac{1}{2}(\mu_\theta+\mu_{f\theta})(\{E:\|A_\bullet(\theta,z)\|_{\pm L}>2\|A(\theta,z)\|T\}).$$

The plan is to show that for any $\eta>1$ and any θ_0 satisfying $(\mu_{\theta_0}+\mu_{f\theta_0})(U)>0$, we have

$$(\mu_{\theta_0}+\mu_{f\theta_0})(\{E:\|A_\bullet(\theta_0,z)\|_{\pm T}>T^\eta\})\gtrsim(\mu_{\theta_0}+\mu_{f\theta_0})(U).$$

Proof of Lemma 3.1. We will prove part (1) in detail. Part (2) will be discussed briefly at the end of this proof.

Fix $\eta>1$ and θ_0 such that $(\mu_{\theta_0}+\mu_{f\theta_0})(U)>0$. Let $\zeta=\frac{1}{2}(\mu_{\theta_0}+\mu_{f\theta_0})(U)$, $D=D_\zeta$ from Lemma 4.1, and $\epsilon=\min(\gamma D/(40M\eta), D/6)$. Then by Lemma 4.1, there exists a set B, $0<|B|<\frac{1}{2}(\mu_{\theta_0}+\mu_{V\theta_0})(U)$, and a sequence $\{n_k\}$ such that $L(E)\geq D$ on $U\setminus B$ and for $E\in U\setminus B,\ k\geq k_0,\ |z-E|<e^{-4\epsilon n_k}$ and

[1] Here we formulate this lemma for operators with potential $V(n)=\phi(f^n\theta)$. This covers arbitrary bounded potentials by taking f to be a corresponding subshift.

any $\theta \in \mathcal{M}$,

$$\min_{\iota\in\{-1,1\}} \max_{\iota j=0,\dots,e^{(5M\epsilon/\gamma)n_k}} \|A_{n_k}(f^j\theta, z)\| > e^{n_k(L(E)-3\epsilon)}.$$

Using that $A_{s+t}(\theta, z) = A_t(f^s(\theta), z)A_s(\theta, z)$, this implies, by the condition on ϵ,

$$\|A_\bullet(\theta, z)\|_{\pm e^{(10M\epsilon/\gamma)n_k}} > e^{\frac{1}{2}n_k(L(E)-3\epsilon)} \geq e^{(10M\epsilon/\gamma)n_k\eta}.$$

If we take $T_k = e^{(10M\epsilon/\gamma)n_k}$, then $U \setminus B \subset \{E : \|A_\bullet(\theta, E)\|_{\pm T_k} > T_k^\eta\}$ for any θ, in particular θ_0. Then by (4-4),

$$P_{\theta_0,T_k^\eta}(T_k) + P_{f\theta_0,T_k^\eta}(T_k) \geq C\tfrac{1}{2}(\mu_{\theta_0} + \mu_{f\theta_0})(\{E : \|A_\bullet(\theta_0, E)\|_{\pm T_k} > T_k^\eta\}) \geq \widetilde{C}\tfrac{1}{2}(\mu_{\theta_0} + \mu_{f\theta_0})(U).$$

This implies $\underline{\xi}_\theta = 0$ for all $\theta \in \mathcal{M}$ such that $(\mu_\theta + \mu_{f\theta})(U) > 0$.

Remark 4.5. Using Lemmas 4.1.1(2) and 4.2.1(2) instead of 4.1.1(1) and 4.2.1(1), part (2) can be proved without taking a subsequence n_k; therefore the conclusion holds for all T large enough rather than a sequence T_k. □

Bounds on β. The key to the bounds on β is to apply the following lemma.

Lemma 4.6 [Damanik and Tcheremchantsev 2007, Theorem 1; 2008, Corollary 1]. *Let H be the Schrödinger operator, with f real-valued and bounded, and $K \geq 4$ such that $\sigma(H) \subset [-K+1, K-1]$. Suppose for all $\rho \in (0, 1)$ we have*

$$\int_{-K}^{K} \left(\min_{\iota\in\{-1,1\}} \max_{1\leq \iota n\leq T^\rho} \left\| A_n\left(E + \frac{i}{T}\right)\right\|^2 \right)^{-1} \mathrm{d}E = O(T^{-\eta}) \tag{4-5}$$

for any $\eta \geq 1$. Then $\beta^+(p) = 0$ for all $p > 0$. If (4-5) is satisfied for a sequence $T_k \to \infty$, then $\beta^-(p) = 0$ for all $p > 0$.

Proof of Lemma 3.2. We will prove part (1) in detail. A modification needed for part (2) is discussed briefly at the end of this proof.

It suffices to consider small $\rho \in (0, 1)$. Fix any $\rho \in (0, 1)$ small and $\eta \geq 1$. Assume $\sigma(H) \subset [-K+1, K-1]$. Since $L(E)$ is continuous in E on a compact set $[-K, K]$, we have $L(E) \geq D > 0$ on $[-K, K]$. Fix $\epsilon_\eta = \min(\rho\gamma D/(20M\eta), D/6)$. By Lemma 4.2.2 there exists a sequence $\{n_{\eta,k}\}$ such that for any $E \in [-K, K]$, $k > k_\eta$, any $|z - E| < e^{-4\epsilon_\eta n_{\eta,k}}$ and any $\theta \in \mathcal{M}$,

$$\min_{\iota\in\{-1,1\}} \max_{\iota j=0,\dots,e^{(5M\epsilon_\eta/\gamma)n_{\eta,k}}} \|A_{n_{\eta,k}}(f^j\theta, z)\| > e^{n_{\eta,k}(L(E)-3\epsilon_\eta)}.$$

Thus

$$\min_{\iota\in\{-1,1\}} \max_{j=0,\dots,e^{(10M\epsilon_\eta/\gamma)n_{\eta,k}}} \|A_j(\theta, z)\|^2 \geq e^{n_{\eta,k}(L(E)-3\epsilon_\eta)} \geq e^{(10M\epsilon_\eta/(\gamma\rho))n_{\eta,k}\eta}$$

holds for any $\theta \in \mathcal{M}$, any $E \in [-K, K]$ and $|z - E| < e^{-4\epsilon_\eta n_{\eta,k}}$. Now we take $T_{\eta,k} = e^{(10M\epsilon_\eta/(\gamma\rho))n_{\eta,k}}$,

$$\left| E + \frac{i}{T_{\eta,k}} - E \right| = \frac{1}{T_{\eta,k}} < e^{-4\epsilon_\eta n_{\eta,k}}.$$

Thus

$$\min_{\iota\in\{-1,1\}} \max_{\iota j=0,\dots,T^{\rho}_{\eta,k}} \left\| A_j\left(\theta, E+\frac{i}{T_{\eta,k}}\right)\right\|^2 \geq T^{\eta}_{\eta,k}$$

holds for any $E \in [-K, K]$. Therefore

$$\int_{-K}^{K} \left(\min_{\iota\in\{-1,1\}} \max_{1\leq \iota n\leq T^{\rho}_{\eta,k}} \left\| A_n\left(\theta, E+\frac{i}{T_{\eta,k}}\right)\right\|^2 \right)^{-1} dE \leq 2K T^{-\eta}_{\eta,k}.$$

Now take a sequence $\{k_i\}$ such that $T_{1,k_1} < T_{2,k_2} < \cdots$. Let $T_m = T_{m,k_m}$. Then

$$\int_{-K}^{K} \left(\min_{\iota\in\{-1,1\}} \max_{1\leq \iota n\leq T^{\rho}_{m}} \left\| A_n\left(\theta, E+\frac{i}{T_{m}}\right)\right\|^2 \right)^{-1} dE \leq 2K T^{-m}_{m}.$$

By (4-5), we have $\beta^-_\theta(p) \leq \rho$ for all $\theta \in \mathcal{M}$, any $\rho \in (0, 1)$ and any $p > 0$; thus $\beta^-_\theta(p) = 0$ for all $\theta \in \mathcal{M}$ and any $p > 0$.

Remark 4.7. Using Lemmas 4.1.2(2) and 4.2.2(2), part (2) follows without taking a subsequence $\{n_{\eta,k}\}$. Therefore the conclusion holds for all T large rather than a sequence T_k. □

5. Skew-shift: proofs of Lemmas 3.8 and 3.9

In this section, we obtain the discrepancy bounds for the skew-shift. While the Diophantine case is likely known, we didn't find this in the literature. We thus present a detailed proof, especially since we build our proof for the Liouvillean case on some of the same considerations.

Skew-shift. Let $f\colon \mathbb{T}^d \to \mathbb{T}^d$ be defined as

$$f(y_1, y_2, \dots, y_d) = (y_1+\alpha, y_2+y_1, \dots, y_d+y_{d-1}).$$

Let $\vec{Y}_n = f^n(y_1, \dots, y_d)$; then

$$\vec{Y}_n = \left(y_1+\binom{n}{1}\alpha,\ y_2+\binom{n}{1}y_1+\binom{n}{2}\alpha, \dots, y_d+\binom{n}{1}y_{d-1}+\cdots+\binom{n}{d}\alpha\right), \tag{5-1}$$

where $\binom{n}{m} = 0$ if $n < m$.

Preparation*: *combinatorial identities.

Lemma 5.1. *Let $r_t \in \mathbb{N}$ for $1 \leq t \leq s$. Then we have*

$$\sum_{\substack{l_t=0,1\\1\leq t\leq s}} (-1)^{s-\sum_{t=1}^{s} l_t} \binom{\sum_{t=1}^{s} l_t r_t}{s-1} = 0, \tag{5-2}$$

$$\sum_{\substack{l_t=0,1\\1\leq t\leq s}} (-1)^{s-\sum_{t=1}^{s} l_t} \binom{\sum_{t=1}^{s} l_t r_t}{s} = \prod_{t=1}^{s} r_t. \tag{5-3}$$

Proof. Let us consider the coefficient C_a of x^a in the product $(1+x)^{r_1}(1+x)^{r_2}\cdots(1+x)^{r_s} = (1+x)^{\sum_{i=1}^{s} r_i}$. Let us define

$$A^{(a)} = \left\{ (\vec{j}_1, \vec{j}_2, \ldots, \vec{j}_s) : \vec{j}_t = (j_{t,1}, j_{t,2}, \ldots, j_{t,r_t}),\ j_{t,k} \in \{0,1\},\ \sum_{t=1}^{s} \sum_{k=1}^{r_t} j_{t,k} = a \right\}. \tag{5-4}$$

Each element in $A^{(a)}$ corresponds to one way of choosing 1 or x in each term of the product $(1+x)^{r_1} \cdot (1+x)^{r_2} \cdots (1+x)^{r_s}$ in order to get x^a, where $j_{t,k} = 0$ means we choose 1 out of the k-th $(1+x)$ from $(1+x)^{r_t}$, and $j_{t,k} = 1$ means we choose x instead of 1. Thus the capacity of $A^{(a)}$, denoted by $|A^{(a)}|$, is equal to $C_\alpha = \binom{\sum_{t=1} r_t}{a}$. Let us further define

$$A_t^{(a)} = A^{(a)} \cap \{\vec{j}_t = \vec{0}\}. \tag{5-5}$$

For $a = s-1$, since it is impossible to obtain x^{s-1} with $\vec{j}_t \neq \vec{0}$ for any $1 \le t \le s$, we have

$$A^{(s-1)} \setminus \left(\bigcup_{t=1}^{s} A_t^{(s-1)} \right) = \varnothing. \tag{5-6}$$

For $a = s$,

$$A^{(s)} \setminus \left(\bigcup_{t=1}^{s} A_t^{(s)} \right) = D, \tag{5-7}$$

where

$$D = \left\{ (\vec{j}_1, \vec{j}_2, \ldots, \vec{j}_t) : \sum_{k=1}^{r_t} j_{t,k} = 1 \text{ for } 1 \le t \le s \right\}. \tag{5-8}$$

Clearly,

$$\left| \bigcup_{t=1}^{s} A_t^{(a)} \right| = \sum_{i=1}^{s} (-1)^{i-1} \sum_{1 \le t_1 < t_2 < \cdots < t_i \le s} \left| \bigcap_{l=1}^{i} A_{t_l}^{(a)} \right|, \tag{5-9}$$

in which

$$\sum_{1 \le t_1 < t_2 < \cdots < t_i \le s} \left| \bigcap_{l=1}^{i} A_{t_l}^{(a)} \right| = \sum_{\sum_{t=1}^{s} l_t = s-i}^{l_t = 0,1} \binom{\sum_{t=1}^{s} l_t r_t}{a}. \tag{5-10}$$

Thus

$$\begin{aligned} \left| A^{(a)} \setminus \left(\bigcup_{t=1}^{s} A_t^{(a)} \right) \right| &= \binom{\sum_{t=1}^{s} r_t}{a} + \sum_{i=1}^{s} (-1)^i \sum_{\sum_{t=1}^{s} l_t = s-i}^{l_t=0,1} \binom{\sum_{t=1}^{s} l_t r_t}{a} \\ &= \sum_{1 \le t \le s}^{l_t = 0,1} (-1)^{s - \sum_{t=1}^{s} l_t} \binom{\sum_{t=1}^{s} l_t r_t}{a}. \end{aligned} \tag{5-11}$$

For $a = s-1$, (5-2) follows directly from (5-6) and (5-11). For $a = s$, (5-3) follows from (5-7), (5-11) and the fact that $|D| = \prod_{t=1}^{s} r_t$. $\square$

Diophantine α.

Proof of Lemma 3.8. For $\alpha\in\mathrm{DC}(\tau)$, we take integers

$$H_j\sim N^{2^j/((2^d-1)(\tau+\epsilon))}\quad\text{for } 0\le j\le d-1. \tag{5-12}$$

By Lemma 2.5,

$$\begin{aligned}D(\vec{Y}_1,\dots,\vec{Y}_N)&\le C_d\biggl(\frac{1}{H_0}+\sum_{0<|\vec{h}|\le H_0}\frac{1}{r(\vec{h})}\biggl|\frac{1}{N}\sum_{n=1}^{N}e^{2\pi i\langle\vec{h},\vec{Y}_n\rangle}\biggr|\biggr)\\&=C_d\biggl(\frac{1}{H_0}+\sum_{0<|\vec{h}|\le H_0}\frac{1}{r(\vec{h})}\biggl|\frac{1}{N}\sum_{n=1}^{N}u_n^{(0)}\biggr|\biggr),\end{aligned} \tag{5-13}$$

where

$$u_n^{(0)}=\exp\biggl\{2\pi i\sum_{j=1}^{d}\Bigl(h_j\alpha+\sum_{r=1}^{d-j}h_{j+r}y_r\Bigr)\binom{n}{j}\biggr\}. \tag{5-14}$$

Let

$$u_{k_1,n}^{(1)}=\exp\biggl\{2\pi i\sum_{j=2}^{d}\Bigl(h_j\alpha+\sum_{r=1}^{d-j}h_{j+r}y_r\Bigr)\sum_{l_1=0}^{1}(-1)^{1-l_1}\binom{n+l_1k_1}{j}\biggr\}. \tag{5-15}$$

In general, if $d\ge 3$, we define the following for $1\le s\le d-2$:

$$u_{k_1,\dots,k_s,n}^{(s)}=\exp\biggl\{2\pi i\sum_{j=s+1}^{d}\Bigl(h_j\alpha+\sum_{r=1}^{d-j}h_{j+r}y_r\Bigr)\sum_{\substack{l_t=0,1\\1\le t\le s}}(-1)^{s-\sum_{t=1}^{s}l_t}\binom{n+\sum_{t=1}^{s}l_tk_t}{j}\biggr\}. \tag{5-16}$$

Next, we illustrate the steps of the proof without details for two simple cases $d=2$ and $d=3$. After that, we give a detailed derivation for arbitrary d.

Applying Lemma 2.6 to the $\bigl|\sum_{n=1}^{N}u_n^{(0)}/N\bigr|$ term in (5-13), we obtain

$$\biggl|\frac{1}{N}\sum_{n=1}^{N}u_n^{(0)}\biggr|^2\lesssim\frac{1}{H_1}+\frac{1}{NH_1^2}\sum_{k_1=1}^{H_1}(H_1-k_1)\biggl|\sum_{n=1}^{N-k_1}u_n^{(0)}\overline{u_{n+k_1}^{(0)}}\biggr|. \tag{5-17}$$

The $d=2$ case: Estimating the $\bigl|\sum_{n=1}^{N-k_1}u_n^{(0)}\overline{u_{n+k_1}^{(0)}}\bigr|$ term on the right-hand-side of (5-17) (see (5-27) with $d=2$) we have

$$\biggl|\sum_{n=1}^{N-k_1}u_n^{(0)}\overline{u_{n+k_1}^{(0)}}\biggr|\lesssim\frac{1}{\|h_2k_1\alpha\|_{\mathbb{T}}}. \tag{5-18}$$

The Diophantine condition on α implies that, see (5-28),

$$\sum_{k_1=1}^{H_1}\frac{1}{\|h_2k_1\alpha\|_{\mathbb{T}}}\lesssim\sum_{j=1}^{H_1}\frac{\prod_{l=0}^{1}H_l^{\tau}}{j}\le H_0^{\tau}H_1^{\tau+\epsilon}. \tag{5-19}$$

Thus combining (5-17), (5-18) with (5-19), we have

$$\left|\frac{1}{N}\sum_{n=1}^{N} u_n^{(0)}\right|^2 \lesssim \frac{1}{H_1} = \frac{1}{H_0^2}.$$

Plugging this estimate into (5-13) yields the claimed result for $d = 2$.

The $d = 3$ case: The difference between the cases $d \geq 3$ and $d = 2$ is that for $d = 2$ we can directly estimate (5-17) via (5-18). However, for $d \geq 3$, we need to iteratively apply Lemma 2.6 to reduce the dimension. Now let us illustrate the proof for $d = 3$.

To estimate the right-hand-side of (5-17), we compute as in (5-23),

$$\left|\sum_{n=1}^{N-k_1} u_n^{(0)}\overline{u_{n+k_1}^{(0)}}\right| = \left|\sum_{n=1}^{N-k_1} u_{k_1,n}^{(1)}\right|. \tag{5-20}$$

Applying Lemma 2.6 to the right-hand-side of the equation above, we obtain

$$\left|\frac{1}{N-k_1}\sum_{n=1}^{N-k_1} u_{k_1,n}^{(1)}\right|^2 \lesssim \frac{1}{H_2} + \frac{1}{(N-k_1)H_2^2}\sum_{k_2=1}^{H_2}(H_2-k_2)\left|\sum_{n=1}^{N-\sum_{t=1}^{2}k_t} u_{k_1,n}^{(1)}\overline{u_{k_1,n+k_2}^{(1)}}\right|.$$

As in (5-27), we compute

$$\left|\sum_{n=1}^{N-\sum_{t=1}^{2}k_t} u_{k_1,n}^{(1)}\overline{u_{k_1,n+k_2}^{(1)}}\right| \lesssim \frac{1}{\|h_3k_1k_2\alpha\|_{\mathbb{T}}}.$$

Proceeding as in the $d = 2$ case via the Diophantine condition, we arrive at

$$\left|\frac{1}{N-k_1}\sum_{n=1}^{N-k_1} u_{k_1,n}^{(1)}\right|^2 \lesssim \frac{1}{H_2} = \frac{1}{H_1^2}.$$

Combining (5-17), (5-20) with the estimate above, we obtain

$$\left|\frac{1}{N}\sum_{n=1}^{N} u_n^{(0)}\right|^2 \lesssim \frac{1}{H_1} = \frac{1}{H_0^2}.$$

This proves the claimed result for $d = 3$.

The general case: As we explained above, the general strategy is to use Lemma 2.6 to reduce

$$u^{(0)} \to u^{(1)} \to u^{(2)} \to \cdots \to u^{(d-2)}.$$

We stop when we reach $u^{(d-2)}$, as we can apply (5-27) to these terms.

With the $u^{(s)}$ terms, $0 \leq s \leq d-3$, defined in (5-16), Lemma 2.6 implies,

$$\begin{aligned}\left|\frac{1}{N-\sum_{t=1}^{s}k_t}\sum_{n=1}^{N-\sum_{t=1}^{s}k_t} u_{k_1,\dots,k_s,n}^{(s)}\right|^2 \\ \lesssim \frac{1}{H_{s+1}} + \frac{1}{\left(N-\sum_{t=1}^{s}k_t\right)}H_{s+1}^2\sum_{k_{s+1}=1}^{H_{s+1}}(H_{s+1}-k_{s+1})\left|\sum_{n=1}^{N-\sum_{t=1}^{s+1}k_t} u_{k_1,\dots,k_s,n}^{(s)}\overline{u_{k_1,\dots,k_s,n+k_{s+1}}^{(s)}}\right|.\end{aligned} \tag{5-21}$$

Here

$$\left|\sum_{n=1}^{N-\sum_{t=1}^{s+1}k_t} u^{(s)}_{k_1,\dots,k_s,n}\overline{u^{(s)}_{k_1,\dots,k_s,n+k_{s+1}}}\right|$$

$$\begin{aligned}
&=\Bigg|\sum_{n=1}^{N-\sum_{t=1}^{s+1}k_t}\exp\Bigg\{2\pi i\sum_{j=s+1}^{d}\Bigg(h_j\alpha+\sum_{r=1}^{d-j}h_{j+r}y_r\Bigg)\\
&\qquad\qquad\sum_{1\le t\le s}^{l_t=0,1}(-1)^{s-\sum_{t=1}^{s}l_t}\Bigg(\binom{n+\sum_{t=1}^{s}l_tk_t}{j}-\binom{n+k_{s+1}+\sum_{t=1}^{s}l_tk_t}{j}\Bigg)\Bigg\}\Bigg|\\
&=\Bigg|\sum_{n=1}^{N-\sum_{t=1}^{s+1}k_t}\exp\Bigg\{2\pi i\sum_{j=s+1}^{d}\Bigg(h_j\alpha+\sum_{r=1}^{d-j}h_{j+r}y_r\Bigg)\sum_{1\le t\le s+1}^{l_t=0,1}(-1)^{s+1-\sum_{t=1}^{s+1}l_t}\binom{n+\sum_{t=1}^{s+1}l_tk_t}{j}\Bigg\}\Bigg|\\
&=\Bigg|\sum_{n=1}^{N-\sum_{t=1}^{s+1}k_t}\exp\Bigg\{2\pi i\sum_{j=s+1}^{d}\Bigg(h_j\alpha+\sum_{r=1}^{d-j}h_{j+r}y_r\Bigg)\sum_{0\le t\le s+1}^{l_t=0,1}(-1)^{s+2-\sum_{t=0}^{s+1}l_t}\binom{l_0n+\sum_{t=1}^{s+1}l_tk_t}{j}\Bigg\}\Bigg|\\
&=\Bigg|\sum_{n=1}^{N-\sum_{t=1}^{s+1}k_t}\exp\Bigg\{2\pi i\sum_{j=s+2}^{d}\Bigg(h_j\alpha+\sum_{r=1}^{d-j}h_{j+r}y_r\Bigg)\sum_{0\le t\le s+1}^{l_t=0,1}(-1)^{s+2-\sum_{t=0}^{s+1}l_t}\binom{l_0n+\sum_{t=1}^{s+1}l_tk_t}{j}\Bigg\}\Bigg|
\end{aligned}\tag{5-22}$$

$$\begin{aligned}
&=\Bigg|\sum_{n=1}^{N-\sum_{t=1}^{s+1}k_t}\exp\Bigg\{2\pi i\sum_{j=s+2}^{d}(h_j\alpha+\sum_{r=1}^{d-j}h_{j+r}y_r)\sum_{1\le t\le s+1}^{l_t=0,1}(-1)^{s+1-\sum_{t=1}^{s+1}l_t}\binom{n+\sum_{t=1}^{s+1}l_tk_t}{j}\Bigg\}\Bigg|\\
&=\Bigg|\sum_{n=1}^{N-\sum_{t=1}^{s+1}k_t}u^{(s+1)}_{k_1,\dots,k_{s+1},n}\Bigg|.
\end{aligned}\tag{5-23}$$

Notice that in (5-22), we applied (5-3),

$$\exp\Bigg\{\Bigg(h_{s+1}\alpha+\sum_{r=1}^{d-s-1}h_{s+1+r}y_r\Bigg)\sum_{0\le t\le s+1}^{l_t=0,1}(-1)^{s+2-\sum_{t=0}^{s+1}l_t}\binom{l_0n+\sum_{t=1}^{s+1}l_tk_t}{s+1}\Bigg\}=1.$$

Combining (5-21) with (5-23), we get for any $0\le s\le d-3$,

$$\begin{aligned}
&\left|\frac{1}{N-\sum_{t=1}^{s}k_s}\sum_{n=1}^{N-\sum_{t=1}^{s}k_t}u^{(s)}_{k_1,\dots,k_s,n}\right|^2\\
&\qquad\le\frac{1}{H_{s+1}}+\frac{1}{(N-\sum_{t=1}^{s}k_t)H_{s+1}^2}\sum_{k_{s+1}=1}^{H_{s+1}}(H_{s+1}-k_{s+1})\Bigg(N-\sum_{t=1}^{s+1}k_t\Bigg)\\
&\qquad\qquad\times\left|\frac{1}{N-\sum_{t=1}^{s+1}k_t}\sum_{n=1}^{N-\sum_{t=1}^{s+1}k_t}u^{(s+1)}_{k_1,\dots,k_{s+1},n}\right|.
\end{aligned}\tag{5-24}$$

By (5-21), for $s = d-2$,

$$\begin{aligned}
&\left|\frac{1}{N-\sum_{l=1}^{d-2}k_l}\sum_{n=1}^{N-\sum_{l=1}^{d-2}k_l} u^{(d-2)}_{k_1,\dots,k_{d-2},n}\right|^2 \\
&\lesssim \frac{1}{H_{d-1}}+\frac{1}{(N-\sum_{l=1}^{d-2}k_l)H_{d-1}^2}\sum_{k_{d-1}=1}^{H_{d-1}}(H_{d-1}-k_{d-1})\left|\sum_{n=1}^{N-\sum_{l=1}^{d-1}k_l} u^{(d-2)}_{k_1,\dots,k_{d-2},n}\overline{u^{(d-2)}_{k_1,\dots,k_{d-2},n+k_{d-1}}}\right| \\
&\lesssim \frac{1}{H_{d-1}}+\frac{1}{(N-\sum_{l=1}^{d-2}k_l)H_{d-1}}\sum_{k_{d-1}=1}^{H_{d-1}}\left|\sum_{n=1}^{N-\sum_{l=1}^{d-1}k_l} u^{(d-2)}_{k_1,\dots,k_{d-2},n}\overline{u^{(d-2)}_{k_1,\dots,k_{d-2},n+k_{d-1}}}\right|,
\end{aligned}\tag{5-25}$$

and

$$\begin{aligned}
&\left|\sum_{n=1}^{N-\sum_{l=1}^{d-1}k_l} u^{(d-2)}_{k_1,\dots,k_{d-2},n}\overline{u^{(d-2)}_{k_1,\dots,k_{d-2},n+k_{d-1}}}\right| \\
&\qquad=\left|\sum_{n=1}^{N-\sum_{l=1}^{d-1}k_l}\exp\left\{2\pi i h_d\alpha\sum_{1\le l\le d-1}^{j_l=0,1}(-1)^{d-1-\sum_{l=1}^{d-1}j_l}\binom{n+\sum_{j=1}^{d-1}j_lk_l}{d}\right\}\right| \\
&\qquad=\left|\sum_{n=1}^{N-\sum_{l=1}^{d-1}k_l}\exp\left\{2\pi i h_d\alpha\sum_{0\le l\le d-1}^{j_l=0,1}(-1)^{d-\sum_{l=0}^{d-1}j_l}\binom{l_0n+\sum_{j=1}^{d-1}j_lk_l}{d}\right\}\right| \\
&\qquad=\left|\sum_{n=1}^{N-\sum_{l=1}^{d-1}k_l}\exp\left\{2\pi i h_d n\alpha\prod_{l=1}^{d-1}k_l\right\}\right| &&(5\text{-}26)\\
&\qquad\lesssim\frac{1}{\|h_d\alpha\prod_{l=1}^{d-1}k_l\|_{\mathbb{T}}}, &&(5\text{-}27)
\end{aligned}$$

where in (5-26) we used (5-3).

Since $\alpha\in \mathrm{DC}(\tau)$, by the property of the Diophantine condition (2-9) and since $|h_i|\le H_0,\ 1\le k_i\le H_i$, we have

$$\sum_{k_{d-1}=1}^{H_{d-1}}\frac{1}{\|h_d\alpha\prod_{l=1}^{d-1}k_l\|_{\mathbb{T}}}\lesssim\sum_{j=1}^{H_{d-1}}\frac{\prod_{l=0}^{d-1}H_l^{\tau}}{j}\le H_{d-1}^{\tau+\epsilon}\prod_{l=0}^{d-2}H_l^{\tau}.\tag{5-28}$$

Thus combining (5-25), (5-27) with (5-28), we have

$$\left|\frac{1}{N-\sum_{l=1}^{d-2}k_l}\sum_{n=1}^{N-\sum_{l=1}^{d-2}k_l}u^{(d-2)}_{k_1,\dots,k_{d-2},n}\right|^2\lesssim\frac{1}{H_{d-1}}+\frac{H_{d-1}^{\tau+\epsilon}\prod_{l=0}^{d-2}H_l^{\tau}}{H_{d-1}(N-\sum_{l=1}^{d-2}H_l)}\lesssim\frac{1}{H_{d-1}}=\frac{1}{H_{d-2}^2}.$$

Lemma 5.2. *For any $\alpha\in\mathbb{T}$, if, for any $1\le k_s\le H_s$,*

$$\left|\frac{1}{N-\sum_{l=1}^{s}k_l}\sum_{n=1}^{N-\sum_{l=1}^{s}k_l}u^{(s)}_{k_1,\dots,k_s,n}\right|^2\lesssim\frac{1}{H_s^2},$$

then for any $0 \le t \le s-1,\ 1 \le k_t \le H_t$, *we have*

$$\left|\frac{1}{N-\sum_{l=1}^{t} k_l}\sum_{n=1}^{N-\sum_{l=1}^{t} k_l} u^{(t)}_{k_1,\dots,k_t,n}\right|^2 \lesssim \frac{1}{H_t^2}.$$

Proof. For $t = s-1$, by (5-24),

$$\begin{aligned}\left|\frac{1}{N-\sum_{l=1}^{s-1} k_l}\sum_{n=1}^{N-\sum_{l=1}^{s-1} k_l} u^{(s-1)}_{k_1,\dots,k_{s-1},n}\right|^2 &\lesssim \frac{1}{H_s}+\frac{1}{\left(N-\sum_{l=1}^{s-1} k_l\right)H_s^2}\sum_{k_s=1}^{H_s}(H_s-k_s)\left(N-\sum_{l=1}^{s} k_l\right)\left|\frac{\sum_{n=1}^{N-\sum_{l=1}^{s} k_l} u^{(s)}_{k_1,\dots,k_s,n}}{\left(N-\sum_{l=1}^{s} k_l\right)}\right| \\ &\lesssim \frac{1}{H_s} = \frac{1}{H_{s-1}^2}.\end{aligned}$$

Then we proceed by reverse induction. □

At the final step we obtain

$$\left|\frac{1}{N}\sum_{n=1}^{N} u_n^{(0)}\right|^2 \lesssim \frac{1}{H_0^2}.$$

Plugging it into (5-13), we have

$$D(\vec{Y}_1,\dots,\vec{Y}_N) \lesssim \frac{1}{H_0} + \sum_{0<|\vec{h}|\le H_0} \frac{1}{r(\vec{h})}\frac{1}{H_0} \lesssim \frac{1}{H_0^{1-\epsilon}} \sim N^{-(1-\epsilon)/((2^d-1)(\tau+\epsilon))}.$$ □

Liouvillean α.

Proof of Lemma 3.9. For $\alpha \notin \mathrm{DC}(d)$, by property (2-11), we can find a subsequence $\{p_n/q_n\}$ of the continued fraction approximants of α such that $q_{n+1} > q_n^d$. In the following we will use q instead of q_n and $\tilde{q}$ instead of q_{n+1} for simplicity. Here we would like to show $D_q(\vec{Y}_1,\dots,\vec{Y}_q) \le q^{-\delta}$ for some $\delta > 0$. Take

$$H_j \sim q^{2^j/2^d} \quad \text{for } 0 \le j \le d-2 \qquad \text{and} \qquad H_{d-1} \sim q^{2^{d-1}(1+\epsilon)/2^d}, \tag{5-29}$$

where $\epsilon > 0$ is small enough such that

$$\prod_{l=0}^{d-1} H_l = q^{(2^d-1+2^{d-1}\epsilon)/2^d} < q. \tag{5-30}$$

Now by Lemma 2.5,

$$\begin{aligned}&D(\vec{Y}_1,\dots,\vec{Y}_q) \\ &\quad\le C_d\left(\frac{1}{H_0} + \sum_{0<|\vec{h}|\le H_0}\frac{1}{r(\vec{h})}\left|\frac{1}{q}\sum_{n=1}^{q}\exp\left\{2\pi i\sum_{j=1}^{d}(h_j\alpha + h_{j+1}y_1+\cdots+h_d y_{d-j})\binom{n}{j}\right\}\right|\right).\end{aligned} \tag{5-31}$$

Consider the difference

$$\begin{aligned}
&\frac{1}{q}\left|\sum_{n=1}^{q}\exp\left\{2\pi i\sum_{j=1}^{d}(h_j\alpha+h_{j+1}y_1+\cdots+h_d y_{d-j})\binom{n}{j}\right\}\right. \\
&\qquad\left.-\sum_{n=1}^{q}\exp\left\{2\pi i\sum_{j=1}^{d}\left(h_j\frac{p}{q}+h_{j+1}y_1+\cdots+h_d y_{d-j}\right)\binom{n}{j}\right\}\right| \\
&\le\frac{1}{q}\sum_{n=1}^{q}\left|\exp\left\{2\pi i\sum_{j=1}^{d}h_j\left(\alpha-\frac{p}{q}\right)\binom{n}{j}\right\}-1\right| \\
&\lesssim\frac{1}{q}\sum_{n=1}^{q}\sum_{j=1}^{d}\binom{n}{j}H_0\left|\alpha-\frac{p}{q}\right| \\
&\lesssim\frac{H_0}{q},
\end{aligned}\tag{5-32}$$

where in the last step we use (2-7),

$$\left|\alpha-\frac{p}{q}\right|\le\frac{1}{q\tilde{q}}<\frac{1}{q^{d+1}}.$$

Then combining (5-31) with (5-32), we have

$$D(\vec{Y}_1,\ldots,\vec{Y}_q)\lesssim C_d\left(\frac{1}{H_0}+\sum_{0<|\vec{h}|\le H_0}\frac{1}{r(\vec{h})}\left|\frac{1}{q}\sum_{n=1}^{q}u_n^{(0)}\right|\right)+\frac{H_0}{q},\tag{5-33}$$

where

$$\tilde{u}_n^{(0)}=\exp\left\{2\pi i\sum_{j=1}^{d}\left(h_j\frac{p}{q}+h_{j+1}y_1+\cdots+h_d y_{d-j}\right)\binom{n}{j}\right\},$$

that is, $u_n^{(0)}$ as in (5-14) with α replaced with p/q. Thus with $\tilde{u}^{(s)}_{k_1,\ldots,k_s,n}$ defined as in (5-16) with α replaced with p/q, similar to (5-25) and (5-26), we have

$$\begin{aligned}
&\left|\frac{1}{N-\sum_{l=1}^{d-2}k_l}\sum_{n=1}^{N-\sum_{l=1}^{d-2}k_l}\tilde{u}^{(d-2)}_{k_1,\ldots,k_{d-2},n}\right|^2 \\
&\qquad\lesssim\frac{1}{H_{d-1}}+\frac{1}{(N-\sum_{l=1}^{d-2}k_l)H_{d-1}}\sum_{k_{d-1}=1}^{H_{d-1}}\left|\sum_{n=1}^{N-\sum_{l=1}^{d-1}k_l}\tilde{u}^{(d-2)}_{k_1,\ldots,k_{d-2},n}\overline{\tilde{u}^{(d-2)}_{k_1,\ldots,k_{d-2},n+k_{d-1}}}\right|,
\end{aligned}\tag{5-34}$$

and

$$\begin{aligned}
\left|\sum_{n=1}^{q-\sum_{l=1}^{d-1}k_l}\tilde{u}^{(d-2)}_{k_1,\ldots,k_{d-2},n}\overline{\tilde{u}^{(d-2)}_{k_1,\ldots,k_{d-2},n+k_{d-1}}}\right|&=\left|\sum_{n=1}^{q-\sum_{l=1}^{d-1}k_l}\exp\left\{2\pi i h_d n\frac{p}{q}\prod_{l=1}^{d-1}k_l\right\}\right| \\
&\lesssim\frac{1}{\left\|h_d(p/q)\prod_{l=1}^{d-1}k_l\right\|_{\mathbb{R}/\mathbb{Z}}}.
\end{aligned}\tag{5-35}$$

Since $|h_d| \le H_0$, $1 \le k_i \le H_i$, and by (5-30), for any $1 \le k \le H_{d-1}$ we have

$$\left\| k h_d \frac{p}{q} \prod_{l=1}^{d-2} k_l \right\|_{\mathbb{R}/\mathbb{Z}} \ge \frac{1}{q}.$$

Thus

$$\sum_{k_{d-1}=1}^{H_{d-1}} \frac{1}{\left\| h_d (p/q) \prod_{l=1}^{d-1} k_l \right\|_{\mathbb{R}/\mathbb{Z}}} \lesssim \sum_{j=1}^{H_{d-1}} \frac{q}{j} \le q \ln H_{d-1}. \tag{5-36}$$

Then combining (5-34), (5-35) with (5-36), we get

$$\begin{aligned} \left| \frac{1}{q - \sum_{l=1}^{d-2} k_l} \sum_{n=1}^{q - \sum_{l=1}^{d-2} k_l} \tilde{u}^{(d-2)}_{k_1,\dots,k_{d-2},n} \right|^2 &\lesssim \frac{1}{H_{d-1}} + \frac{q \ln H_{d-1}}{\left(q - \sum_{l=1}^{d-2} H_l\right) H_{d-1}} \\ &\lesssim \frac{1}{H_{d-1}^{1/(1+\epsilon)}} = \frac{1}{H_{d-2}^2}. \end{aligned} \tag{5-37}$$

By Lemma 5.2,

$$\left| \frac{1}{q} \sum_{n=1}^{q} \tilde{u}^{(0)}_n \right|^2 \lesssim \frac{1}{H_0}.$$

Plugging it into (5-33), we get

$$D(\vec{Y}_1, \dots, \vec{Y}_q) \lesssim \frac{1}{H_0} + \frac{(\log H_0)^d}{H_0} + \frac{H_0}{q} \lesssim \frac{1}{q^{(1-\epsilon)/2^d}}. \qquad \square$$

6. Bounded remainder sets

Most of the material covered in this section comes from [Grepstad and Lev 2015]. We briefly discuss it here for completeness and readers' convenience. From now on we restrict our attention to irrational rotation on $\mathbb{T}^d$. For a measurable set $U \subset \mathbb{T}^d$, consider the function

$$A_N(U, \vec{x}) - N|U| := A(U, \{\vec{x} + n\alpha\}_{n=0}^{N-1}) - N|U| = \sum_{n=0}^{N-1} \chi_U(\vec{x} + n\alpha) - N|U|.$$

We will say U is a *bounded remainder set* (BRS) with respect to α if there exists a constant $C(U, \alpha) > 0$ such that

$$|A_N(U, \vec{x}) - N|U|| \le C(U, \alpha)$$

for any N and a.e. $\vec{x} \in \mathbb{T}^d$. We will call a measurable function g on $\mathbb{T}^d$ a transfer function for U if its characteristic function satisfies

$$\chi_U(\vec{x}) - |U| = g(\vec{x}) - g(\vec{x} - \alpha) \quad \text{a.e.}$$

Obviously if g is a transfer function for U, then its Fourier coefficients satisfy

$$\hat{g}(\vec{m}) = \frac{\hat{\chi}_U(\vec{m})}{1 - e^{-2\pi i \langle \vec{m}, \alpha \rangle}}, \quad \vec{m} \ne 0. \tag{6-1}$$

Proposition 6.1 [Grepstad and Lev 2015]. *For a measurable set $U \subset \mathbb{T}^d$, the following are equivalent*:

- *U is a bounded remainder set.*
- *U has a bounded transfer function g.*

Theorems 9, 10 and Corollary 6.2 are presented in [Grepstad and Lev 2015] without explicit bounds on the transfer functions. We present the proofs in order to extract the needed estimates.

Theorem 9. *Any interval $I \subset \mathbb{T}$ of length $0 < |q\alpha - p| < 1$ is a BRS with respect to α. Furthermore its transfer function g satisfies $\|g\|_\infty \le |q|$.*

Proof. Without loss of generality, we consider an interval $I = [0, \kappa]$, where $\kappa = q\alpha - p > 0$. Then

$$\begin{aligned}\chi_I(x) - |I| &= -\{x\} + \{x - \kappa\}\\ &= -\{x\} + \{x - q\alpha\}\\ &= (-\{x\} - \cdots - \{x - (q-1)\alpha\}) + (\{x - \alpha\} + \cdots + \{x - q\alpha\})\\ &= g(x) - g(x - \alpha),\end{aligned}$$

where $g(x) = -\sum_{j=0}^{q-1}\{x - j\alpha\}$, $\|g\|_\infty \le |q|$. □

Theorem 10. *Let $\vec{v} = (v_1, v_2, \ldots, v_d) = q\alpha - \vec{p} \in \mathbb{Z}\alpha + \mathbb{Z}^d$, $v \notin \mathbb{Z}^d$, and let $\Sigma \in \mathbb{T}^{d-1}$ be a BRS with respect to the vector $(v_1/v_d, v_2/v_d, \ldots, v_{d-1}/v_d)$ with transfer function h. Then the set*

$$U = U(\Sigma, \vec{v}) = \{(\vec{x}, 0) + t\vec{v} : \vec{x} \in \Sigma,\ 0 \le t < 1\}$$

is a BRS with respect to α, whose transfer function g satisfies $\|g\|_\infty \le |q|(\|h\|_\infty + 1)$.

Proof. Let $\vec{v}_0 = (v_1, \ldots, v_{d-1})$ be the vector in $\mathbb{T}^{d-1}$ which consists of the first $d-1$ entries of $\vec{v}$. First, we wish to find a bounded function $\tilde{g}$ on $\mathbb{T}^d$ satisfying the cohomological equation

$$\chi_U(\vec{x}, y) - |U| = \tilde{g}(\vec{x}, y) - \tilde{g}(\vec{x} - \vec{v}_0, y - v_d) \quad \text{for a.e. } (\vec{x}, y) \in \mathbb{T}^{d-1} \times \mathbb{T}.$$

This means the Fourier coefficients satisfy the equation

$$\hat{\tilde{g}}(\vec{m}, n)(1 - e^{-2\pi i(\langle \vec{m}, \vec{v}_0\rangle + n v_d)}) = \int_0^{v_d}\int_{\Sigma + (y/v_d)\vec{v}_0} e^{-2\pi i\langle \vec{m}, \vec{x} + (y/v_d)\vec{v}_0\rangle}\, \mathrm{d}\vec{x}\, e^{-2\pi i n y}\, \mathrm{d}y, \quad (\vec{m}, n) \ne (\vec{0}, 0), \tag{6-2}$$

which implies

$$\hat{\tilde{g}}(\vec{m}, n) = \frac{\hat{\chi}_\Sigma(\vec{m})}{2\pi i(\langle \vec{m}, \vec{v}_0\rangle / v_d + n)}, \quad (\vec{m}, n) \ne (\vec{0}, 0). \tag{6-3}$$

We know Σ is a BRS with respect to $\vec{v}_0/v_d$; by (6-1) its transfer function $h : \mathbb{T}^{d-1} \to \mathbb{R}$ satisfies

$$\hat{h}(\vec{m}) = \frac{\hat{\chi}_\Sigma(\vec{m})}{1 - e^{-2\pi i\langle \vec{m}, \vec{v}_0\rangle / v_d}}, \quad \vec{m} \ne 0.$$

It is straightforward to check that the bounded function $\tilde{g}$ defined by

$$\tilde{g}(\vec{x}, y) = h\left(\vec{x} - \frac{\vec{v}_0}{v_d}\{y\}\right) - |\Sigma| \cdot \{y\}$$

satisfies the cohomological equation (6-3). Hence $\tilde{g}$ is a bounded transfer function for U with respect to $\vec{v}$.

Indeed, $\|\tilde{g}\|_\infty \le \|h\|_\infty + 1$. Since $\vec{v} = q\alpha - \vec{p}$, letting $g(\vec{x}) = \tilde{g}(\vec{x}) + \tilde{g}(\vec{x} - \alpha) + \cdots + \tilde{g}(\vec{x} - (q-1)\alpha)$ we have that U is a BRS with respect to α with bounded transfer function g satisfying $\|g\|_\infty \le |q|\|\tilde{g}\|_\infty \le |q|(\|h\|_\infty + 1)$. □

The following corollary will be used several times in Section 7.

Corollary 6.2. *Let $U \subset \mathbb{T}^2$ be the parallelogram spanned by two vectors*

$$m(\alpha_1, \alpha_2) - (l_1, l_2) \quad \text{and} \quad \left(q\frac{m\alpha_1 - l_1}{m\alpha_2 - l_2} - p,\ 0\right).$$

Then U is a BRS with respect to (α_1, α_2) with transfer function g satisfying $\|g\|_\infty \le |m|(|q|+1) \le 2|mq|$.

Proof. In this case

$$v = (v_1, v_2) = m(\alpha_1, \alpha_2) - (l_1, l_2) \in \mathbb{Z}\alpha + \mathbb{Z}^2, \quad \Sigma = \left[0, q\frac{v_1}{v_2} - p\right] \times \{0\}.$$

We know the transfer function h of Σ with respect to v_1/v_2 satisfies $\|h\|_\infty \le |q|$. Thus $\|g\|_\infty \le |m|(|q|+1) \le 2|mq|$. □

7. 2-dimensional irrational rotation with weak Diophantine frequencies

In this section we deal with 2-dimensional weakly Diophantine frequencies. Our goal is to prove Lemma 3.5.

Proof of Lemma 3.5. Assume $(\alpha_1, \alpha_2) \in \mathrm{WDC}(c_0, \tau/4)$, for some $\tau > 4$ and $c_0 > 0$. We divide the discussion into two parts.

First, we introduce the coprime Diophantine condition:

$$\begin{aligned} \mathrm{PDC}(\tau) &= \bigcup_{c>0} \mathrm{PDC}(c, \tau) \\ &= \bigcup_{c>0} \left\{ (\alpha_1, \alpha_2) : \|\langle \vec{h}, \boldsymbol{\alpha} \rangle\|_{\mathbb{T}} \ge \frac{c}{|\vec{h}|^\tau} \text{ for any } \gcd(h_1, h_2) = 1 \text{ or } h_1 h_2 = 0 \text{ but } \vec{h} \ne \vec{0} \right\}. \end{aligned} \tag{7-1}$$

Obviously if $\alpha \in \mathrm{PDC}(c, \tau)$, both α_1 and α_2 belong to $\mathrm{DC}(c, \tau)$.

Next we will distinguish two different cases: PDC or non-PDC. Roughly speaking, in the PDC setting, we use bounded remainder sets technique presented in Section 6 and work directly with the 2-dimensional problem. In the non-PDC but WDC setting, we are able to reduce the 2-dimensional problem to the 1-dimensional problem, which is much easier to analyze.

Case A: $(\alpha_1, \alpha_2) \in \mathrm{PDC}(c_1, \tau)$ for some $c_1 > 0$. Let's take the best simultaneous approximation $\{(l_{1,n}/m_n, l_{2,n}/m_n)\}$ of (α_1, α_2). It has the following property.

Lemma 7.1 [Lagarias 1982, Theorem 3.5]. *If $\{1, \alpha_1, \alpha_2\}$ is linearly independent over $\mathbb{Q}$, then there are infinitely many n_k such that*

$$\begin{vmatrix} m_{n_k} & l_{1,n_k} & l_{2,n_k} \\ m_{n_k+1} & l_{1,n_k+1} & l_{2,n_k+1} \\ m_{n_k+2} & l_{1,n_k+2} & l_{2,n_k+2} \end{vmatrix} \neq 0.$$

Now we take $r_k > 0$ such that

$$m_{n_k} \le \frac{4}{\pi} r_k^{-2} < m_{n_k+1}. \tag{7-2}$$

By (2-12), the choice of r_k guarantees that for $n \ge n_k$,

$$(m_n\alpha_1 - l_{1,n}, m_n\alpha_2 - l_{2,n}) \in B_{r_k}(0,0), \tag{7-3}$$

where

$$B_r(x_1, x_2) := \{y = (y_1, y_2) \in \mathbb{T}^2 : \|y_1 - x_1\|_{\mathbb{T}}^2 + \|y_2 - x_2\|_{\mathbb{T}}^2 < r_k^2\}.$$

Let $\{p_{n,s}/q_{n,s}\}_{s=1}^{\infty}$ be the continued fraction approximants of $(m_n\alpha_1 - l_{1,n})/(m_n\alpha_2 - l_{2,n})$. For each n choose s_n such that

$$q_{n,s_n} \le r_k^{-1} < q_{n,s_n+1}. \tag{7-4}$$

By (2-7), the choice of s_n guarantees that

$$\left(q_{n,s_n}\frac{m_n\alpha_1 - l_{1,n}}{m_n\alpha_2 - l_{2,n}} - p_{n,s_n}, 0\right) \in B_{r_k}(0,0). \tag{7-5}$$

By (2-12) and (2-14) we have

$$\frac{c_0}{m_n^{\tau/4}} \le \max\{|m_n\alpha_1 - l_{1,n}|, |m_n\alpha_2 - l_{2,n}|\} \le \frac{2}{\sqrt{\pi}\sqrt{m_{n+1}}}, \tag{7-6}$$

and by (7-2) we have $m_{n_k} \le (4/\pi) r_k^{-2}$. Thus

$$\max\,(m_{n_k}, m_{n_k+1}, m_{n_k+2}) \le C_{c_0,\tau} r_k^{-\tau^2/2}. \tag{7-7}$$

We have:

Lemma 7.2. *For some $n \in \{n_k, n_k+1, n_k+2\}$, we have $q_{n,s_n+1} \le r_k^{-2\tau^4}$.*

Let us postpone the proof of this lemma and finish the proof of Case A first.

Let U be the parallelogram spanned by the two vectors

$$m_n(\alpha_1, \alpha_2) - (l_{1,n}, l_{2,n}) \quad \text{and} \quad \left(q_{n,s_n}\frac{m_n\alpha_1 - l_{1,n}}{m_n\alpha_2 - l_{2,n}} - p_{n,s_n},\ 0\right).$$

By (7-3) and (7-5), $U\subset B_{2r_k}(0,0)$. Corollary 6.2 implies

$$\left|\sum_{j=0}^{M-1}\chi_U(x+j\alpha_1,y+j\alpha_2)-M|U|\right|\le 4|m_nq_{n,s_n}|$$

for a.e. (x,y). Thus as long as $M>4|m_nq_{n,s_n}|/|U|$, we have $\bigcup_{j=0}^{M-1}U-(j\alpha_1,j\alpha_2)$ covers the whole $\mathbb{T}^2$ up to a measure zero set. Then

$$\mathbb{T}^2\subseteq\bigcup_{j=0}^{M-1}B_{2r_k}(-j\alpha_1,-j\alpha_2)\quad\text{for } M>\frac{4|m_nq_{n,s_n}|}{|U|}. \tag{7-8}$$

Now we want to estimate $|U|$. Since $\alpha_2\in\mathrm{DC}(c_1,\tau)$, by (2-9) we have

$$|U|=|m_n\alpha_2-l_{2,n}|\cdot\left|q_{n,s_n}\frac{m_n\alpha_1-l_{1,n}}{m_n\alpha_2-l_{2,n}}-p_{n,s_n}\right|\ge\frac{c_1}{|m_n|^\tau}\frac{1}{2q_{n,s_n+1}}.$$

Thus by (7-4) and (7-7),

$$\frac{4|m_n|q_{n,s_n}}{|S|}\le\frac{8}{c_1}|m_n|^{1+\tau}q_{n,s_n}q_{n,s_n+1}\le C_{c_0,c_1,\tau}r_k^{-3\tau^4}.$$

This means it takes $B_{2r_k}(0,0)$ at most $C_{\alpha_1,\alpha_2,\tau}r_k^{-3\tau^4}$ steps to cover the whole $\mathbb{T}^2$.

Proof of Lemma 7.2. We will show it is impossible to have $q_{n,s_n+1}>r_k^{-2\tau^4}$ for all $n\in\{n_k,n_k+1,n_k+2\}$. In this case by (2-7), (2-12) and (7-2), we have

$$\begin{aligned}|q_{n,s_n}m_n\alpha_1-p_{n,s_n}m_n\alpha_2+M_n|&=|m_n\alpha_2-l_{2,n}|\cdot\left|q_{n,s_n}\frac{m_n\alpha_1-l_{1,n}}{m_n\alpha_2-l_{2,n}}-p_{n,s_n}\right|\\&<\frac{2}{\sqrt{\pi}\sqrt{|m_{n+1}|}q_{n,s_n}}<r_k^{2\tau^4+1},\end{aligned} \tag{7-9}$$

where $M_n=p_{n,s_n}l_{2,n}-q_{n,s_n}l_{1,n}$.

We have the following estimates on the upper bounds of p_{n,s_n} and M_n. Combining (2-9), (7-2), (7-4), (7-6) with (7-7),

$$|p_{n,s_n}|\le q_{n,s_n}\left|\frac{m_n\alpha_1-l_{1,n}}{m_n\alpha_2-l_{2,n}}\right|+\frac{1}{q_{n,s_n+1}}\le\frac{2q_{n,s_n}|m_n|^\tau}{c_1\sqrt{\pi}\sqrt{|m_{n+1}|}}+r_k^{2\tau^4}\le C_{c_0,c_1,\tau}r_k^{-\tau^3/2}. \tag{7-10}$$

By (7-9), (7-2), (7-7), (7-4) and (7-10),

$$|M_n|<|q_{n,s_n}m_n\alpha_1-p_{n,s_n}m_n\alpha_2|+r_k^{2\tau^4}\le C_{c_0,c_1,\tau}r_k^{-\tau^3}. \tag{7-11}$$

<u>Case 1</u>: If $p_{n,s_n}=0$ for some $n\in\{n_k,n_k+1,n_k+2\}$, then by (2-7), (2-12) and (7-1), (2-9),(7-2), (7-7), we have

$$r_k^{2\tau^4}>\frac{1}{q_{n,s_n+1}}\ge\left|q_{n,s_n}\frac{m_n\alpha_1-l_{1,n}}{m_n\alpha_2-l_{2,n}}\right|\ge\frac{c_1\sqrt{\pi}\sqrt{|m_{n+1}|}}{2m_n^\tau}\ge C_{c_0,c_1,\tau}r_k^{\tau^3/2+1},$$

which is a contradiction.

Case 2: If $M_n = 0$ for some $n \in \{n_k, n_k+1, n_k+2\}$, then by (7-9), (7-2), (7-10), and the fact that $(\alpha_1, \alpha_2) \in \mathrm{PDC}(c_1, \tau)$, we have

$$r_k^{2\tau^4} > |m_n|\,|q_{n,s_n}\alpha_1 - p_{n,s_n}\alpha_2| \ge \frac{c_1|m_n|}{\max(p_{n,s_n}, q_{n,s_n})^\tau} \ge C_{c_0,c_1,\tau} r_k^{\tau^4/2},$$

again a contradiction.

Case 3: If $p_{n,s_n} \neq 0$ and $M_n \neq 0$ for any $n \in \{n_k, n_k+1, n_k+2\}$, then for any $i, j \in \{n_k, n_k+1, n_k+2\}$, we have

$$\begin{aligned}
&|(q_{i,s_i}m_iM_j - q_{j,s_j}m_jM_i)\alpha_1 - (p_{i,s_i}m_iM_j - p_{j,s_j}m_jM_i)\alpha_2| \\
&\qquad \le |(q_{i,s_i}m_i\alpha_1 - p_{i,s_i}m_i\alpha_2 + M_i)M_j| + |(q_{j,s_j}m_j\alpha_1 - p_{j,s_j}m_j\alpha_2 + M_j)M_i| \\
&\qquad < (|M_i| + |M_j|)r_k^{2\tau^4}.
\end{aligned} \tag{7-12}$$

Case 3.1: $(q_{i,s_i}m_iM_j - q_{j,s_j}m_jM_i,\ p_{i,s_i}m_iM_j - p_{j,s_j}m_jM_i) \neq (0,0)$ for some $i, j \in \{n_k, n_k+1, n_k+2\}$. In this case let $h = \gcd(q_{i,s_i}m_iM_j - q_{j,s_j}m_jM_i,\ p_{i,s_i}m_iM_j - p_{j,s_j}m_jM_i)$ be the greatest common divisor of the two numbers if they are both nonzero, and $h = 1$ otherwise. Then by (7-12),

$$\left|\frac{q_{i,s_i}m_iM_j - q_{j,s_j}m_jM_i}{h}\alpha_1 - \frac{p_{i,s_i}m_iM_j - p_{j,s_j}m_jM_i}{h}\alpha_2\right| < \frac{|M_i| + |M_j|}{h} r_k^{2\tau^4}.$$

However on one hand by (7-11),

$$\frac{|M_i| + |M_j|}{h} r_k^{2\tau^4} \le (|M_i| + |M_j|) r_k^{2\tau^4} \le C_{c_0,c_1,\tau} r_k^{2\tau^4 - \tau^3}.$$

On the other hand, by the fact that $(\alpha_1, \alpha_2) \in \mathrm{PDC}(c_1, \tau)$ and (7-2), (7-7), (7-10), (7-11),

$$\begin{aligned}
&\left|\frac{q_{i,s_i}m_iM_j - q_{j,s_j}m_jM_i}{h}\alpha_1 - \frac{p_{i,s_i}m_iM_j - p_{j,s_j}m_jM_i}{h}\alpha_2\right| \\
&\qquad \ge \frac{c_1 h^\tau}{|(q_{i,s_i}m_iM_j - q_{j,s_j}m_jM_i,\ p_{i,s_i}m_iM_j - p_{j,s_j}m_jM_i)|^\tau} \\
&\qquad \ge C_{c_0,c_1,\tau} r_k^{7\tau^4/4},
\end{aligned}$$

a contradiction.

Case 3.2: For any $i, j \in \{n_k, n_k+1, n_k+2\}$

$$q_{i,s_i}m_iM_j = q_{j,s_j}m_jM_i,$$
$$p_{i,s_i}m_iM_j = p_{j,s_j}m_jM_i.$$

Then for $n = n_k$,

$$\frac{p_{n,s_n}}{q_{n,s_n}} = \frac{p_{n+1,s_{n+1}}}{q_{n+1,s_{n+1}}} = \frac{p_{n+2,s_{n+2}}}{q_{n+2,s_{n+2}}}.$$

Hence we can let $p = p_{n,s_n} = p_{n+1,s_{n+1}} = p_{n+2,s_{n+2}}$ and $q = q_{n,s_n} = q_{n+1,s_{n+1}} = q_{n+2,s_{n+2}}$. Then we have (after plugging in $M_n = ql_{1,n} - pl_{2,n}$)

$$q(m_nl_{1,n+1} - m_{n+1}l_{1,n}) = p(m_nl_{2,n+1} - m_{n+1}l_{2,n}), \tag{7-13}$$

$$q(m_n l_{1,n+2} - m_{n+2} l_{1,n}) = p(m_n l_{2,n+2} - m_{n+2} l_{2,n}), \tag{7-14}$$
$$q(m_{n+1} l_{1,n+2} - m_{n+2} l_{1,n+1}) = p(m_{n+1} l_{2,n+2} - m_{n+2} l_{2,n+1}). \tag{7-15}$$

Then considering (7-13) $\cdot (-l_{1,n+2})$ + (7-14) $\cdot\, l_{1,n+1}$ + (7-15) $\cdot (-l_{1,n})$, we get

$$p \cdot \begin{vmatrix} m_{n_k} & l_{1,n_k} & l_{2,n_k} \\ m_{n_k+1} & l_{1,n_k+1} & l_{2,n_k+1} \\ m_{n_k+2} & l_{1,n_k+2} & l_{2,n_k+2} \end{vmatrix} = q \cdot 0 = 0,$$

a contradiction with the choice of n_k. □

Case B: $(\alpha_1, \alpha_2) \notin \mathrm{PDC}(\tau)$. By the definition of $\mathrm{PDC}(\tau)$, the sequence $\vec{h}_n = (h_{1,n}, h_{2,n})$ for which (7-1) fails has to satisfy either $\gcd(h_{1,n}, h_{2,n}) = 1$ (Case B.1) or $h_{1,n} h_{2,n} = 0$ (Case B.2).

Case B.1: We can find a sequence $\{n_j\}$ such that $|\vec{h}_{n_j}| = \max(|h_{1,n_j}|, |h_{2,n_j}|) \to \infty$ as $j \to \infty$, $\gcd(h_{1,n_j}, h_{2,n_j}) = 1$ and

$$\|h_{1,n_j}\alpha_1 + h_{2,n_j}\alpha_2\|_{\mathbb{T}} < \frac{1}{|\vec{h}_{n_j}|^{\tau}}.$$

Without loss of generality, we can assume $|h_{1,n_j}| = |\vec{h}_{n_j}|$. In this case we can take $r_{n_j} = 1/|h_{1,n_j}|$. For simplicity we will denote n_j by n.

Now that $\|h_{1,n}\alpha_1 + h_{2,n}\alpha_2\|_{\mathbb{T}} < 1/|h_{1,n}|^{\tau}$, we can find $l_{1,n}, l_{2,n} \in \mathbb{Z}$ such that $|h_{1,n}(\alpha_1 - l_{1,n}) + h_{2,n}(\alpha_2 - l_{2,n})| < 1/|h_{1,n}|^{\tau}$. Since replacing (α_1, α_2) with $(\alpha_1 + l_{1,n}, \alpha_2 + l_{2,n})$ does not change anything, we will assume $|h_{1,n}\alpha_1 + h_{2,n}\alpha_2| < 1/|h_{1,n}|^{\tau}$. Then

$$\left| \frac{\alpha_2}{\alpha_1} - \left(-\frac{h_{1,n}}{h_{2,n}} \right) \right| < \frac{1}{|h_{1,n}|^{\tau}\alpha_1}. \tag{7-16}$$

We consider the following two lines on $\mathbb{T}^2$:

$$l_1(t) = \left(\{t\}, \left\{ \frac{\alpha_2}{\alpha_1} t \right\} \right) \quad \text{and} \quad l_2(t) = \left(\{t\}, \left\{ -\frac{h_{1,n}}{h_{2,n}} t \right\} \right).$$

These two lines are close to each other in the sense that for $|t| \le |h_{1,n}|^{3\tau/4}$, by (7-16),

$$\left\| \left\{ \frac{\alpha_2}{\alpha_1} t \right\} - \left\{ -\frac{h_{1,n}}{h_{2,n}} t \right\} \right\|_{\mathbb{T}} \le \left| \frac{\alpha_2}{\alpha_1} t + \frac{h_{1,n}}{h_{2,n}} t \right| \le \frac{|t|}{|h_{1,n}|^{\tau}\alpha_1} \le \frac{1}{|h_{1,n}|^{\tau/4}\alpha_1}.$$

The graph of $l_2(t)$ is the hypotenuse of a right triangle with two legs of lengths $|h_{1,n}|$ and $|h_{2,n}|$ (mod $\mathbb{Z}^2$). We consider the orbit of $(\alpha_1, -(h_{1,n}/h_{2,n})\alpha_1)$ under the rotation $(\alpha_1, -(h_{1,n}/h_{2,n})\alpha_1)$. These points lie on $l_2(t)$. Under this rotation the point moves a distance $(\sqrt{h_{1,n}^2 + h_{2,n}^2}/|h_{2,n}|)\alpha_1$ at each step by a big interval with length $\sqrt{h_{1,n}^2 + h_{2,n}^2}$. Let $\{p_m/q_m\}_{m=1}^{\infty}$ be the continued fraction approximants of $\alpha_1/h_{2,n}$. Choose m such that

$$q_{m-1} \le |h_{1,n}|\sqrt{h_{1,n}^2 + h_{2,n}^2} < q_m. \tag{7-17}$$

Then it would take a point on $\mathbb{T}$ at most $q_m + q_{m-1}$ steps (under the $(\alpha_1/h_{2,n})$-rotation) to enter each interval of length $1/(|h_{1,n}|\sqrt{h_{1,n}^2 + h_{2,n}^2})$ on $\mathbb{T}$, see, e.g., [Jitomirskaya and Last 2000], which means it would take a point on $l_2(t)$ at most $q_m+q_{m-1}-1$ steps (under the $(\sqrt{h_{1,n}^2+h_{2,n}^2}\alpha_1/|h_{2,n}|)$-rotation) to enter each interval of length $1/|h_{1,n}|=r_n$ on the graph of $l_2(t)$. Moreover, it is easy to see that the distance from any $x\in\mathbb{T}^2$ to $l_2(t)$ is bounded by $1/\sqrt{h_{1,n}^2+h_{2,n}^2}<r_n$. Thus

$$\mathbb{T}^2\subseteq\bigcup_{k=0}^{q_m+q_{m-1}} B_{2r_n}\Big(k\alpha_1,-\frac{h_{1,n}}{h_{2,n}}k\alpha_1\Big). \tag{7-18}$$

By (2-7) and (7-16),

$$\Big|p_{m-1}+q_{m-1}\frac{\alpha_2}{h_{1,n}}\Big|=\Big|p_{m-1}-q_{m-1}\frac{\alpha_1}{h_{2,n}}+q_{m-1}\Big(\frac{\alpha_1}{h_{2,n}}+\frac{\alpha_2}{h_{1,n}}\Big)\Big|\le\frac{1}{q_m}+\frac{q_{m-1}}{|h_{1,n}|^{\tau-1}}.$$

This implies, by (2-7) and (7-17),

$$\begin{aligned}\|q_{m-1}\alpha_1\|_{\mathbb{T}}&\le|q_{m-1}\alpha_1-h_{2,n}p_{m-1}|\le\frac{|h_{2,n}|}{q_m},\\ \|q_{m-1}\alpha_2\|_{\mathbb{T}}&\le\frac{|h_{1,n}|}{q_m}+\frac{2}{|h_{1,n}|^{\tau-4}}.\end{aligned}$$

Then by the fact that $\alpha\in\mathrm{WDC}(c_0,\tau/4)$ and (7-17),

$$\max\Big\{\frac{|h_{2,n}|}{q_m},\frac{|h_{1,n}|}{q_m}+\frac{2}{|h_{1,n}|^{\tau-4}}\Big\}\ge\max\big(\|q_{m-1}\alpha_1\|_{\mathbb{T}},\|q_{m-1}\alpha_2\|_{\mathbb{T}}\big)\ge\frac{c_0}{q_{m-1}^{\tau/4}}\ge\frac{c_0}{2^{\tau/4}|h_{1,n}|^{\tau/2}}.$$

This implies

$$q_m+q_{m-1}<2q_m\le\frac{2^{\tau/4+2}}{c_0}|h_{1,n}|^{\tau/2+1}. \tag{7-19}$$

Since

$$0\le k\le\frac{2^{\tau/4+2}}{c_0}|h_{1,n}|^{\tau/2+1}<r_n^{-3\tau/4},$$

by (7-16) the points $(k\alpha_1,k\alpha_2)$ and $(k\alpha_1,-(h_{1,n}/h_{2,n})k\alpha_1)$ differ at most by $r_n^{\tau/4}$, so we obtain using (7-18) and (7-19),

$$\mathbb{T}^2\subseteq\bigcup_{k=0}^{r_n^{-3\tau/4}}B_{3r_n}(k\alpha_1,k\alpha_2).$$

<u>Case B.2</u>: We can find a sequence $\{n_j\}$ such that $h_{2,n_j}\equiv 0$ and $|h_{1,n_j}|\to\infty$ such that

$$\|h_{1,n_j}\alpha_1\|_{\mathbb{T}}<\frac{1}{|h_{1,n_j}|^{\tau}}. \tag{7-20}$$

For simplicity we will replace n_j with n. We can find M_n such that $|h_{1,n}\alpha_1-M_n|<1/|h_{1,n}|^{\tau}$. Let $d_n=\gcd(h_{1,n},M_n)$ be the greatest common divisor. Let $\tilde h_{1,n}=h_{1,n}/d_n$ and $\widetilde M_n=M_n/d_n$. We have

$$\Big|\alpha_1-\frac{\widetilde M_n}{\tilde h_{1,n}}\Big|<\frac{1}{|h_{1,n}|^{\tau+1}}\to 0. \tag{7-21}$$

If $\tilde{h}_{1,n}$ is bounded in n, then α_1 can be approximated arbitrarily closely by rationals with bounded denominators, which is impossible. Thus $|\tilde{h}_{1,n}| \to \infty$. Now take radius $r_n = 1/|\tilde{h}_{1,n}|$. For each $0 \le i \le \tilde{h}_{1,n} - 1$ consider $\{(i\alpha_1 + k\tilde{h}_{1,n}\alpha_1,\ i\alpha_2 + k\tilde{h}_{1,n}\alpha_2)\}_{k=0}^{\infty}$. Let $\{p_m/q_m\}_{m=1}^{\infty}$ be the continued fraction approximants of $\tilde{h}_{1,n}\alpha_2$. Choose m such that

$$q_{m-1} \le |\tilde{h}_{1,n}| = r_n^{-1} < q_m. \tag{7-22}$$

Then it takes any point on $\mathbb{T}$ at most $q_m + q_{m-1} - 1$ steps (under the $\tilde{h}_{1,n}\alpha_2$–rotation) to enter each interval of length r_n; see, e.g., [Jitomirskaya and Last 2000]. By (2-7),

$$|p_{m-1} - q_{m-1}\tilde{h}_{1,n}\alpha_2| \le \frac{1}{q_m}. \tag{7-23}$$

By (7-20), (7-22) and since $\tau > 4$, we have

$$\|q_{m-1}\tilde{h}_{1,n}\alpha_1\| \le \frac{q_{m-1}}{|\tilde{h}_{1,n}|^{\tau}} < \frac{c_0}{(q_{m-1}|\tilde{h}_{1,n}|)^{\tau/4}}.$$

By the fact that $\alpha \in \mathrm{WDC}(c_0, \tau/4)$,

$$\|q_{m-1}\tilde{h}_{1,n}\alpha_2\| \ge \frac{c_0}{(q_{m-1}|\tilde{h}_{1,n}|)^{\tau/4}}.$$

By (7-23) and (7-22), we have

$$q_m \le \frac{1}{c_0}|\tilde{h}_{1,n}|^{\tau/2}. \tag{7-24}$$

Now for $0 \le k \le q_m + q_{m-1} - 1$, by (7-21), (7-20) and (7-24),

$$\left\| i\alpha_1 + k\tilde{h}_{1,n}\alpha_1 - \frac{i\widetilde{M}_n}{\tilde{h}_{1,n}} \right\|_{\mathbb{T}} \le \frac{C}{|\tilde{h}_{1,n}|^{\tau/2}} = C r_n^{\tau/2}.$$

Since $\gcd(\tilde{h}_{1,n}, \widetilde{M}_n) = 1$, any interval of length $r_n = 1/|\tilde{h}_{1,n}|$ contains $i\widetilde{M}_n/\tilde{h}_{1,n}$ for some $0 \le i \le \tilde{h}_{1,n} - 1$. Thus

$$\mathbb{T}^2 \subseteq \bigcup_{k=0}^{(q_m+q_{m-1})|\tilde{h}_{1,n}|} B_{r_n}(k\alpha_1, k\alpha_2).$$

By (7-24), $(q_m + q_{m-1})|\tilde{h}_{1,n}| \le r_n^{-\tau}$, so we have

$$\mathbb{T}^2 \subseteq \bigcup_{k=0}^{r_n^{-\tau}} B_{r_n}(k\alpha_1, k\alpha_2), \tag{7-25}$$

completing the proof of Case B.2 and thus of Lemma 3.5. □

Appendix

Proof of Lemma 3.7. For sufficiently small $\epsilon > 0$, fix an integer $H_0 \sim N^{1/(d(\tau-1)+1+d\epsilon)}$, define $g(n) = 1/(n(n+1))$ for $1 \le n < H_0$ and $g(H_0) = 1/H_0$. For $(n_1, \ldots, n_d) \in \mathbb{Z}^d$ with $1 \le n_i \le H_0$, define $f(n_1,\dots,n_d)=\prod_{i=1}^d g(n_i)$. By Lemma 2.5, we have

$$\begin{aligned}
D_N(\theta) &\le C_d\left(\frac{1}{H_0}+\sum_{0<|h|\le H_0}\frac{1}{r(\vec{h})}\left|\frac{1}{N}\sum_{n=1}^{N}e^{2\pi i\langle\vec{h},\alpha\rangle n}\right|\right)\\
&\le \widetilde{C}_d\left(\frac{1}{H_0}+\frac{1}{N}\sum_{0<|h|\le H_0}\frac{1}{r(\vec{h})}\frac{1}{\|\langle\vec{h},\alpha\rangle\|_{\mathbb{T}}}\right)\\
&= \widetilde{C}_d\left(\frac{1}{H_0}+\frac{1}{N}\sum_{n_1,\dots,n_d=1}^{H_0}f(n_1,\dots,n_d)\sum_{\vec{h}=(h_1,\dots,h_d)\neq\vec{0},\,|h_j|\le n_j}\frac{1}{\|\langle\vec{h},\alpha\rangle\|_{\mathbb{T}}}\right)\\
&\le \widetilde{C}_d\left(\frac{1}{H_0}+\frac{1}{N}\sum_{n_1,\dots,n_d=1}^{H_0}f(n_1,\dots,n_d)\sum_{j=1}^{3^d r(\vec{n})}\frac{r(\vec{n})^\tau}{j}\right)\\
&\le \widetilde{C}_d\left(\frac{1}{H_0}+\frac{1}{N}\sum_{n_1,\dots,n_d=1}^{H_0}f(n_1,\dots,n_d)r(\vec{n})^\tau\log r(\vec{n})\right)\\
&\le \widetilde{C}_d\left(\frac{1}{H_0}+\frac{{H_0}^{d(\tau-1+\epsilon)}}{N}\right)\\
&\lesssim N^{-1/(d(\tau-1)+1+d\epsilon)}. \qquad\square
\end{aligned}$$

Acknowledgement

Han would like to thank Anton Gorodetski for valuable discussions. This research was partially supported by the NSF grants DMS-1401204 and DMS-1800689. Jitomirskaya would like to thank the support of the Simons Foundation where she was a Fellow in 2014–15. We are also grateful to the Isaac Newton Institute for Mathematical Sciences, Cambridge, for its hospitality, supported by EPSRC grant number EP/K032208/1, during the programme Periodic and Ergodic Spectral Problems, where this work was started.

References

[Avila et al. 2014] A. Avila, D. Damanik, and Z. Zhang, "Singular density of states measure for subshift and quasi-periodic Schrödinger operators", *Comm. Math. Phys.* **330**:2 (2014), 469–498. MR Zbl

[Berger 2003] M. Berger, *A panoramic view of Riemannian geometry*, Springer, 2003. MR Zbl

[Bourgain and Goldstein 2000] J. Bourgain and M. Goldstein, "On nonperturbative localization with quasi-periodic potential", *Ann. of Math.* (2) **152**:3 (2000), 835–879. MR Zbl

[Bourgain and Jitomirskaya 2000] J. Bourgain and S. Jitomirskaya, "Anderson localization for the band model", pp. 67–79 in *Geometric aspects of functional analysis*, edited by V. D. Milman and G. Schechtman, Lecture Notes in Math. **1745**, Springer, 2000. MR Zbl

[Bourgain et al. 2001] J. Bourgain, M. Goldstein, and W. Schlag, "Anderson localization for Schrödinger operators on $\mathbb{Z}$ with potentials given by the skew-shift", *Comm. Math. Phys.* **220**:3 (2001), 583–621. MR Zbl

[do Carmo 1992] M. P. do Carmo, *Riemannian geometry*, Birkhäuser, Boston, 1992. MR Zbl

[Damanik and Tcheremchantsev 2003] D. Damanik and S. Tcheremchantsev, "Power-law bounds on transfer matrices and quantum dynamics in one dimension", *Comm. Math. Phys.* **236**:3 (2003), 513–534. MR Zbl

[Damanik and Tcheremchantsev 2007] D. Damanik and S. Tcheremchantsev, "Upper bounds in quantum dynamics", *J. Amer. Math. Soc.* **20**:3 (2007), 799–827. MR Zbl

[Damanik and Tcheremchantsev 2008] D. Damanik and S. Tcheremchantsev, "Quantum dynamics via complex analysis methods: general upper bounds without time-averaging and tight lower bounds for the strongly coupled Fibonacci Hamiltonian", *J. Funct. Anal.* **255**:10 (2008), 2872–2887. MR Zbl

[Del Rio et al. 1994] R. Del Rio, N. Makarov, and B. Simon, "Operators with singular continuous spectrum, II: Rank one operators", *Comm. Math. Phys.* **165**:1 (1994), 59–67. MR Zbl

[Grepstad and Lev 2015] S. Grepstad and N. Lev, "Sets of bounded discrepancy for multi-dimensional irrational rotation", *Geom. Funct. Anal.* **25**:1 (2015), 87–133. MR Zbl

[Jitomirskaya and Last 1999] S. Jitomirskaya and Y. Last, "Power-law subordinacy and singular spectra, I: Half-line operators", *Acta Math.* **183**:2 (1999), 171–189. MR Zbl

[Jitomirskaya and Last 2000] S. Y. Jitomirskaya and Y. Last, "Power law subordinacy and singular spectra, II: Line operators", *Comm. Math. Phys.* **211**:3 (2000), 643–658. MR Zbl

[Jitomirskaya and Mavi 2017] S. Jitomirskaya and R. Mavi, "Dynamical bounds for quasiperiodic Schrödinger operators with rough potentials", *Int. Math. Res. Not.* **2017**:1 (2017), 96–120. MR Zbl

[Jitomirskaya and Zhang 2015] S. Jitomirskaya and S. Zhang, "Quantitative continuity of singular continuous spectral measures and arithmetic criteria for quasiperiodic Schrödinger operators", preprint, 2015. arXiv

[Jitomirskaya et al. 2003] S. Jitomirskaya, H. Schulz-Baldes, and G. Stolz, "Delocalization in random polymer models", *Comm. Math. Phys.* **233**:1 (2003), 27–48. MR Zbl

[Khinchin 1964] A. Y. Khinchin, *Continued fractions*, The University of Chicago Press, 1964. MR

[Killip et al. 2003] R. Killip, A. Kiselev, and Y. Last, "Dynamical upper bounds on wavepacket spreading", *Amer. J. Math.* **125**:5 (2003), 1165–1198. MR Zbl

[Koksma 1950] J. F. Koksma, *Some theorems on Diophantine inequalities*, Stichting Mathematisch Centrum **5**, Math. Centrum Amsterdam, 1950. MR Zbl

[Kuipers and Niederreiter 1974] L. Kuipers and H. Niederreiter, *Uniform distribution of sequences*, Wiley-Interscience, New York, 1974. Pure and Applied Mathematics. MR Zbl

[Lagarias 1982] J. C. Lagarias, "Best simultaneous Diophantine approximations, II: Behavior of consecutive best approximations", *Pacific J. Math.* **102**:1 (1982), 61–88. MR Zbl

[Last 1996] Y. Last, "Quantum dynamics and decompositions of singular continuous spectra", *J. Funct. Anal.* **142**:2 (1996), 406–445. MR Zbl

[Nicolaescu 2007] L. I. Nicolaescu, *Lectures on the geometry of manifolds*, 2nd ed., World Scientific, Hackensack, NJ, 2007. MR Zbl

[del Rio et al. 1996] R. del Rio, S. Jitomirskaya, Y. Last, and B. Simon, "Operators with singular continuous spectrum, IV: Hausdorff dimensions, rank one perturbations, and localization", *J. Anal. Math.* **69** (1996), 153–200. MR Zbl

[Simon 2007] B. Simon, "Equilibrium measures and capacities in spectral theory", *Inverse Probl. Imaging* **1**:4 (2007), 713–772. MR Zbl

[Yomdin 1987] Y. Yomdin, "Volume growth and entropy", *Israel J. Math.* **57**:3 (1987), 285–300. MR Zbl

Received 6 Dec 2016. Revised 30 May 2018. Accepted 5 Jul 2018.

RUI HAN: rui.han@math.gatech.edu
Department of Mathematics, Georgia Institute of Technology, Atlanta, GA, United States

SVETLANA JITOMIRSKAYA: szhitomi@math.uci.edu
Department of Mathematics, University of California, Irvine, CA, United States

ANALYSIS AND PDE
Vol. 12, No. 4, 2019
dx.doi.org/10.2140/apde.2019.12.903

TWO-DIMENSIONAL GRAVITY WATER WAVES WITH CONSTANT VORTICITY I: CUBIC LIFESPAN

MIHAELA IFRIM AND DANIEL TATARU

This article is concerned with the incompressible, infinite-depth water wave equation in two space dimensions, with gravity and constant vorticity but with no surface tension. We consider this problem expressed in position-velocity potential holomorphic coordinates, and prove local well-posedness for large data, as well as cubic lifespan bounds for small-data solutions.

1. Introduction

The motion of water in contact with air is well described by the incompressible Euler equations in the fluid domain, combined with two boundary conditions on the free surface, i.e., the interface with air. In special, but still physically relevant cases, the equations of motion can be viewed as evolution equations for the free surface. These equations are commonly referred to as the water wave equations. Most notably, this is the case for irrotational flows. However, in two space dimensions there is a natural extension of these equations to flows with constant vorticity.

In previous work [Hunter, Ifrim and Tataru 2016; Ifrim and Tataru 2016] we have considered the local and long-time behavior for the irrotational gravity wave equations with infinite depth in two space dimensions. In this article we take a first step toward understanding the more difficult question of the long-time behavior of gravity waves with infinite depth and constant vorticity, either in $\mathbb{R} \times \mathbb{R}$ or in the periodic case $\mathbb{R} \times \mathbb{T}$. We begin by establishing a local well-posedness result. Then we consider the lifespan of small-data solutions, where, like in the zero vorticity case [Hunter, Ifrim and Tataru 2016],

M. Ifrim was supported by the Simons Foundation. D. Tataru was partially supported by the NSF grant DMS-1266182 as well as by a Simons Investigator Grant from the Simons Foundation.
MSC2010: primary 35Q35; secondary 42B37, 76B15.
Keywords: gravity waves, constant vorticity, normal forms, modified energy.

we are able to prove cubic lifespan bounds. To the best of our knowledge, this is the first long-time well-posedness result for this problem.

We remark that it is of further interest to consider small localized data, and establish an almost global in time result, as it was done in the irrotational case in [Hunter, Ifrim and Tataru 2016], improving an earlier result of [Wu 2009]. However, in the constant vorticity case this problem presents some interesting new challenges. We hope to consider this in subsequent work.

The motivation to study the constant vorticity problem comes from multiple sources. On one hand, from a mathematical perspective, it provides us with the possibility to consider vorticity effects in a framework where the equations of motion can still be described in terms of the water/air interface, while allowing for a larger range of dynamic behavior, which is particularly interesting over large time scales. On the other hand, from a practical perspective, constant vorticity flows are good models for the water motion in the presence of countercurrents. An interesting example of this type is provided by tidal effects.

The constant vorticity problem is a subset of the full vorticity problem, and as such, local well-posedness for regular enough data can be viewed as a consequence of results for the general problem. For this we refer the reader to [Christodoulou and Lindblad 2000; Lannes 2005; Lindblad 2005; Coutand and Shkoller 2007; Shatah and Zeng 2008; Zhang and Zhang 2008]. There has also been a considerable body of work devoted to the study of solitary waves in constant vorticity flows. A good source of information in this direction is provided by several recent articles [Constantin and Varvaruca 2011; Kozlov, Kuznetsov and Lokharu 2014; Constantin, Kalimeris and Scherzer 2015], as well as the survey article [Strauss 2010]. Various ways of formulating the equations have been described in [Wahlén 2007; Ehrnström 2008; Ashton and Fokas 2011].

The conformal formulation for two-dimensional water waves, which we adopt here, following our previous work [Hunter, Ifrim and Tataru 2016], originates in early work on traveling waves; see, e.g., [Levi-Civita 1925]. For the dynamical problem, to the best of our knowledge it first appears in [Ovsjannikov 1974], but was better developed later in [Wu 1997] and also in [Dyachenko, Kuznetsov, Spector and Zakharov 1996]. It has been widely used since then in order to study a variety of water wave problems. However, as far as we know, this is the first article where this formulation is fully implemented in the constant vorticity case.

1A. ***Equations of motion.*** The water flow is governed by the incompressible Euler equations in the fluid domain Ω_t, with a dynamic and a kinematic boundary condition on the free surface of the fluid Γ_t. Denoting the fluid velocity by $\boldsymbol{u}(t,x,y)=(u(t,x,y),v(t,x,y))$, and the pressure by $p(t,x,y)$, the equations of motion are the equation of mass conservation

$$u_x+v_y=0 \quad \text{in } \Omega(t), \tag{1-1}$$

and Euler's equations also in $\Omega(t)$

$$\begin{cases} u_t+uu_x+vu_y=-p_x, \\ v_t+uv_x+vv_y=-p_y-g, \end{cases} \tag{1-2}$$

where g is the gravitational constant. The boundary conditions for capillary-gravity waves are the dynamic boundary condition

$$p = p_0 \quad \text{on } \Gamma(t), \tag{1-3}$$

p_0 being the constant atmospheric pressure, and the kinematic boundary condition, which asserts that the free boundary $\Gamma(t)$ is transported along the flow. Since we are in the two-dimensional case, the vorticity will also be transported along the flow. This makes it possible to study flows with a nonzero constant vorticity field,

$$\omega = u_y - v_x = -c, \qquad \text{where } c \text{ is a constant.}$$

Then the velocity field $\boldsymbol{u}$ can be represented as

$$\boldsymbol{u} = (cy + \varphi_x, \varphi_y),$$

where $\varphi(t, x, y)$ is called the (generalized) velocity potential. Here φ it is defined up to an arbitrary function of time and, by the incompressibility condition, it satisfies the Laplace equation in $\Omega(t)$

$$\triangle\varphi(t, x, y) = 0.$$

This brings us to our boundary condition on the bottom, which asserts that

$$\lim_{y\to-\infty} \varphi(x, y) = 0, \qquad \text{uniformly in } x.$$

Then φ is uniquely determined by its values on the free surface $\Gamma(t)$.

Introducing its harmonic conjugate $\theta(t, x, y)$,

$$\varphi_x = \theta_y, \qquad \varphi_y = -\theta_x,$$

we can rewrite the equations in (1-2) as a single scalar equation in the fluid domain:

$$\nabla\big(\varphi_t - c\theta + cy\varphi_x + \tfrac{1}{2}(\varphi_x^2 + \varphi_y^2) + gy\big) = 0.$$

Since φ is only defined up to an arbitrary function of time, we can absorb a function of time into φ. Using the fact that the pressure is constant on $\Gamma(t)$, we obtain the following analogue of Bernoulli's equation:

$$\varphi_t - c\theta + cy\varphi_x + \tfrac{1}{2}(\varphi_x^2 + \varphi_y^2) + gy = 0 \quad \text{on } \Gamma(t). \tag{1-4}$$

This is the equation that makes possible reduction of dimensionality of the problem, in the same manner as for purely potential flows. This is in combination with the kinematic boundary condition, which is already restricted to $\Gamma(t)$. We remark that expressing θ and the full gradient $\nabla\varphi$ on $\Gamma(t)$ in terms of the restriction of φ to $\Gamma(t)$ requires using the Dirichlet-to-Neumann map associated to the fluid domain, and in turn makes our equations nonlocal.

1B. ***The equations in holomorphic coordinates.*** The first issue we need to address is the choice of coordinates. Here we take our cue from [Hunter, Ifrim and Tataru 2016], henceforth abbreviated [HIT16], and work in holomorphic coordinates. There are also other ways of expressing the equations, for instance in Cartesian coordinates using the Dirichlet-to-Neumann map associated to the water domain; see, e.g., [Alazard, Burq and Zuily 2011; 2014]. However, we prefer the holomorphic coordinates due to the simpler form of the equations; in particular, in these coordinates the Dirichlet-to-Neumann map is diagonalized and given in terms of the standard Hilbert transform.

We obtain a system which models the time evolution of the free surface, described via a function W as the graph of the function $\alpha \to W(\alpha)+\alpha$, and that of the holomorphic velocity potential $Q=\varphi+i\theta$ restricted to the free surface. The derivation of the equations is relegated to Appendix B, also using some of the analysis from the Appendix in [HIT16]. Some minor changes are needed for the periodic case; these are also described in [HIT16, Appendix B]. The outcome of this computation is the following set of equations:

$$\begin{cases} W_t+(W_\alpha+1)\underline{F}+i\dfrac{c}{2}W=0, \\ Q_t-igW+\underline{F}\,Q_\alpha+icQ+\boldsymbol{P}\left[\dfrac{|Q_\alpha|^2}{J}\right]-i\dfrac{c}{2}T_1=0, \end{cases} \tag{1-5}$$

where $J:=|1+W_\alpha|^2$, $\boldsymbol{P}$ is the projection onto negative wave numbers

$$\boldsymbol{P}=\tfrac{1}{2}(I-iH), \quad \text{with } H \text{ the Hilbert transform},$$

and

$$F:=\boldsymbol{P}\left[\frac{Q_\alpha-\bar{Q}_\alpha}{J}\right], \quad F_1:=\boldsymbol{P}\left[\frac{W}{1+\overline{W}_\alpha}+\frac{\overline{W}}{1+W_\alpha}\right],$$
$$\underline{F}:=F-i\frac{c}{2}F_1, \qquad T_1:=\boldsymbol{P}\left[\frac{W\bar{Q}_\alpha}{1+\overline{W}_\alpha}-\frac{\overline{W}Q_\alpha}{1+W_\alpha}\right].$$

These equations are considered either in $\mathbb{R}\times\mathbb{R}$ or in $\mathbb{R}\times\mathbb{S}^1$. They model an evolution in the space of functions which admit bounded holomorphic extensions to the lower half-space; equivalently, their Fourier transform is supported on the negative real line. By a slight abuse, we call such functions holomorphic.

This is a Hamiltonian system, where the Hamiltonian has the form

$$\mathcal{E}(W,Q)=\Re\int g|W|^2(1+W_\alpha)-iQ\bar{Q}_\alpha+cQ_\alpha(\Im W)^2-\frac{c^3}{2i}|W|^2W(1+W_\alpha)\,d\alpha. \tag{1-6}$$

A second conserved quantity is the horizontal momentum,

$$\mathcal{P}(W,Q)=\int\left\{\frac{1}{i}(\bar{Q}W_\alpha-Q\overline{W}_\alpha)-c|W|^2+\frac{c}{2}(W^2\overline{W}_\alpha+\overline{W}^2W_\alpha)\right\}d\alpha, \tag{1-7}$$

which is the Hamiltonian for the group of horizontal translations.

We remark that if $c=0$ then the equations (1-5) are the gravity water wave equations studied in [HIT16]. The sign of c is not important, as it can be switched via the flip $\alpha\to-\alpha$. For convenience we assume $c>0$. The space-time scaling

$$(W(t,\alpha),Q(t,\alpha))\to(\lambda^{-2}W(\lambda t,\lambda^2x),\lambda^{-3}Q(\lambda t,\lambda^2x))$$

is a symmetry for gravity waves (thus it leaves g unchanged). Here it changes $c \to \lambda c$. Finally, the purely spatial scaling

$$(W(t,\alpha), Q(t,\alpha)) \to (\lambda^{-2} W(t, \lambda^2 x), \lambda^{-3} Q(t, \lambda^2 x))$$

has the effect of leaving c unaffected, but it changes $g \to \lambda^{-1} g$. Thus one could use scaling considerations to set both $c = 1$ and $g = 1$. However, we choose not to do that, and instead use the coefficients c and g to keep better track of the expressions arising in our analysis. In this context, it is useful to observe that one can interpret c^2/g as an (inverse) semiclassical parameter, so that all energy expressions can be viewed as homogeneous.

We further remark that the terms involving c are lower-order, though they cannot be viewed as bounded. Thus, it is natural to expect that the local theory for this problem is not very different from the gravity waves; indeed, our results in this regard are similar to [HIT16]. However, we will see that the long-time behavior is quite different in the constant vorticity case.

To further motivate our expectations concerning this system, we note that the linearization of the system (1-5) around the zero solution is

$$\begin{cases} w_t + q_\alpha = 0, \\ q_t + icq - igw = 0, \end{cases} \tag{1-8}$$

restricted to holomorphic functions (in our terminology, these are functions with negative spectrum). It is not difficult to see that this is a dispersive equation. Expressed as a second-order equation this becomes

$$w_{tt} + icw_t + igw_\alpha = 0. \tag{1-9}$$

From here we can deduce the associated dispersion relation,

$$\tau^2 + c\tau + g\xi = 0, \quad \xi \le 0. \tag{1-10}$$

This is the intersection of a lateral parabola with the left half-space. It has two branches, intersecting the axis $\xi = 0$ at $\tau = 0$ and $\tau = -c$.

The conserved energy and momentum for (1-8) are, respectively,

$$\mathcal{E}_0(w, r) = \int |w|^2 - iq\bar{q}_\alpha \, d\alpha = \|w\|_{L^2}^2 + \|q\|_{\dot{H}^{1/2}}^2,$$

$$\mathcal{P}_0(w, r) = \int \frac{1}{i}(\bar{q} w_\alpha - q\bar{w}_\alpha) - c|w|^2 \, d\alpha.$$

The former motivates the functional setting where we will study the equations (1-5). The system (1-8) is a well-posed linear evolution in the space $\dot{\mathcal{H}}_0$ of holomorphic functions endowed with the $L^2 \times \dot{H}^{1/2}$-norm. To measure higher regularity we will use the spaces $\dot{\mathcal{H}}_n$ endowed with the norm

$$\|(w, r)\|_{\dot{\mathcal{H}}_n}^2 := \sum_{k=0}^{n} \|\partial_\alpha^k (w, r)\|_{L^2 \times \dot{H}^{1/2}}^2,$$

where $n \ge 1$.

1C. *The differentiated equations and diagonalization.* As the system (1-5) is fully nonlinear, a standard procedure is to convert it into a quasilinear system by differentiating it. In the case of gravity waves this yields a self-contained first-order quasilinear system for (W_α, Q_α). This is no longer true here, precisely due to the contributions from F_1 and T_1.

To write the differentiated system, as well as the linearized system later on, we introduce several notations. First, as in [HIT16], we set

$$\mathbf{W} = W_\alpha, \quad R = \frac{Q_\alpha}{1+\mathbf{W}}, \quad Y = \frac{\mathbf{W}}{1+\mathbf{W}}.$$

The expression R has an intrinsic meaning; namely it is the complex velocity on the water surface. Y, on the other hand, is introduced for computational reasons only, in order to avoid rational expressions in many places in the sequel.

We also need two key auxiliary real functions. The first is $\underline{b}$, which we call the *advection velocity*, and is given by

$$\underline{b} := b - i\frac{c}{2}b_1, \quad b := \mathbf{P}\left[\frac{Q_\alpha}{J}\right] + \bar{\mathbf{P}}\left[\frac{\bar{Q}_\alpha}{J}\right], \quad b_1 := \mathbf{P}\left[\frac{W}{1+\bar{\mathbf{W}}}\right] - \bar{\mathbf{P}}\left[\frac{\bar{W}}{1+\mathbf{W}}\right]. \tag{1-11}$$

The second is the *frequency-shift* a, given by

$$\underline{a} := a + \frac{c}{2}a_1, \quad a := i(\bar{\mathbf{P}}[\bar{R}R_\alpha] - \mathbf{P}[R\bar{R}_\alpha]), \quad a_1 = R + \bar{R} - N, \tag{1-12}$$

where

$$N := \mathbf{P}[W\bar{R}_\alpha - \bar{\mathbf{W}}R] + \bar{\mathbf{P}}[\bar{W}R_\alpha - \mathbf{W}\bar{R}]. \tag{1-13}$$

The functions $\underline{a}$ and $\underline{b}$ are the leading-order coefficients in the linearized equation, and thus fully describe the quasilinear nature of the problem. The linearized equation is more involved and is described in full later, but, as a baseline, the reader should keep in mind the linear system

$$\begin{cases} w_t + bw_\alpha + q_\alpha = 0, \\ q_t + icq - i(g+a)w = 0. \end{cases} \tag{1-14}$$

The real function $g + \underline{a}$ has a physical interpretation as the normal derivative of the pressure on the interface; this is proved in Appendix A. For more regular irrotational waves (i.e., $c = 0$) this was proved by Wu [1997] to be positive; an alternate proof was provided in [HIT16] in the context of holomorphic coordinates, assuming only scale-invariant regularity $(\mathbf{W}, R) \in \dot{\mathcal{H}}_{1/2}$. This positivity, called the *Taylor sign condition* [1950], was crucial for the well-posedness of the water wave system, both with gravity [Wu 1997; Hunter, Ifrim and Tataru 2016] and with surface tension [Ifrim and Tataru 2017].

For the present problem we still need to know that $g + \underline{a}$ is nonnegative, which corresponds with the normal derivative of the pressure being bounded away from zero. But it is no longer the case that this comes for free. Thus, we will impose the positivity condition on the normal derivative of the pressure and solve the equations for as long as this condition holds uniformly. We remark that this is always the case if we assume that R is small in L^∞, which is the case for our small-data result.

Differentiating with respect to α yields a system for (W_α, Q_α), which turns out to be degenerate hyperbolic with a double speed $\underline{b}$. This is explained in detail in the context of irrotational gravity waves in [HIT16], and easily carries over here as the highest-order terms in the equations are the same. Then it is natural to diagonalize it. This is also done exactly as in [HIT16], using the operator

$$\boldsymbol{A}(w, q) := (w, q - Rw), \qquad R := \frac{Q_\alpha}{1 + W_\alpha}. \tag{1-15}$$

Noting that

$$\boldsymbol{A}(W_\alpha, Q_\alpha) = (\boldsymbol{W}, R), \qquad \boldsymbol{W} := W_\alpha,$$

it follows that the pair $(\boldsymbol{W}, R)$ diagonalizes the differentiated system. Thus, rather than repeating the more extensive computations in [HIT16], here we directly take advantage of this knowledge to arrive more efficiently at the differentiated system for the diagonal variables $(\boldsymbol{W}, R)$.

We first introduce $\underline{b}$ into the equations using the relations

$$F = b - \frac{\bar{R}}{1 + \boldsymbol{W}}, \qquad F_1 = b_1 + \frac{\bar{W}}{1 + \boldsymbol{W}}, \tag{1-16}$$

so that the system (1-5) is written in the form

$$\begin{cases} W_t + \underline{b}(1 + W_\alpha) + i\frac{c}{2}W = \bar{R} + i\frac{c}{2}\bar{W}, \\ Q_t + \underline{b}Q_\alpha - igW + icQ - i\frac{c}{2}\bar{R}W = \bar{\boldsymbol{P}}[|R|^2] - i\frac{c}{2}\bar{\boldsymbol{P}}[W\bar{R} - \bar{W}R]. \end{cases}$$

Here the terms on the right are antiholomorphic and disappear when the equations are projected onto the holomorphic space.

Next we differentiate the equations. For the first equation we need the expression for $\underline{b}_\alpha$, for which we introduce one last set of quadratic auxiliary functions M, M_1 and $\underline{M}$ as follows:

$$\begin{aligned} \underline{M} &:= M - i\frac{c}{2}M_1, \\ M &:= \frac{R_\alpha}{1 + \bar{\boldsymbol{W}}} + \frac{\bar{R}_\alpha}{1 + \boldsymbol{W}} - b_\alpha = \bar{\boldsymbol{P}}[\bar{R}Y_\alpha - R_\alpha\bar{Y}] + \boldsymbol{P}[R\bar{Y}_\alpha - \bar{R}_\alpha Y], \\ M_1 &:= \boldsymbol{W} - \bar{\boldsymbol{W}} - b_{1,\alpha} = \boldsymbol{P}[W\bar{Y}]_\alpha - \bar{\boldsymbol{P}}[\bar{W}Y]_\alpha. \end{aligned} \tag{1-17}$$

Thus, we can substitute $\underline{b}_\alpha$ with

$$\underline{b}_\alpha = \frac{R_\alpha}{1 + \bar{\boldsymbol{W}}} + \frac{\bar{R}_\alpha}{1 + \boldsymbol{W}} - i\frac{c}{2}(\boldsymbol{W} - \bar{\boldsymbol{W}}) - \underline{M}. \tag{1-18}$$

For the second equation we switch directly to R, and then $\underline{b}_\alpha$ is no longer needed in view of the identity

$$(\partial_t + \underline{b}\partial_\alpha)R = \partial_\alpha(\partial_t + \underline{b}\partial_\alpha)Q - R\partial_\alpha(W_t + \underline{b}(1 + W_\alpha)).$$

Taking the above discussion into account, after some straightforward computations, one arrives at the system

$$\begin{cases} W_t + \underline{b} W_\alpha + \dfrac{(1+W)R_\alpha}{1+\overline{W}} = (1+W)\underline{M} + i\dfrac{c}{2} W(W-\overline{W}), \\ R_t + \underline{b} R_\alpha + icR - i\dfrac{gW-a}{1+W} = i\dfrac{c}{2}\dfrac{RW+\bar{R}W+N}{1+W}. \end{cases} \tag{1-19}$$

This governs an evolution in the space of holomorphic functions, and will be used both directly and in its projected version. Obviously the two forms are algebraically equivalent.

We remark that while the transport coefficient $\underline{b}$ appears explicitly in these equations, the frequency shift $\underline{a}$ does not. This is due to the fact that the right-hand side of the second equation is still fully nonlinear in W. To understand better the role played by $\underline{a}$ one needs either to consider the linearized equations, which are discussed in Section 3, or to further differentiate (1-19), as in Section 4.

1D. *The main results.* To describe the lifespan of the solutions we begin with the control norms in [HIT16], namely

$$A := \|W\|_{L^\infty} + \|Y\|_{L^\infty} + \||D|^{1/2}R\|_{L^\infty \cap B^{0,\infty}_2}, \tag{1-20}$$

$$B := \||D|^{1/2}W\|_{\mathrm{BMO}} + \|R_\alpha\|_{\mathrm{BMO}}, \tag{1-21}$$

where $|D|$ represents the multiplier with symbol $|\xi|$. In order to estimate lower-order terms introduced in conjunction with c we also introduce

$$A_{-1/2} := \||D|^{1/2}W\|_{L^\infty} + \|R\|_{L^\infty}, \tag{1-22}$$

$$A_{-1} := \|W\|_{L^\infty}. \tag{1-23}$$

It is also useful to introduce the notations

$$\begin{aligned} \underline{B} &:= B + cA + c^2 A_{-1/2}, \\ \underline{A} &:= A + cA_{-1/2} + c^2 A_{-1}. \end{aligned} \tag{1-24}$$

Here A is a scale-invariant quantity, while B corresponds to the homogeneous $\dot{\mathcal{H}}_1$-norm of (W, R) and $A_{-1/2}$ corresponds to the homogeneous $\dot{\mathcal{H}}_0$-norm of (W, R). We note that B, and all but the Y-component of A are controlled by the energy and $\dot{\mathcal{H}}_1$-norm of (W, R).

Now we are ready to state our main results. We begin with the local well-posedness result:

Theorem 1. *Let $n \geq 1$. The system (1-5) is locally well-posed for initial data (W_0, Q_0) with the regularity*

$$(W_0, Q_0) \in \dot{\mathcal{H}}_0, \quad (W_0, R_0) \in \dot{\mathcal{H}}_1,$$

and satisfying the pointwise constraints

$$|W(\alpha)+1| > \delta > 0 \quad (\textit{no interface singularities}), \tag{1-25}$$

$$g + \underline{a}(\alpha) > \delta > 0 \quad (\textit{Taylor sign condition}). \tag{1-26}$$

Further, the solution can be continued for as long as $\underline{A}$ and $\underline{B}$ remain bounded and the pointwise conditions above hold uniformly. The same result holds in the periodic setting.

The well-posedness above should be interpreted in the sense of Hadamard. To be more precise, it means that there exists some time $T > 0$, depending only on the initial data size and on the constant δ in the above pointwise constraints, so that the following properties hold:

(1) (regular data) For each data (W_0, Q_0), which is as above, but with additional regularity $(\mathbf{W}_0, R_0) \in \dot{\mathcal{H}}_n$, with $n \geq 2$, there exists a unique solution (W, Q) in $[0, T]$, with the property that

$$\|(\mathbf{W}, R)\|_{C[0,T;\dot{\mathcal{H}}_k]} \lesssim \|(\mathbf{W}_0, R_0)\|_{\dot{\mathcal{H}}_k}, \quad 0 \leq k \leq n.$$

(2) (rough data) For each data (W_0, Q_0) as above there exists a solution (W, Q) in $[0, T]$, with the property that

$$\|(\mathbf{W}, R)\|_{C[0,T;\dot{\mathcal{H}}_k]} \lesssim \|(\mathbf{W}_0, R_0)\|_{\dot{\mathcal{H}}_k}, \quad k = 0, 1.$$

Further this solution is the unique limit of regular solutions, and depends continuously on the initial data.

(3) (weak Lipschitz dependence) The solution (W, Q) has a Lipschitz dependence on the initial data in the $\dot{\mathcal{H}}_1$ topology.

The implicit constants in all the estimates above depend only on the initial data size and on the constant δ in the above pointwise constraints. Further, the last part of the Theorem 1 asserts that in effect these constants can alternatively be estimated purely in terms of our uniform control parameters A and B, rather than the full Sobolev norm of the data.

Our second result in this paper is a cubic lifespan bound for the small-data problem:

Theorem 2. *Let (W, Q) be a solution for the system (1-5) whose initial data satisfies*

$$\|(W_0, Q_0)\|_{\dot{\mathcal{H}}_0} + \|(\mathbf{W}_0, R_0)\|_{\dot{\mathcal{H}}_1} \leq \epsilon \ll 1. \tag{1-27}$$

Then the solution exists for a time $T_\epsilon \approx \epsilon^{-2}$, with bounds

$$\|(W, Q)(t)\|_{\dot{\mathcal{H}}_0} + \|(\mathbf{W}, R)(t)\|_{\dot{\mathcal{H}}_1} \lesssim \epsilon, \quad |t| < T_\epsilon. \tag{1-28}$$

Further, higher regularity is also preserved,

$$\|(\mathbf{W}, R)(t)\|_{\dot{\mathcal{H}}_n} \lesssim \|(\mathbf{W}, R)(0)\|_{\dot{\mathcal{H}}_n}, \quad |t| < T_\epsilon, \tag{1-29}$$

whenever the norm on the right is finite.

To the best of our knowledge this is the first nontrivial lifespan bound for solutions to this problem. A similar result for the zero-vorticity problem was proved in our earlier article [HIT16]. The problem here is considerably more difficult than the one in [HIT16], both technically and conceptually. At the technical level, the normal form for the vorticity problem is much more involved, and quite nontrivial to compute (see the next section). Qualitatively, here we have stronger quadratic interactions at low frequency, which, unlike in [HIT16], prevent us from obtaining cubic bounds for the linearized equation (and thus, for differences of solutions).

We further remark that in [HIT16] we also provide a proof of an almost global result for small localized data for gravity waves. Our aim is to also provide a similar result in this context. However, the ideas in [HIT16] do not directly carry over to this case, due first, to the lack of a scaling symmetry, and secondly, to the lack of cubic estimates for the linearized equation. We hope to be able to address these issues in subsequent work.

We note that the periodic case is almost identical and is not discussed separately. The only difference in the analysis is in how the constant functions are treated. This is discussed in detail in the Appendix to [HIT16], and carries over to the present paper without any change.

1E. ***Outline of the paper.*** There are three key steps in our analysis, which eventually provide all the ingredients which are necessary in order to prove our main results. These are as follows:

(i) *The normal-form analysis.* The constant vorticity water wave equation has many quadratic interactions, yet we seek to prove small-data lifespan bounds as if the nonlinearities were cubic. At least formally the key to this is the normal-form analysis [Shatah 1985], which allows us to replace quadratic nonlinearities with cubic ones. While the normal-form transformation for gravity waves is quite straightforward, in the presence of constant vorticity, this is no longer the case.

Indeed, the normal form turns out to be unbounded both at low frequency and at high frequency. This computation is fairly involved, and is carried out in the next section. Its redeeming feature is that its outcome is also quite explicit.

The normal form we calculate here is not directly used in any of the estimates we derive later on. However, it is crucially used in order to construct modified energies with cubic estimates, which is the base of our *quasilinear modified energy method* [Hunter, Ifrim, Tataru and Wong 2015; 2016].[1] Even though the normal form is badly unbounded, it has enough of a "null structure", or antisymmetry, so that the cubic energy corrections it generates are all of bounded type.

(ii) *The analysis of the linearized equation in Section 3*. This is a critical part of any local well-posedness result for a quasilinear problem. The derivation of the equations is also interesting, as it clarifies the quasilinear structure and the roles played by the advection coefficient $\underline{b}$, and the frequency shift $\underline{a}$.

As for the gravity waves in [HIT16], we are able to prove that the linearized problem is well-posed in our base space $\dot{\mathcal{H}}_0$. Further, the bounds we prove are in terms of our control parameters $\underline{A}$ and $\underline{B}$, and not in terms of the full Sobolev norm of the solution.

Unlike in [HIT16], we are no longer able to prove cubic estimates for the linearized equation. This is due to the unbounded low-frequency part of the normal-form transformation, which loses its skew-adjoint structure after linearization. Because of this, we are able to use the bounds for the linearized equation in the proof of local well-posedness, but only partially in the proof of the cubic lifespan result.

(iii) *The cubic energy estimates in Section 4*. Since we already have the conserved Hamiltonian, which controls the $\dot{\mathcal{H}}_0$-norm of (W, Q), our task here is to successively provide bounds for $(\boldsymbol{W}, R)$ in the

[1] For earlier attempts to adapt normal forms to quasilinear problems we refer the reader to [Wu 2009; Germain and Masmoudi 2014; Ionescu and Pusateri 2015]. Other equally successful approaches are the *paradiagonalization method* in [Alazard and Delort 2015], and the flow method of [Hunter and Ifrim 2012].

$\dot{\mathcal{H}}_k$ spaces for $k = 0, 1, \dots$. In all cases, we control the evolution of these norms using our pointwise control parameters $\underline{A}$ and $\underline{B}$.

These bounds come in two flavors: (a) local bounds, which apply for large data, and are needed for the local well-posedness result, and (b) cubic long-time bounds for small data, which are used for the cubic lifespan result. The former are obtained largely by differentiating the equation, and then by applying the bounds for the linearized equation. The latter, however, requires computing cubic energy corrections, and there, we rely heavily on the normal form.

The crucial step is the one for the $\dot{\mathcal{H}}_1$-norm of $(\boldsymbol{W}, R)$ (i.e., $k = 1$), as this is the level where we have our well-posedness result; for this reason, we describe this case in detail. The case $k = 0$ is simpler since $(\boldsymbol{W}, R)$ solves the linearized equation, so we already have the local bound. The case $k \geq 2$ is discussed last, without explicitly computing the modified energy.

Once the bounds for the linearized equation and for the differentiated equation are established, the remaining arguments in the proofs of our main theorems are a fairly straightforward repetition of arguments in [HIT16]. We outline this in the last section of the paper.

Finally, the Appendix plays two roles. In Appendix B we outline the derivation of the constant-vorticity gravity wave equation in holomorphic coordinates. Last, but not least, in Appendix A we collect a number of bilinear Coifman–Meyer and nonlinear Moser-type estimates, some from [HIT16], and prove the ones which are new in this paper.

2. The normal-form transformation

The nonlinear evolution (1-5) contains quadratic terms, yet for our problem we seek to prove cubic lifespan bounds, as if the nonlinearity is at least cubic. In this section we consider the question of finding a normal transformation, whose aim is to replace the original variables (W, Q) with normal-form variables $(\widetilde{W}, \widetilde{Q})$ of the form

$$\begin{cases} \widetilde{W} &= W + \boldsymbol{P} W_{[2]}, \\ \widetilde{Q} &= Q + \boldsymbol{P} Q_{[2]}, \end{cases} \tag{2-1}$$

where $W_{[2]}$ and $Q_{[2]}$ are quadratic forms in (W, Q), so that the normal-form variables satisfy an equation with only cubic-and-higher terms,

$$\begin{cases} \widetilde{W}_t + \widetilde{Q}_\alpha = \text{cubic and higher}, \\ \widetilde{Q}_t - ig\widetilde{W} + ic\widetilde{Q} = \text{cubic and higher}. \end{cases} \tag{2-2}$$

We will indeed show that such a normal-form transformation exists; as it turns out, it is highly unbounded, both at low and at high frequencies. However, this is expected, and it does not cause any difficulties. This is because we do not use the normal form directly, but only as a tool to help us construct modified energies in our quasilinear modified energy method. As it turns, even though the normal-form transformation is unbounded, it has some favorable structure, so that the modified cubic energies are nevertheless bounded.

2A. ***The resonance analysis.*** We begin by examining whether our evolution has quadratic resonant interactions. Taking into account the possible complex conjugations and the dispersion relation (1-10), the quadratic resonant interactions are associated with three pairs of characteristic frequencies (τ_1, ξ_1), (τ_2, ξ_2) and (τ, ξ) so that $(\tau_1, \xi_1) + (\tau_2, \xi_2) = (\tau, \xi)$. Combining this with the dispersion relation we obtain $\tau_1^2 + \tau_2^2 = \tau^2$, which leads to $\tau_1 \tau_2 = 0$. Hence, we either have $(\tau_1, \xi_1) = (0, 0)$ or $(\tau_2, \xi_2) = (0, 0)$. Thus, resonant interactions occur only when either one of the inputs or the output is at frequency zero. In terms of the normal-form transformation, this indicates that at most we will have singularities when either input is at frequency zero, or the output is at frequency zero. We will see that, due to the form of the quadratic terms in the equation, the former scenario happens, but the latter does not.

2B. ***The normal-form computation.*** We begin with the quadratic and expansion in the equation (1-5), and then we compute the normal-form transformation which eliminates the quadratic terms from the equation. Starting with

$$F \approx Q_\alpha - Q_\alpha W_\alpha + \boldsymbol{P}[\bar{Q}_\alpha W_\alpha - Q_\alpha \overline{W}_\alpha] + \boldsymbol{P}[(Q_\alpha - \bar{Q}_\alpha)(4|\Re W_\alpha|^2 - |W_\alpha|^2)],$$

we compute the multilinear expansion

$$\begin{cases} W_t + Q_\alpha = G^{(2)} + G^{(3+)}, \\ Q_t - igW + icQ = K^{(2)} + K^{(3+)}, \end{cases} \tag{2-3}$$

where the quadratic terms $(G^{(2)}, K^{(2)})$ are given by

$$\begin{cases} G^{(2)} = -\boldsymbol{P}[\bar{Q}_\alpha W_\alpha - Q_\alpha \overline{W}_\alpha] - i\dfrac{c}{2}\boldsymbol{P}[W\overline{W}_\alpha + \overline{W} W_\alpha] + i\dfrac{c}{2} W W_\alpha, \\ K^{(2)} = -Q_\alpha^2 - \boldsymbol{P}[|Q_\alpha|^2] + i\dfrac{c}{2} W Q_\alpha + i\dfrac{c}{2}\boldsymbol{P}[W\bar{Q}_\alpha - \overline{W} Q_\alpha]. \end{cases}$$

The role of the normal-form transformation is to eliminate the quadratic terms $(G^{(2)}, K^{(2)})$ from the equation (2-3). We can divide the quadratic terms above into two classes:

(a) holomorphic, i.e., those which are the product of two holomorphic functions,

(b) mixed, i.e., those which are the (projected) product of one holomorphic function with another's conjugate.

Thus, we expect the normal form to have a similar structure,

$$W_{[2]} = W_{[2]}^h + W_{[2]}^a, \qquad Q_{[2]} = Q_{[2]}^h + Q_{[2]}^a, \tag{2-4}$$

where, allowing for all possible combinations, the above components must have the form

$$\begin{aligned} W_{[2]}^h &= B^h(W, W) + C^h(Q, Q) + D^h(W, Q), \\ W_{[2]}^a &= B^a(W, \overline{W}) + C^a(Q, \bar{Q}) + D^a(W, \bar{Q}) + E^a(Q, \overline{W}), \\ Q_{[2]}^h &= F^h(W, W) + H^h(Q, Q) + A^h(W, Q), \\ Q_{[2]}^a &= F^a(W, \overline{W}) + H^a(Q, \bar{Q}) + A^a(W, \bar{Q}) + G^a(Q, \overline{W}). \end{aligned} \tag{2-5}$$

All the above expressions are translation-invariant bilinear forms, whose symbols we need to compute. Our main result is as follows:

Proposition 2.1. *The equation* (2-3) *admits a normal-form transformation* (2-1) *as in* (2-4)–(2-5), *where the symbols for the bilinear operators in* (2-5) *are given by* (2-12), (2-13), *and* (2-16).

For later use, i.e., for computing the modified energy cubic corrections, we also translate the symbols (2-12), (2-13), and (2-16) into the spatial description. Precisely, the holomorphic terms have the form

$$\begin{cases} B^h(W, W) = -W\mathbf{W} + i\dfrac{c^2}{2g}(\mathbf{W}\partial_\alpha^{-1}W + W^2) + \dfrac{c^4}{4g^2}W\partial_\alpha^{-1}W, \\ C^h(Q, Q) = -\dfrac{c^2}{4g^2}QQ_\alpha, \\ D^h(W, Q) = -\dfrac{c}{2g}(\mathbf{W}Q + WQ_\alpha) + i\dfrac{c^3}{4g}(WQ + \partial_\alpha^{-1}WQ_\alpha), \\ F^h(W, W) = i\dfrac{c}{4}W^2 + \dfrac{c^3}{4g}W\partial_\alpha^{-1}W, \\ H^h(Q, Q) = -\dfrac{c}{2g}QQ_\alpha, \\ A^h(W, Q) = -WQ_\alpha + i\dfrac{c^2}{2g}\partial_\alpha^{-1}WQ_\alpha + i\dfrac{c^2}{4g}WQ, \end{cases} \tag{2-6}$$

and the antiholomorphic counterparts are given by

$$\begin{cases} B^a(W, \overline{W}) = -\mathbf{W}\overline{W} - i\dfrac{c^2}{2g}\mathbf{W}\partial_\alpha^{-1}\overline{W} + i\dfrac{c^2}{4g}|W|^2 - \dfrac{c^4}{4g^2}W\partial_\alpha^{-1}\overline{W}, \\ C^a(Q, \bar{Q}) = -\dfrac{c^2}{4g^2}\bar{Q}Q_\alpha, \\ D^a(W, \bar{Q}) = -\dfrac{c}{2g}\mathbf{W}\bar{Q} + i\dfrac{c^3}{4g^2}W\bar{Q}, \\ G^a(Q, \overline{W}) = -Q_\alpha\overline{W} - i\dfrac{c^2}{2g}Q_\alpha\partial_\alpha^{-1}\overline{W}, \\ E^a(Q, \overline{W}) = -\dfrac{c}{2g}Q_\alpha\overline{W} - i\dfrac{c^3}{4g^2}Q_\alpha\partial_\alpha^{-1}\overline{W}, \\ H^a(Q, \bar{Q}) = -\dfrac{c}{2g}\bar{Q}Q_\alpha, \\ A^a(W, \bar{Q}) = i\dfrac{c^2}{4g}W\bar{Q}, \\ F^a(W, \overline{W}) = i\dfrac{c}{2}|W|^2 - \dfrac{c^3}{4g}W\partial_\alpha^{-1}\overline{W}. \end{cases} \tag{2-7}$$

We note that while for computational purposes it is convenient to separate the two components of the normal form, in order to see the antisymmetric structure of the (low frequency) unbounded part, one has to consider them together; see the computations in Section 4. Toward that goal, we rewrite the quadratic normal-form components in the following form:

$$\begin{aligned} W^{[2]} &= -(W+\overline{W})W_\alpha - \frac{c}{2g}[(Q+\bar{Q})W_\alpha + (W+\overline{W})Q_\alpha] \\ &\quad + i\frac{c^2}{2g}\big[(\partial^{-1}W - \partial^{-1}\overline{W})W_\alpha + W^2 + \tfrac{1}{2}|W|^2\big] - \frac{c^2}{4g^2}(Q+\bar{Q})Q_\alpha \\ &\quad + i\frac{c^3}{4g^2}[(Q+\bar{Q})W + (\partial^{-1}W - \partial^{-1}\overline{W})Q_\alpha] + \frac{c^4}{4g^2}(\partial^{-1}W - \partial^{-1}\overline{W})W, \end{aligned} \tag{2-8}$$

$$\begin{aligned} Q^{[2]} &= -(W+\overline{W})Q_\alpha - \frac{c}{2g}(Q+\bar{Q})Q_\alpha + i\frac{c}{4}(W^2 + 2|W|^2) \\ &\quad + i\frac{c^2}{2g}\big[(\partial^{-1}W - \partial^{-1}\overline{W})Q_\alpha + \tfrac{1}{2}(Q+\bar{Q})W\big] + \frac{c^3}{4g}(\partial^{-1}W - \partial^{-1}\overline{W})W. \end{aligned}$$

Proof of Proposition 2.1. A priori, computing the normal form, i.e., all symbols of the bilinear expressions above, might appear quite involved. However, there are several observations which bring this analysis to a more manageable level:

- The analysis for the holomorphic products, and for the mixed terms is completely separate.
- The system we obtain for the symbols has polynomial coefficients, so the solutions are rational functions.
- Counting c as one half of a derivative; the problem is homogeneous. Thus, organizing symbols based on the powers of c, each such term will have a specific homogeneity. Further, at each power of c, we will encounter only half the terms, as there is a half-derivative difference between the scaling of W and that of Q.
- The terms without c are already known from the gravity wave problem [HIT16].

Given the above considerations, the natural strategy is to split the analysis into the holomorphic and the mixed part, and in each of these cases to successively solve for increasing powers of c. The computation stops at c^4.

(i) *Holomorphic terms*. Here we seek a normal form for the system

$$\begin{cases} W_t + Q_\alpha = i\frac{c}{2}WW_\alpha, \\ Q_t - igW + icQ = -Q_\alpha^2 + i\frac{c}{2}WQ_\alpha + \text{cubic}. \end{cases}$$

By checking parity, our normal form must be

$$\begin{cases} \widetilde{W} = W + B^h(W,W) + C^h(Q,Q) + D^h(W,Q), \\ \widetilde{Q} = Q + F^h(W,W) + H^h(Q,Q) + A^h(W,Q), \end{cases}$$

where B^h, C^h, F^h and H^h are symmetric bilinear forms with symbols $B^h(\xi,\eta)$, $C^h(\xi,\eta)$, $F^h(\xi,\eta)$, $H^h(\xi,\eta)$ and A^h and D^H are arbitrary. We compute

$$\begin{aligned} \widetilde{W}_t + \widetilde{Q}_\alpha &= -2B^h(Q_\alpha, W) + 2igC^h(W,Q) - 2icC^h(Q,Q) - D^h(Q_\alpha,Q) + igD^h(W,W), \\ &\quad - icD^h(W,Q) + \partial_\alpha(F^h(W,W) + H^h(Q,Q) + A^h(W,Q)) + i\frac{c}{2}WW_\alpha + \text{cubic}, \end{aligned}$$

$$\begin{aligned}\widetilde{Q}_t - ig\widetilde{W} + ic\widetilde{Q} = -2F^h(Q_\alpha, W) + 2igH^h(W, Q) - icH^h(Q, Q) - A^h(Q_\alpha, Q) + igA^h(W, W)\\ - igB^h(W, W) - igC^h(Q, Q) - igD^h(W, Q) + icF^h(W, W) - Q_\alpha^2\\ + i\frac{c}{2}WQ_\alpha + \text{cubic}.\end{aligned}$$

We denote the two input frequencies by ξ and η, both of which are negative. Then, matching like terms, we obtain the following linear system for the symbols:

$$\begin{cases} 2\eta B^h - 2gC^h + cD^h - (\xi+\eta)A^h = 0,\\ 2cC^h + [\xi D^h]_{\text{sym}} - (\xi+\eta)H^h = 0,\\ g[D^h]_{\text{sym}} + (\xi+\eta)F^h = -i\frac{c}{4}(\xi+\eta),\\ 2\eta F^h - 2gH^h + gD^h = i\frac{c}{2}\eta,\\ cH^h + [\xi A^h]_{\text{sym}} + gC^h = -i\xi\eta,\\ g[A^h]_{\text{sym}} - gB^h + cF^h = 0,\end{cases}$$

where "sym" stands for symmetrization. Using the first equation in the system above helps us to determine the symmetrized symbol of A^h,

$$A^h = \frac{1}{\xi+\eta}[cD^h - 2gC^h + 2\eta B^h],$$

which implies

$$\begin{aligned}[A^h]_{\text{sym}} &= \frac{1}{\xi+\eta}\big[c[D^h]_{\text{sym}} - 2gC^h + (\eta+\xi)B^h\big],\\ [\xi A^h]_{\text{sym}} &= \frac{1}{\xi+\eta}\big[c[\xi D^h]_{\text{sym}} - g(\xi+\eta)C^h + 2\xi\eta B^h\big].\end{aligned} \tag{2-9}$$

The fourth equation provides an expression for the symmetrized symbol of D^h,

$$D^h = -\frac{2}{g}\eta F^h + 2H^h + i\frac{c}{2g}\eta,$$

which implies

$$\begin{aligned} g[D^h]_{\text{sym}} &= -(\xi+\eta)F^h + 2gH^h + i\frac{c}{4}(\xi+\eta),\\ [\xi D^h]_{\text{sym}} &= -\frac{2}{g}\xi\eta F^h + (\xi+\eta)H^h + i\frac{c}{2g}\xi\eta,\end{aligned} \tag{2-10}$$

Using (2-10) in (2-9) gives

$$\begin{aligned}[A^h]_{\text{sym}} &= \frac{1}{\xi+\eta}\bigg[-\frac{c}{g}(\xi+\eta)F^h + 2cH^h + i\frac{c^2}{4g}(\xi+\eta) - 2gC^h + (\eta+\xi)B^h\bigg],\\ [\xi A^h]_{\text{sym}} &= \frac{1}{\xi+\eta}\bigg[-\frac{2c}{g}\xi\eta F^h + (\xi+\eta)cH^h + i\frac{c^2}{2g}\xi\eta - g(\xi+\eta)C^h + 2\xi\eta B^h\bigg].\end{aligned} \tag{2-11}$$

Thus, we return to the following system

$$\begin{cases} 2cC^h + [\xi D^h]_{\text{sym}} - (\xi+\eta)H^h = 0,\\ g[D^h]_{\text{sym}} + (\xi+\eta)F^h = -i\frac{c}{4}(\xi+\eta),\\ cH^h + [\xi A^h]_{\text{sym}} + gC^h = -i\xi\eta,\\ g[A^h]_{\text{sym}} - gB^h + cF^h = 0,\end{cases}$$

where we substitute the corresponding values from (2-10), (2-11), and obtain

$$\begin{cases} 2cC^h-\dfrac{2}{g}\xi\eta F^h+(\xi+\eta)H^h+i\dfrac{c}{2g}\xi\eta-(\xi+\eta)H^h=0,\\ -(\xi+\eta)F^h+2gH^h+i\dfrac{c}{4}(\xi+\eta)+(\xi+\eta)F^h=-i\dfrac{c}{4}(\xi+\eta),\\ cH^h+\dfrac{1}{\xi+\eta}\Big[-\dfrac{2c}{g}\xi\eta F^h+(\xi+\eta)cH^h+i\dfrac{c^2}{2g}\xi\eta-g(\xi+\eta)C^h+2\xi\eta B^h\Big]+gC^h=-i\xi\eta,\\ \dfrac{1}{\xi+\eta}\Big[-c(\xi+\eta)F^h+2cgH^h+i\dfrac{c^2}{4}(\xi+\eta)-2g^2C^h+(\eta+\xi)gB^h\Big]-gB^h+cF^h=0, \end{cases}$$

so

$$\begin{cases} 2cC^h-\dfrac{2}{g}\xi\eta F^h=-i\dfrac{c}{2g}\xi\eta,\\ H^h=-i\dfrac{c}{4g}(\xi+\eta),\\ cH^h-\dfrac{c}{g}\dfrac{\xi\eta}{\xi+\eta}F^h+\dfrac{\xi\eta}{\xi+\eta}B^h=-i\dfrac{\xi\eta}{2}-i\dfrac{c^2}{4g}\dfrac{\xi\eta}{\xi+\eta},\\ cH^h-gC^h=-i\dfrac{c^2}{8g}(\xi+\eta). \end{cases}$$

The solution is

$$\begin{aligned} B^h&=-\frac{i}{2}(\xi+\eta)+i\frac{c^2}{4g}\frac{(\xi+\eta)^2}{\xi\eta}-i\frac{c^4}{8g^2}\frac{\xi+\eta}{\xi\eta},\\ C^h&=-i\frac{c^2}{8g^2}(\xi+\eta),\\ F^h&=i\frac{c}{4}-i\frac{c^3}{8g}\frac{\xi+\eta}{\xi\eta},\\ H^h&=-i\frac{c}{4g}(\xi+\eta). \end{aligned}\tag{2-12}$$

It remains to find A^h and D^h, which we obtain from

$$\begin{cases} 2\eta B^h-2gC^h+cD^h-(\xi+\eta)A^h=0,\\ 2\eta F^h-2gH^h+gD^h=i\dfrac{c}{2}\eta. \end{cases}$$

Thus,

$$\begin{aligned} A^h&=-i\eta+i\frac{c^2}{2g}\eta\xi^{-1}+i\frac{c^2}{4g},\\ D^h&=i\frac{c^3}{4g^2}\frac{\xi+\eta}{\xi}-i\frac{c}{2g}(\xi+\eta). \end{aligned}\tag{2-13}$$

(ii) *Mixed terms*. Here we need a normal form for the system

$$\begin{cases} W_t+Q_\alpha=-\boldsymbol{P}[\bar{Q}_\alpha W_\alpha-Q_\alpha\overline{W}_\alpha]-i\dfrac{c}{2}\boldsymbol{P}[W\overline{W}_\alpha+\overline{W}W_\alpha]+\text{cubic},\\ Q_t-igW+icQ=-\boldsymbol{P}[|Q_\alpha|^2]+i\dfrac{c}{2}\boldsymbol{P}[W\bar{Q}_\alpha-\overline{W}Q_\alpha]+\text{cubic}. \end{cases}$$

The general expression for our normal form is

$$\begin{cases} \widetilde{W}=W+B^a(W,\overline{W})+C^a(Q,\bar{Q})+D^a(W,\bar{Q})+E^a(Q,\overline{W}),\\ \widetilde{Q}=Q+F^a(W,\overline{W})+H^a(Q,\bar{Q})+A^a(W,\bar{Q})+G^a(Q,\overline{W}), \end{cases}\tag{2-14}$$

where no symmetry assumption is required. We compute

$$\begin{aligned}\widetilde{W}_t+\widetilde{Q}_\alpha = &-B^a(Q_\alpha,\overline{W})-B^a(W,\bar{Q}_\alpha)+igC^a(W,\bar{Q})-igC^a(Q,\overline{W})-D^a(Q_\alpha,\bar{Q})\\ &-igD^a(W,\overline{W})+icD^a(W,\bar{Q})+igE^a(W,\overline{W})-icE^a(Q,\overline{W})-E^a(Q,\bar{Q}_\alpha)\\ &-\boldsymbol{P}[\bar{Q}_\alpha W_\alpha-Q_\alpha\overline{W}_\alpha]+\partial_\alpha[F^a(W,\overline{W})+H^a(Q,\bar{Q})+A^a(W,\bar{Q})+G^a(Q,\overline{W})]\\ &-i\frac{c}{2}\boldsymbol{P}[W\overline{W}_\alpha+\overline{W}W_\alpha]+\text{cubic},\end{aligned}$$

$$\begin{aligned}\widetilde{Q}_t-ig\widetilde{W}+ic\widetilde{Q} = &-F^a(Q_\alpha,\overline{W})-F^a(W,\bar{Q}_\alpha)+igH^a(W,\bar{Q})-igH^a(Q,\overline{W})-A^a(Q_\alpha,\bar{Q})\\ &-igA^a(W,\overline{W})+icA^a(W,\bar{Q})-G^a(Q,\bar{Q}_\alpha)+igG^a(W,\overline{W})-igB^a(W,\overline{W})\\ &-igC^a(Q,\bar{Q})-igD^a(W,\bar{Q})-igE^a(Q,\overline{W})+icF^a(W,\overline{W})+icH^a(Q,\bar{Q})\\ &+icA^a(W,\bar{Q})-\boldsymbol{P}[|Q_\alpha|^2]+i\frac{c}{2}\boldsymbol{P}[W\bar{Q}_\alpha-\overline{W}Q_\alpha]+\text{cubic}.\end{aligned}$$

Now we denote by ξ the frequency of the holomorphic input and by η the frequency of the conjugated input (both negative). Matching again like terms, it remains to solve the system

$$\begin{cases}\xi B^a+gC^a+cE^a-(\xi-\eta)G^a=-i\xi\eta,\\ \eta B^a+gC^a+(\xi-\eta)A^a+cD^a=-i\xi\eta,\\ \xi D^a-\eta E^a-(\xi-\eta)H^a=0,\\ gD^a-gE^a-(\xi-\eta)F^a=i\frac{c}{2}(\eta-\xi),\\ \xi F^a+gH^a+gE^a=-i\frac{c}{2}\xi,\\ \eta F^a+gH^a+2cA^a-gD^a=i\frac{c}{2}\eta,\\ \xi A^a+gC^a-cH^a-\eta G^a=i\xi\eta,\\ gA^a+gB^a-cF^a-gG^a=0.\end{cases} \tag{2-15}$$

To avoid solving an 8×8 system we expand the result in terms of powers of c. A homogeneity analysis shows that B^a, C^a, G^a and A^a contain only even powers, while the other four symbols contain only odd powers. We solve for the coefficients of increasing powers of c, noting that at each stage we only need to solve a 4×4 system. The outcome of the first step (i.e., if $c=0$) is already known from gravity waves [HIT16]. The computations are somewhat tedious, but quite elementary. We omit them and only write the final result:

$$\begin{aligned}C^a(\xi,\eta)&=-i\frac{c^2}{4g^2}\xi, & B^a(\xi,\eta)&=-i\frac{c^4}{4g^2}\eta^{-1}+i\frac{c^2}{2g}\xi\eta^{-1}+i\frac{c^2}{4g}-i\xi,\\ D^a(\xi,\eta)&=i\frac{c^3}{4g^2}-i\frac{c}{2g}\xi, & E^a(\xi,\eta)&=i\frac{c^3}{4g^2}\xi\eta^{-1}-i\frac{c}{2g}\xi,\\ A^a(\xi,\eta)&=i\frac{c^2}{4g}, & F^a(\xi,\eta)&=-i\frac{c^3}{4g}\eta^{-1}+i\frac{c}{2},\\ H^a(\xi,\eta)&=-i\frac{c}{2g}\xi, & G^a(\xi,\eta)&=i\frac{c^2}{2g}\xi\eta^{-1}-i\xi.\end{aligned} \tag{2-16}$$

□

3. The linearized equation

In this section we derive the linearized water wave equations, and prove energy estimates for them. We recall that for the similar problem in [HIT16] we were able to prove quadratic energy estimates in $\dot{\mathcal{H}}_0$, which apply for the large-data problem, but also cubic energy estimates for the small-data problem. By contrast, here we are only able to prove the quadratic energy estimates. This suffices for the local well-posedness theory, but is not so useful in order to establish improved lifespan bounds. It appears unlikely that cubic energy estimates hold in $\dot{\mathcal{H}}_0$ for the linearized problem; in any case, we leave this question open.

3A. ***Computing the linearization.*** The solutions for the linearized water wave equation around a profile (W, Q) are denoted by (w, q). However, it will be more convenient to immediately switch to diagonal variables (w, r), where

$$r := q - Rw.$$

We first recall the equations,

$$\begin{cases} W_t + \underline{F}(W_\alpha + 1) + i\frac{c}{2}W = 0, \\ Q_t - igW + \underline{F}Q_\alpha + icQ + \boldsymbol{P}[|R|^2] - i\frac{c}{2}T_1 = 0, \end{cases} \tag{3-1}$$

where $\underline{F} = F - i\frac{c}{2}F_1$ with

$$F = \boldsymbol{P}\left[\frac{R}{1+\overline{\boldsymbol{W}}} - \frac{\bar{R}}{1+\boldsymbol{W}}\right], \quad F_1 = \boldsymbol{P}\left[\frac{W}{1+\overline{\boldsymbol{W}}} + \frac{\overline{W}}{1+\boldsymbol{W}}\right], \quad T_1 = \boldsymbol{P}[W\bar{R} - \overline{W}R].$$

The linearization of R is

$$\delta R = \frac{q_\alpha - Rw_\alpha}{1+\boldsymbol{W}} = \frac{r_\alpha + R_\alpha w}{1+\boldsymbol{W}},$$

and that of $|R|^2$ is

$$\delta|R|^2 = n + \bar{n}, \quad n := \bar{R}\delta R = \frac{\bar{R}(r_\alpha + R_\alpha w)}{1+\boldsymbol{W}}.$$

The linearizations of F, F_1 and T_1 can be expressed in the form

$$\delta F = \boldsymbol{P}[m - \bar{m}], \quad \delta F_1 = \boldsymbol{P}[m_1 + \bar{m}_1], \quad \delta T_1 = \boldsymbol{P}[m_2 - \bar{m}_2],$$

where the auxiliary variables m, m_1, m_2 correspond to differentiating F, F_1 and T_1 with respect to the holomorphic variables,

$$m := \frac{q_\alpha - Rw_\alpha}{J} + \frac{\bar{R}w_\alpha}{(1+\boldsymbol{W})^2} = \frac{r_\alpha + R_\alpha w}{J} + \frac{\bar{R}w_\alpha}{(1+\boldsymbol{W})^2},$$

$$m_1 := \frac{1}{1+\overline{\boldsymbol{W}}}w - \frac{\overline{W}}{(1+\boldsymbol{W})^2}w_\alpha, \quad m_2 := \bar{R}w - \frac{\overline{W}r_\alpha + \overline{W}R_\alpha w}{1+\boldsymbol{W}}.$$

Given all of the above, it follows that the linearized water wave equations take the form

$$\begin{cases} w_t + \underline{F} w_\alpha + (1+\boldsymbol{W})\Big(\boldsymbol{P}[m-\bar{m}] - i\dfrac{c}{2}\boldsymbol{P}[m_1+\bar{m}_1]\Big) + i\dfrac{c}{2}w = 0, \\ q_t + \underline{F} q_\alpha + Q_\alpha\Big(P[m-\bar{m}] - i\dfrac{c}{2}\boldsymbol{P}[m_1+\bar{m}_1]\Big) - igw + icq + \boldsymbol{P}[n+\bar{n}] - i\dfrac{c}{2}\boldsymbol{P}[m_2-\bar{m}_2] = 0. \end{cases}$$

Now we transition in both equations from $\underline{F}$ to the real advection coefficient $\underline{b}$ using the relations (1-16). We also move to the right all the terms which we expect to be perturbative. These are terms like $\bar{P}m$, $\bar{P}n$, $\bar{P}m_1$, $\bar{P}m_2$, which are lower-order since the differentiated holomorphic variables have to be lower-frequency. The same applies to their conjugates. Then, our equations are rewritten as

$$\begin{cases} (\partial_t + \underline{b}\partial_\alpha)w + \left[i\dfrac{c}{2}w + (1+\boldsymbol{W})\Big(m - i\dfrac{c}{2}m_1\Big) - \dfrac{\bar{R}w_\alpha}{1+\boldsymbol{W}} - i\dfrac{c}{2}\dfrac{\bar{W}w_\alpha}{1+\boldsymbol{W}}\right] = \underline{\mathcal{G}}_0 \\ (\partial_t + \underline{b}\partial_\alpha)q - igw + icq + \left[Q_\alpha\Big(m - i\dfrac{c}{2}m_1\Big) + n - i\dfrac{c}{2}m_2 - \dfrac{\bar{R}q_\alpha}{1+\boldsymbol{W}} - i\dfrac{c}{2}\dfrac{\bar{W}q_\alpha}{1+\boldsymbol{W}}\right] = \underline{\mathcal{K}}_0, \end{cases}$$

where

$$\begin{cases} \underline{\mathcal{G}}_0 = (1+\boldsymbol{W})\Big((\boldsymbol{P}\bar{m} + \bar{\boldsymbol{P}}m) + i\dfrac{c}{2}(\boldsymbol{P}\bar{m}_1 - \bar{\boldsymbol{P}}m_1)\Big), \\ \underline{\mathcal{K}}_0 = Q_\alpha\Big((\boldsymbol{P}\bar{m} + \bar{\boldsymbol{P}}m) + i\dfrac{c}{2}(\boldsymbol{P}\bar{m}_1 - \bar{\boldsymbol{P}}m_1)\Big) + (\bar{\boldsymbol{P}}n - \boldsymbol{P}\bar{n}) - i\dfrac{c}{2}(\boldsymbol{P}\bar{m}_2 + \bar{\boldsymbol{P}}m_2). \end{cases}$$

Taking advantage of algebraic cancellations in the square brackets above we are left with

$$\begin{cases} (\partial_t + \underline{b}\partial_\alpha)w + \dfrac{1}{1+\bar{\boldsymbol{W}}}\Big(r_\alpha + R_\alpha w + i\dfrac{c}{2}(\bar{\boldsymbol{W}} - \boldsymbol{W})w\Big) = \underline{\mathcal{G}}_0, \\ (\partial_t + \underline{b}\partial_\alpha)q - igw + icq + \dfrac{R}{1+\bar{\boldsymbol{W}}}\Big(r_\alpha + R_\alpha w + i\dfrac{c}{2}(\bar{\boldsymbol{W}} - \boldsymbol{W})w\Big) - i\dfrac{c}{2}(R+\bar{R})w = \underline{\mathcal{K}}_0. \end{cases}$$

Now we can switch from q to $r = q - Rw$ and obtain a diagonalized system, namely

$$\begin{cases} (\partial_t + \underline{b}\partial_\alpha)w + \dfrac{1}{1+\bar{\boldsymbol{W}}}\Big(r_\alpha + R_\alpha w + i\dfrac{c}{2}(\bar{\boldsymbol{W}} - \boldsymbol{W})w\Big) = \underline{\mathcal{G}}_0, \\ (\partial_t + \underline{b}\partial_\alpha)r + icr - i\Big(g + \dfrac{c}{2}(\bar{R} - R) + i(\partial_t + \underline{b}\partial_\alpha)R\Big)w = \underline{\mathcal{K}}_0 - R\underline{\mathcal{G}}_0. \end{cases}$$

For the expression $(\partial_t + \underline{b}\partial_\alpha)R$ in the second equation we use (1-19) and compute the coefficient of w as follows:

$$\begin{aligned} g + \frac{c}{2}(\bar{R} - R) - i(\partial_t + \underline{b}\partial_\alpha)R &= g + \frac{c}{2}(\bar{R} - R) + cR - \frac{g\boldsymbol{W} - a}{1+\boldsymbol{W}} - \frac{c}{2}\frac{R\boldsymbol{W} + \bar{R}\boldsymbol{W} + N}{1+\boldsymbol{W}} \\ &= \frac{g + a + (c/2)(R + \bar{R} - N)}{1+\boldsymbol{W}}. \end{aligned}$$

This motivates the definition of $\underline{a}$ in (1-12). With this notation, we write the final form of the linearized equations as

$$\begin{cases} (\partial_t + \underline{b}\partial_\alpha)w + \dfrac{r_\alpha}{1+\bar{\boldsymbol{W}}} + \dfrac{R_\alpha w}{1+\bar{\boldsymbol{W}}} = \underline{\mathcal{G}}(w,r), \\ (\partial_t + \underline{b}\partial_\alpha)r + icr - i\dfrac{g+\underline{a}}{1+\boldsymbol{W}}w = \underline{\mathcal{K}}(w,r), \end{cases} \tag{3-2}$$

where $\underline{\mathcal{G}}(w,r)$, and $\underline{\mathcal{K}}(w,r)$ are given by

$$\underline{\mathcal{G}}(w,r) = \mathcal{G}(w,r) - i\frac{c}{2}\mathcal{G}_1(w,r), \quad \underline{\mathcal{K}}(w,r) = \mathcal{K}(w,r) - i\frac{c}{2}\mathcal{K}_1(w,r),$$

where

$$\mathcal{G}(w,r) = (1+\boldsymbol{W})(\boldsymbol{P}\bar{m} + \bar{\boldsymbol{P}}m), \quad \mathcal{G}_1(w,r) = -(1+\boldsymbol{W})(\boldsymbol{P}\bar{m}_1 - \bar{\boldsymbol{P}}m_1) + \frac{(\bar{\boldsymbol{W}} - \boldsymbol{W})w}{1+\bar{\boldsymbol{W}}}, \tag{3-3}$$

$$\mathcal{K}(w,r) = \bar{\boldsymbol{P}}n - \boldsymbol{P}\bar{n}, \qquad \mathcal{K}_1(w,r) = \boldsymbol{P}\bar{m}_2 + \bar{\boldsymbol{P}}m_2. \tag{3-4}$$

If $c = 0$, then these equations coincide with those in [HIT16]. The fact that $g + \underline{a}$ is real and positive is crucial for the well-posedness of the linearized system.

We remark that while (w, r) are holomorphic, it is not directly obvious that the above evolution preserves the space of holomorphic states. To remedy this, one can also project the linearized equations onto the space of holomorphic functions via the projection $\boldsymbol{P}$. Then we obtain the equations

$$\begin{cases} (\partial_t + \mathfrak{M}_{\underline{b}}\partial_\alpha)w + \boldsymbol{P}\left[\dfrac{1}{1+\bar{\boldsymbol{W}}}r_\alpha\right] + \boldsymbol{P}\left[\dfrac{R_\alpha}{1+\bar{\boldsymbol{W}}}w\right] = \boldsymbol{P}\underline{\mathcal{G}}(w,r), \\ (\partial_t + \mathfrak{M}_{\underline{b}}\partial_\alpha)r + icr - i\boldsymbol{P}\left[\dfrac{g+\underline{a}}{1+\boldsymbol{W}}w\right] = \boldsymbol{P}\underline{\mathcal{K}}(w,r). \end{cases} \tag{3-5}$$

Since the original set of equations (1-5) is fully holomorphic, it follows that the two sets of equations (3-2) and (3-5) are algebraically equivalent.

In order to investigate the possibility of cubic linearized energy estimates it is also of interest to separate the quadratic parts $\underline{\mathcal{G}}^2$ and $\underline{\mathcal{K}}^2$ of $\underline{\mathcal{G}}$ and $\underline{\mathcal{K}}$. The holomorphic quadratic parts of $\mathcal{G}$ and $\mathcal{K}$, which also appear in [HIT16], are given by

$$\boldsymbol{P}\mathcal{G}^{(2)}(w,r) = -P[\boldsymbol{W}\bar{r}_\alpha] + P[R\bar{w}_\alpha], \quad \boldsymbol{P}\mathcal{K}^{(2)}(w,r) = -\boldsymbol{P}[R\bar{r}_\alpha].$$

Next we compute the similar decomposition for $\mathcal{G}_1^2$ and $\mathcal{K}_1^2$, since the rest was done in [HIT16]. These are split into quadratic and cubic-and-higher terms as shown below:

$$\mathcal{G}_1 = \mathcal{G}_1^{(2)} + \mathcal{G}_1^{(3+)}, \quad \mathcal{K}_1 = \mathcal{K}_1^{(2)} + \mathcal{K}_1^{(3+)}.$$

For the quadratic parts we have the holomorphic components

$$\begin{cases} \boldsymbol{P}\mathcal{G}_1^{(2)}(w,r) = \boldsymbol{P}[\boldsymbol{W}\bar{w}] + \boldsymbol{P}[W\bar{w}_\alpha] + \boldsymbol{P}[\bar{\boldsymbol{W}}w] - \boldsymbol{P}[\boldsymbol{W}w], \\ \boldsymbol{P}\mathcal{K}_1^{(2)}(w,r) = \boldsymbol{P}[W\bar{r}_\alpha] - \boldsymbol{P}[R\bar{w}], \end{cases} \tag{3-6}$$

and the antiholomorphic components

$$\begin{cases} \bar{\boldsymbol{P}}\mathcal{G}_1^{(2)}(w,r) = -\bar{\boldsymbol{P}}[\bar{\boldsymbol{W}}w] - \bar{\boldsymbol{P}}[\bar{W}w_\alpha] + \bar{\boldsymbol{P}}[\bar{\boldsymbol{W}}w], \\ \bar{\boldsymbol{P}}\mathcal{K}_1^{(2)}(w,r) = \bar{\boldsymbol{P}}[\bar{W}r_\alpha] - \bar{\boldsymbol{P}}[\bar{R}w]. \end{cases}$$

The cubic terms have the form

$$\begin{cases} \mathcal{G}_1^{(3+)}(w,r) = -\boldsymbol{W}(\boldsymbol{P}\bar{m}_1 - \bar{\boldsymbol{P}}m_1) - (\boldsymbol{P}\bar{m}_1^{(3+)} - \bar{\boldsymbol{P}}m_1^{(3+)}) - \bar{Y}(\bar{\boldsymbol{W}} - \boldsymbol{W})w, \\ \mathcal{K}_1^{(3+)}(w,r) = \bar{\boldsymbol{P}}m_2^{(3+)} + \boldsymbol{P}\bar{m}_2^{(3+)}. \end{cases}$$

For the purpose of simplifying nonlinear estimates, it is convenient to express $\mathcal{G}_1^{(3)}$ and $\mathcal{K}_1^{(3)}$ in a polynomial fashion. This is done using the variable $Y = \boldsymbol{W}/(1+\boldsymbol{W})$. Then we have

$$\begin{aligned}\bar{\boldsymbol{P}}m_1 &= -\bar{\boldsymbol{P}}[\bar{Y}w + (Y^2 - 2Y + 1)\overline{W}w_\alpha],\\ \bar{\boldsymbol{P}}m^{(3+)} &= \bar{\boldsymbol{P}}[(2Y - Y^2)\overline{W}w_\alpha],\\ \bar{\boldsymbol{P}}m_2^{(3+)} &= \bar{\boldsymbol{P}}[\overline{W}Yr_\alpha - \overline{W}R_\alpha(1-Y)w].\end{aligned}$$

3B. ***Quadratic estimates for large data.*** Our goal here is to study the well-posedness of the system (3-5) in $L^2 \times \dot{H}^{1/2}$. We begin with a more general version of the system (3-5), namely

$$\begin{cases}(\partial_t + \mathfrak{M}_{\underline{b}}\partial_\alpha)w + \boldsymbol{P}\left[\dfrac{1}{1+\overline{\boldsymbol{W}}}r_\alpha\right] + \boldsymbol{P}\left[\dfrac{R_\alpha}{1+\overline{\boldsymbol{W}}}w\right] = G,\\ (\partial_t + \mathfrak{M}_{\underline{b}}\partial_\alpha)r + icr - i\boldsymbol{P}\left[\dfrac{g+\underline{a}}{1+\boldsymbol{W}}w\right] = K,\end{cases} \tag{3-7}$$

and define the associated linear energy

$$E_{\text{lin}}^{(2)}(w,r) = \int_{\mathbb{R}} (g+\underline{a})|w|^2 + \Im(r\bar{r}_\alpha)\,d\alpha.$$

We note that the positivity of the energy is closely related to the Taylor sign condition (1-26). Therefore, when $\underline{a}$ is positive, we have

$$E_{\text{lin}}^{(2)}(w,r) \approx_A \mathcal{E}_0(w,r).$$

Our first result uses the control parameters $A_{-1/2}$, A and B defined in (1-22), (1-20), and(1-21), respectively in order to establish (nearly) cubic bounds for the system (3-7):

Proposition 3.1. *The linear equation (3-7) is well-posed in $\dot{\mathcal{H}}_0$, and the following estimate holds*:

$$\frac{d}{dt}E_{\text{lin}}^{(2)}(w,r) = 2\Re\int_{\mathbb{R}} (g+\underline{a})\bar{w}G - i\bar{r}_\alpha K + c^2\Im R|w|^2\,d\alpha + O_A(\underline{A}\underline{B})E_{\text{lin}}^{(2)}(w,r). \tag{3-8}$$

We will also need a weighted version of this:

Lemma 3.2. *Let f be a real function and (w,r) solutions to (3-7). Then for the difference*

$$I := \frac{d}{dt}\Re\int f((g+\underline{a})|w|^2 - i\bar{r}_\alpha r)\,d\alpha - \Re\int f((g+\underline{a})\bar{w}F - i\bar{r}_\alpha G) + (\partial_t + b\partial_\alpha)f((g+\underline{a})|w|^2 - i\bar{r}_\alpha r)\,d\alpha$$

we have the estimate

$$|I| \lesssim_A (\underline{B}\|f\|_{L^\infty} + \underline{A}\||D|^{1/2}f\|_{\text{BMO}})\,\|(w,r)\|_{L^2\times\dot{H}^{1/2}}^2. \tag{3-9}$$

Our main use for the result in Proposition 3.1 is to apply it to the linearized equation (3-5):

Proposition 3.3. *The linearized equation (3-5) is well-posed in $L^2 \times \dot{H}^{1/2}$, and the following estimate holds*:

$$\frac{d}{dt}E_{\text{lin}}^{(2)}(w,r) \lesssim_A (\underline{B} + c\underline{A})E_{\text{lin}}^{(2)}(w,r). \tag{3-10}$$

Proof of Proposition 3.1. A direct computation yields

$$\frac{d}{dt}\int (g+\underline{a})|w|^2\,d\alpha = 2\Re\int (g+\underline{a})\bar{w}(\partial_t+\mathfrak{M}_{\underline{b}}\partial_\alpha)w+\underline{a}\bar{w}[\underline{b},\boldsymbol{P}]w_\alpha\,d\alpha+\int[\underline{a}_t+((g+\underline{a})\underline{b})_\alpha]|w|^2\,d\alpha.$$

A similar computation shows that

$$\frac{d}{dt}\int \Im(r\partial_\alpha\bar{r})\,d\alpha = 2\Im\int(\partial_t+\mathfrak{M}_{\underline{b}}\partial_\alpha)r\partial_\alpha\bar{r}\,d\alpha.$$

Adding the two and using the equations (3-7), the quadratic $\Re(w\bar{r}_\alpha)$-term cancels modulo another commutator term, and we obtain

$$\frac{d}{dt}E^{(2)}_{\mathrm{lin}}(w,r) = 2\Re\int (g+\underline{a})\bar{w}G - i\bar{r}_\alpha K\,d\alpha + c^2\int \Im R|w|^2\,d\alpha + \underline{\mathrm{err}}_1, \tag{3-11}$$

where

$$\begin{aligned}\underline{\mathrm{err}}_1 = \int[\underline{a}_t+((g+\underline{a})\underline{b})_\alpha - c^2\Im R]|w|^2\,d\alpha - 2\Re\int (g+\underline{a})\frac{R_\alpha}{1+\overline{\boldsymbol{W}}}|w|^2\,d\alpha\\ -2\Re\int \underline{a}\bar{w}[\bar{Y},\boldsymbol{P}](r_\alpha+R_\alpha w)\,d\alpha - 2\Re\int \underline{a}\bar{w}[\boldsymbol{P},\underline{b}]w_\alpha\,d\alpha.\end{aligned}$$

Using the auxiliary function $\underline{M}$ in (1-17), we rewrite it as

$$\begin{aligned}\underline{\mathrm{err}}_1 = \int (\underline{a}_t+\underline{b}\underline{a}_\alpha - c^2\Im R)|w|^2 - (g+\underline{a})\Big(i\frac{c}{2}(\boldsymbol{W}-\overline{\boldsymbol{W}})+\underline{M}\Big)|w|^2\,d\alpha\\ -2\Re\int \underline{a}\bar{w}([\bar{Y},\boldsymbol{P}](r_\alpha+R_\alpha w)+[\boldsymbol{P},\underline{b}]w_\alpha)\,d\alpha.\end{aligned}$$

At this point we need to re-express the error in terms of a, a_1, b, b_1, M, and M_1. The terms which do not have any c-factors were already estimated in [HIT16], so we only need to worry about the remaining components. There is only one exception: the first term in $\underline{\mathrm{err}}_1$, whose counterpart was estimated separately in [HIT16], and which we also estimate separately here. Thus, the error we want to bound is composed of err_1 (the same as in [HIT16]) and additional terms, which we call $\mathrm{err}_{1,1}$:

$$\underline{\mathrm{err}}_1 = \mathrm{err}_1 + \mathrm{err}_{1,1},$$

where $\mathrm{err}_{1,1}$ is further separated into terms

$$\mathrm{err}_{1,1} := \mathrm{err}^1_{1,1}+\mathrm{err}^2_{1,1}+\mathrm{err}^3_{1,1}-\mathrm{err}^4_{1,1}+\mathrm{err}^5_{1,1}+\mathrm{err}^6_{1,1}, \tag{3-12}$$

which will be estimated separately, and are listed below:

$$\begin{aligned}\mathrm{err}^1_{1,1} &:= \int (a_t+\underline{b}a_\alpha)|w|^2\,d\alpha,\\ \mathrm{err}^2_{1,1} &:= \frac{c}{2}\int (a_{1,t}+\underline{b}a_{1,\alpha} - ig(\boldsymbol{W}-\overline{\boldsymbol{W}}) - 2c\Im R)|w|^2\,d\alpha,\\ \mathrm{err}^3_{1,1} &:= \frac{c}{2}\int (i(g+a)M_1 - a_1M - ia(\boldsymbol{W}-\overline{\boldsymbol{W}}))|w|^2\,d\alpha,\\ \mathrm{err}^4_{1,1} &:= 2\Re\int i\frac{c}{2}a\bar{w}[\boldsymbol{P},b_1]w_\alpha\,d\alpha,\end{aligned}$$

$$\mathrm{err}_{1,1}^5 := 2\Re \int \frac{c}{2} a_1 \bar{w} \{[\bar{Y}, \boldsymbol{P}](r_\alpha + R_\alpha w) + [\boldsymbol{P}, b] w_\alpha\} \, d\alpha,$$
$$\mathrm{err}_{1,1}^6 := \Re \int i \frac{c^2}{4} a_1 \bar{w} (2[P, b_1] w_\alpha + (M_1 - \boldsymbol{W} + \overline{\boldsymbol{W}}) w) \, d\alpha.$$

To conclude the proof of (3-8) we want to show that

$$|\underline{\mathrm{err}}_1| \lesssim_A (A + cA_{-1/2})(B + cA) E^{(2)}_{\mathrm{lin}}(w, r). \tag{3-13}$$

From [HIT16] we have the bound

$$|\mathrm{err}_1| \lesssim_A AB E^{(2)}_{\mathrm{lin}}(w, r),$$

so it remains to establish (3-13) for each of the $\mathrm{err}_{1,1}$ components. The positive constant c has the role of a scaling parameter; therefore it makes sense to group terms accordingly.

The bound for $\mathrm{err}_{1,1}^1$. This follows by Proposition A.3 in Appendix A.

The bound for $\mathrm{err}_{1,1}^2$. This follows by Proposition A.4, which we included in Appendix A.

The bound for $\mathrm{err}_{1,1}^3$. Here we use the pointwise estimates

$$\|a\|_{L^\infty} \lesssim A^2, \quad \|a_1\|_{L^\infty} \lesssim_A A_{-1/2}, \quad \|M\|_{L^\infty} \lesssim_A AB, \quad \|M_1\|_{L^\infty} \lesssim_A A^2.$$

The first and the third are from [HIT16], while for the second and fourth we use Lemma A.2 and Proposition A.4. This yields

$$|\mathrm{err}_{1,1}^2| \lesssim c(A^2 + ABA_{-1/2} + A^3 + A^4) E^{(2)}_{\mathrm{lin}}(w, r).$$

The bound for $\mathrm{err}_{1,1}^4$. For the terms in $\mathrm{err}_{1,1}^3$ we use the pointwise bounds of a and $b_{1,\alpha}$ (recall that $b_{1,\alpha} = \boldsymbol{W} - \overline{\boldsymbol{W}} - M_1$), which were obtained in [HIT16, Proposition 8.6] and in Lemma A.2, respectively:

$$|a| \lesssim A^2, \quad \|b_{1,\alpha}\|_{L^\infty} \lesssim A + A^2.$$

To estimate the commutator in L^2 we use a Coifman–Meyer bound, see, e.g., Lemma 8.1 in [HIT16]:

$$\|[\boldsymbol{P}, b_1] w_\alpha\|_{L^2} \lesssim \|b_{1,\alpha}\|_{L^\infty} \|w\|_{L^2}.$$

Combining the results above leads to

$$|\mathrm{err}_{1,1}^3| \lesssim c(A^3 + A^4) E^{(2)}_{\mathrm{lin}}(w, r).$$

The bound for $\mathrm{err}_{1,1}^5$. After using the above pointwise bound on a_1 it remains to estimate the commutators in L^2 (as above). The only difficulty here is that we need to move half of a derivative from r_α onto Y. Such estimates were already considered in [HIT16],

$$\|[\bar{Y}, \boldsymbol{P}] r_\alpha\|_{L^2} \lesssim \| |D|^{1/2} Y\|_{\mathrm{BMO}} \|r\|_{\dot{H}^{1/2}}, \qquad \|[\boldsymbol{P}, b] w_\alpha\|_{L^2} \lesssim \|b_\alpha\|_{\mathrm{BMO}} \|w\|_{L^2},$$

and suffice due to the bounds for b, a_1 and Y in Lemmas 2.5 and 2.7 in [HIT16], and Proposition A.4. For the remaining term in $\mathrm{err}_{1,1}^4$ we write $[\bar{Y}, \boldsymbol{P}](R_\alpha w) = [\overline{\boldsymbol{P}}, \overline{\boldsymbol{P}}[\bar{Y} R_\alpha]] w$ and use the Coifman–Meyer lemma [1976] (also discussed in Appendix B in [HIT16]) to estimate

$$\|\bar{\boldsymbol{P}}[\bar{\boldsymbol{P}}[\bar{Y} R_\alpha]w]\|_{L^2} \lesssim \|w\|_{L^2} \|\bar{\boldsymbol{P}}[\bar{Y} R_\alpha]\|_{\mathrm{BMO}} \lesssim \|w\|_{L^2} \||D|^{1/2} Y\|_{\mathrm{BMO}} \||D|^{1/2} R\|_{\mathrm{BMO}},$$

where the bilinear bound in the second step follows after a bilinear Littlewood–Paley decomposition (again, see [HIT16]). Hence,

$$|\mathrm{err}^4_{1,1}| \lesssim_A cA_{-1/2} ABE^{(2)}_{\mathrm{lin}}(w,r).$$

The bound for $\mathrm{err}^6_{1,1}$. Here we use the pointwise bounds on a_1, M_1 and $b_{1,\alpha}$, along with the Coifman–Meyer commutator estimate.

This concludes the proof of (3-13). □

Proof of Lemma 3.2. We rewrite I in the form

$$I := D_1 + D_2 + D_3 + D_4,$$

where

$$D_1 = \int f(\partial_t + \underline{b}\partial_\alpha)\underline{a}|w|^2\, d\alpha, \quad D_3 = \int f\left(b_\alpha |w|^2 - 2\Re\left(\bar{w}\boldsymbol{P}\left[\frac{R_\alpha}{1+\bar{W}} w\right]\right)\right) d\alpha,$$
$$D_2 = \Re \int f\bar{\boldsymbol{P}}[\underline{b} w_\alpha]\bar{w}\, d\alpha, \qquad D_4 = \int if\left(\bar{r}_\alpha \boldsymbol{P}\left[\frac{(g+a)w}{1+W}\right] - (g+a)\bar{w}\boldsymbol{P}\left[\frac{r_\alpha}{1+W}\right]\right) d\alpha.$$

For the first term we use the pointwise bounds in Propositions A.3 and A.4.

For the second term we directly use a Coifman–Meyer estimate to bound the middle factor in L^2.

For the third term we use the pointwise bound on b_1, and then harmlessly replace b_α by $\boldsymbol{P}[R/(1+\bar{W})]_\alpha$. Then it remains to estimate in L^2 the difference

$$\boldsymbol{P}\left[\frac{R}{1+\bar{W}}\right]_\alpha w - \boldsymbol{P}\left[\frac{R_\alpha}{1+\bar{W}} w\right].$$

If w is the low-frequency factor in either term, then we only need its cofactor in BMO and we win. Else we can drop the first projection, and we are left with estimating the difference

$$\boldsymbol{P}\left[\frac{R}{1+\bar{W}}\right]_\alpha - \frac{R_\alpha}{1+\bar{W}}$$

in L^∞, which was done in [HIT16].

Finally, the last term cancels if we drop the projections. Hence we are left with estimating in L^2 the expression

$$\bar{\boldsymbol{P}}\left[\frac{r_\alpha}{1+W}\right],$$

which is done by a Coifman–Meyer estimate. In the first term we bound the expression

$$\bar{\boldsymbol{P}}\left[\frac{aw}{1+W}\right]$$

in L^2 and then move a half-derivative on f. □

Proof of Proposition 3.3. To estimate the terms involving $\mathcal{G}$ and $\mathcal{K}$ we separate the quadratic and cubic parts, but more importantly we group these expressions keeping track of the scaling parameter c. In our previous paper [HIT16] we have already established the bounds for the components without c, namely

$$\|\boldsymbol{P}\mathcal{G}^{(2)}(w,r)\|_{L^2} + \|\boldsymbol{P}\mathcal{K}^{(2)}(w,r)\|_{\dot{H}^{1/2}} \lesssim B(\|w\|_{L^2} + \|r\|_{\dot{H}^{1/2}}),$$

while the cubic-and-higher terms satisfy

$$\|\boldsymbol{P}\mathcal{G}^{(3+)}(w,r)\|_{L^2} + \|\boldsymbol{P}\mathcal{K}^{(3+)}(w,r)\|_{\dot{H}^{1/2}} \lesssim_A AB(\|w\|_{L^2} + \|r\|_{\dot{H}^{1/2}}).$$

Hence it suffices to estimate the terms with the c-factor, and show that the quadratic terms satisfy

$$\|\boldsymbol{P}\mathcal{G}_1^{(2)}(w,r)\|_{L^2} + \|\boldsymbol{P}\mathcal{K}_1^{(2)}(w,r)\|_{\dot{H}^{1/2}} \lesssim A(\|w\|_{L^2} + \|r\|_{\dot{H}^{1/2}}), \tag{3-14}$$

while the cubic-and-higher terms satisfy

$$\|\boldsymbol{P}\mathcal{G}_1^{(3+)}(w,r)\|_{L^2} + \|\boldsymbol{P}\mathcal{K}_1^{(3+)}(w,r)\|_{\dot{H}^{1/2}} \lesssim_A A^2(\|w\|_{L^2} + \|r\|_{\dot{H}^{1/2}}). \tag{3-15}$$

In order to obtain the estimates claimed in (3-14), (3-15) we use the Coifman–Meyer-type commutator estimates [1976] described in [HIT16, Appendix B, Lemma 8.1],

The bounds for all terms in $\boldsymbol{P}\mathcal{G}_1^{(2)}(w,r)$ are immediate, except for the second, where we use (B.10) in [HIT16] with $s=\frac{1}{2}$ and $\sigma=\frac{1}{2}$, to write

$$\|[\boldsymbol{P}, W]\bar{w}_\alpha\|_{L^2} \lesssim \|\boldsymbol{W}\|_{L^\infty}\,\|w\|_{L^2}.$$

For $\boldsymbol{P}\mathcal{K}_1^{(2)}(w,r)$ we use again (B.10) in [HIT16] with $s=\frac{1}{2}$, and $\sigma=\frac{1}{2}$, and conclude that

$$\|[\boldsymbol{P}, W]\bar{r}_\alpha\|_{\dot{H}^{1/2}} \lesssim \||D|^{1/2}W\|_{L^\infty}\,\|r\|_{\dot{H}^{1/2}}, \qquad \|[\boldsymbol{P}, R]\bar{w}\|_{\dot{H}^{1/2}} \lesssim \||D|^{1/2}R\|_{L^\infty}\,\|w\|_{L^2}.$$

Thus (3-14) follows.

For the cubic-and-higher parts of $\mathcal{G}_1$ and $\mathcal{K}_1$ we apply the same type of commutator estimates, as well as the BMO bounds in [Proposition 8.2, HIT16]. Precisely, in $\mathcal{G}_1^{(3+)}$ there are three nontrivial terms to be estimated in L^2, namely

$$\boldsymbol{P}[\bar{w}_\alpha(\bar{Y}^2-2\bar{Y})W], \quad W\boldsymbol{P}[\bar{w}_\alpha(\bar{Y}^2-2\bar{Y}+1)W], \quad \boldsymbol{P}[W\bar{\boldsymbol{P}}[w_\alpha(Y^2-2Y+1)\bar{W}]].$$

For the first two we use (B.12) and (B.14) from [HIT16] as follows:

$$\|\boldsymbol{P}[\bar{w}_\alpha(\bar{Y}^2-2\bar{Y})W]\|_{L^2} \lesssim \|w\|_{L^2}\,\|\partial_\alpha\boldsymbol{P}[(\bar{Y}^2-2\bar{Y})W]\|_{\mathrm{BMO}} \lesssim_A \|\boldsymbol{W}\|_{\mathrm{BMO}}\,\|Y\|_{L^\infty}\,\|w\|_{L^2}.$$

Similarly, for the remaining terms we have

$$\|\boldsymbol{P}[\bar{w}_\alpha(\bar{Y}^2-2\bar{Y}+1)W]\|_{L^2} \lesssim_A \|\boldsymbol{W}\|_{\mathrm{BMO}}\|w\|_{L^2}.$$

Finally, we estimate the cubic component of $\mathcal{K}_1$; we again use (B.12) and (B.15) from [HIT16], and obtain

$$\||D|^{1/2}\boldsymbol{P}[\bar{r}_\alpha\bar{Y}W]\|_{L^2} \lesssim \|r\|_{\dot{H}^{1/2}}\,\|\partial_\alpha\boldsymbol{P}[\bar{Y}W]\|_{\mathrm{BMO}} \lesssim_A \|r\|_{\dot{H}^{1/2}}\,\|Y\|_{L^\infty}\,\|\boldsymbol{W}\|_{L^\infty},$$

and also

$$\| |D|^{1/2} \boldsymbol{P}[\bar{w}(1-\bar{Y})\bar{R}_\alpha W]\|_{L^2} \lesssim \|w(1-Y)\|_{L^2} \, \| |D|^{1/2} \boldsymbol{P}[\bar{R}_\alpha W]\|_{\mathrm{BMO}}$$
$$\lesssim_A \|w\|_{L^2} \, \| |D|^{1/2} R\|_{\mathrm{BMO}} \, \|\boldsymbol{W}\|_{L^\infty}.$$

This concludes the proof of (3-15), and thus of the proposition. □

3C. ***Nearly cubic estimates for small data.*** Our goal here is to investigate the possibility of obtaining cubic estimates for the system (3-5) in $L^2 \times \dot{H}^{1/2}$. Unlike in [HIT16], this is no longer possible. Instead, our more limited goal in this section is to identify a main portion of the linearized equation for which cubic estimates are valid, precisely up to c^2-terms. This will come in very handy later on in the proof of cubic estimates for the differentiated equation.

Our model problem this time is the following subset of (3-5):

$$\begin{cases} (\partial_t + \mathfrak{M}_{\underline{b}}\partial_\alpha)w + \boldsymbol{P}\left[\dfrac{1}{1+\overline{\boldsymbol{W}}} r_\alpha\right] + \boldsymbol{P}\left[\dfrac{R_\alpha}{1+\overline{\boldsymbol{W}}} w\right] = -\boldsymbol{P}[\boldsymbol{W}\bar{r}_\alpha] + \boldsymbol{P}[R\bar{w}_\alpha] + G, \\ (\partial_t + \mathfrak{M}_{\underline{b}}\partial_\alpha)r + icr - i\boldsymbol{P}\left[\dfrac{g+\underline{a}}{1+\boldsymbol{W}} w\right] = -\boldsymbol{P}[R\bar{r}_\alpha] + K. \end{cases} \tag{3-16}$$

In our previous work [HIT16], for the case $c=0$, we identified a cubic correction to $E^{(2)}_{\mathrm{lin}}(w,r)$ for which cubic estimates hold for solutions to the linearized equation, namely

$$E^{(3)}_{\mathrm{lin}}(w,r) := \int_{\mathbb{R}} (g+\underline{a})|w|^2 + \Im(r\bar{r}_\alpha) + 2\Im(\bar{R}wr_\alpha) - 2\Re(\overline{\boldsymbol{W}}w^2)\, d\alpha.$$

Our next result asserts that in our case the time derivative of $E^{(3)}_{\mathrm{lin}}$ will be quartic at the leading order, but will have a cubic terms with a coefficient of c^2.

Proposition 3.4. *Assume that $\underline{A}$ is small. Then we have the energy equivalence*

$$E^{(3)}_{\mathrm{lin}}(w,r) \approx E^{(2)}_{\mathrm{lin}}(w,r). \tag{3-17}$$

Further, the linear equation (3-7) is well-posed in $\dot{\mathcal{H}}_0$, and the following estimate holds:

$$\frac{d}{dt} E^{(3)}_{\mathrm{lin}}(w,r) = 2\Re \int_{\mathbb{R}} [(g+\underline{a})\bar{w} - i\bar{R}_\alpha r_\alpha - 2\overline{\boldsymbol{W}}w]G - i[\bar{r}_\alpha - (\bar{R}w)_\alpha]K\, d\alpha$$
$$+ c^2 \Im R|w|^2\, d\alpha + O_A(\underline{A}\,\underline{B}) E^{(2)}_{\mathrm{lin}}(w,r). \tag{3-18}$$

Applying this to the linearized equation (3-5) we obtain:

Proposition 3.5. *The linearized equation (3-5) is well-posed in $L^2 \times \dot{H}^{1/2}$, and the following estimate holds*:

$$\frac{d}{dt} E^{(3)}_{\mathrm{lin}}(w,r) = 2\Re \int_{\mathbb{R}} c[gw\boldsymbol{P}\mathcal{G}^{(2)}_1(w,r) - i\bar{r}_\alpha \boldsymbol{P}\mathcal{K}^{(2)}_1(w,r)] - i\frac{c^2}{2} R|w|^2\, d\alpha + O_A(\underline{A}\,\underline{B}) E^{(2)}_{\mathrm{lin}}(w,r), \tag{3-19}$$

where $\boldsymbol{P}\mathcal{G}^{(2)}_1(w,r)$, $\boldsymbol{P}\mathcal{K}^{(2)}_1(w,r)$ are as in (3-6).

Proof of Proposition 3.4 . The energy equivalence is immediate due to the estimates in [HIT16] for the added cubic terms, and the pointwise bounds on $\underline{a}$ in Lemmas A.3 and A.4.

To prove the estimate in (3-18) we compute the time derivative of the cubic component of the energy $E^{(3)}_{\text{lin}}(w,r)$, using the projected equations for w and r and the unprojected equations for R and $\boldsymbol{W}$:

$$\frac{d}{dt}\biggl(\Im\int \bar{R}wr_\alpha\,d\alpha-\Re\int \overline{\boldsymbol{W}}w^2\,d\alpha\biggr)=\Im\int -ig\overline{\boldsymbol{W}}wr_\alpha-\bar{R}r_\alpha r_\alpha+ig\bar{R}ww_\alpha+\bar{R}r_\alpha G+\bar{R}wK_\alpha\,d\alpha \\ +\Re\int \bar{R}_\alpha w^2+2\overline{\boldsymbol{W}}wr_\alpha-2\overline{\boldsymbol{W}}wG\,d\alpha+\underline{\text{err}}_2,$$

where

$$\begin{aligned}\underline{\text{err}}_2=\Im\int\biggl\{&\biggl(i\biggl(\frac{g\overline{\boldsymbol{W}}^2+a}{1+\overline{\boldsymbol{W}}}\biggr)-\underline{b}\bar{R}_\alpha-i\frac{c}{2}\frac{\bar{R}\overline{\boldsymbol{W}}+R\overline{\boldsymbol{W}}+\bar{N}}{1+\overline{\boldsymbol{W}}}\biggr)wr_\alpha\\
&-\bar{R}w\partial_\alpha\biggl(\mathfrak{M}_{\underline{b}}r_\alpha-i\boldsymbol{P}\biggl[\frac{\underline{a}-\boldsymbol{W}}{1+\boldsymbol{W}}w\biggr]+\boldsymbol{P}[R\bar{r}_\alpha]\biggr)\\
&-\bar{R}r_\alpha\biggl(\mathfrak{M}_{\underline{b}}w_\alpha-\boldsymbol{P}\biggl[\frac{\overline{\boldsymbol{W}}}{1+\overline{\boldsymbol{W}}}r_\alpha\biggr]+\boldsymbol{P}\biggl[\frac{R_\alpha}{1+\overline{\boldsymbol{W}}}w\biggr]+\boldsymbol{P}[\boldsymbol{W}\bar{r}_\alpha-R\bar{w}_\alpha]\biggr)\biggr\}\,d\alpha\\
+\Re\int\biggl\{&w^2\biggl(\underline{b}\overline{\boldsymbol{W}}_\alpha+\frac{\overline{\boldsymbol{W}}-\boldsymbol{W}}{1+\boldsymbol{W}}\bar{R}_\alpha-(1+\overline{\boldsymbol{W}})\underline{M}+i\frac{c}{2}\overline{\boldsymbol{W}}(\overline{\boldsymbol{W}}-\boldsymbol{W})\biggr)\\
&+2\overline{\boldsymbol{W}}w\biggl(\mathfrak{M}_{\underline{b}}w_\alpha-\boldsymbol{P}\biggl[\frac{\overline{\boldsymbol{W}}}{1+\overline{\boldsymbol{W}}}r_\alpha\biggr]+\boldsymbol{P}\biggl[\frac{R_\alpha}{1+\overline{\boldsymbol{W}}}w\biggr]+\boldsymbol{P}[\boldsymbol{W}\bar{r}_\alpha-R\bar{w}_\alpha]\biggr)\biggr\}\,d\alpha.\end{aligned}\tag{3-20}$$

We now separate (3-20) into two parts, namely err_2, which was already estimated in [HIT16], and $\text{err}_{2,1}$:

$$\underline{\text{err}}_2:=\text{err}_2+i\frac{c}{2}\text{err}_{2,1},$$

where

$$\begin{aligned}\text{err}_{2,1}:=\Im\int\biggl\{&\biggl(b_1\bar{R}_\alpha-\frac{\bar{R}\overline{\boldsymbol{W}}+R\overline{\boldsymbol{W}}+\bar{N}}{1+\overline{\boldsymbol{W}}}\biggr)wr_\alpha+\bar{R}w\partial_\alpha\boldsymbol{P}[b_1r_\alpha]-\bar{R}w\boldsymbol{P}\biggl[\frac{a_1}{1+\boldsymbol{W}}w\biggr]+\bar{R}r_\alpha\boldsymbol{P}[b_1w_\alpha]\biggr\}\,d\alpha\\
+\Re\int&\{w^2[-b_1\overline{\boldsymbol{W}}_\alpha-(1+\overline{\boldsymbol{W}})\bar{M}_1+\overline{\boldsymbol{W}}(\overline{\boldsymbol{W}}-\boldsymbol{W})]-2\overline{\boldsymbol{W}}w\boldsymbol{P}[b_1w_\alpha]\}\,d\alpha.\end{aligned}\tag{3-21}$$

We still need to estimate this last error.

Adding $\underline{\text{err}}_2$ to (3-18) (but applied to solutions to (3-16)) gives us

$$\begin{aligned}\frac{d}{dt}E^{(3)}_{\text{lin}}(w,r)&=\frac{d}{dt}E^{(2)}_{\text{lin}}(w,r)+2\frac{d}{dt}\biggl\{\biggl(\Im\int \bar{R}wr_\alpha\,d\alpha-\Re\int \overline{\boldsymbol{W}}w^2\,d\alpha\biggr)\biggr\}\\
&=2\Re\int_{\mathbb{R}}(g+\underline{a})\bar{w}\{G-\boldsymbol{P}[\boldsymbol{W}\bar{r}_\alpha]+\boldsymbol{P}[R\bar{w}_\alpha]\}-i\bar{r}_\alpha\{K-\boldsymbol{P}[R\bar{r}_\alpha]\}\,d\alpha\\
&\quad+c^2\Im\int_{\mathbb{R}}R|w|^2\,d\alpha+\underline{\text{err}}_1\\
&\quad+2\Im\int -ig\overline{\boldsymbol{W}}wr_\alpha-\bar{R}r_\alpha r_\alpha+ig\bar{R}ww_\alpha+\bar{R}r_\alpha G+\bar{R}wK_\alpha\,d\alpha\\
&\quad+2\Re\int \bar{R}_\alpha w^2+2\overline{\boldsymbol{W}}wr_\alpha-2\overline{\boldsymbol{W}}wG\,d\alpha+2\underline{\text{err}}_2.\end{aligned}$$

Rewriting the above expression leads to

$$\begin{aligned}\frac{d}{dt}E^{(3)}_{\mathrm{lin}}(w,r)&=\frac{d}{dt}E^{(2)}_{\mathrm{lin}}(w,r)+2\frac{d}{dt}\left\{\left(\Im\int\bar{R}wr_\alpha\,d\alpha-\Re\int\overline{\boldsymbol{W}}w^2\,d\alpha\right)\right\}\\&=2\Re\int_{\mathbb{R}}(g+\underline{a})\bar{w}G-i\bar{r}_\alpha K\,d\alpha+c^2\Im\int_{\mathbb{R}}R|w|^2\,d\alpha+\underline{\mathrm{err}}_1\\&\quad+2\Re\int_{\mathbb{R}}(g+\underline{a})\bar{w}\{\boldsymbol{P}[R\bar{w}_\alpha]-\boldsymbol{P}[\boldsymbol{W}\bar{r}_\alpha]\}+i\bar{r}_\alpha\boldsymbol{P}[R\bar{r}_\alpha]\,d\alpha\\&\quad+2\Im\int-ig\overline{\boldsymbol{W}}wr_\alpha-\bar{R}r_\alpha r_\alpha+ig\bar{R}ww_\alpha+\bar{R}r_\alpha G+\bar{R}wK_\alpha\,d\alpha\\&\quad+2\Re\int\bar{R}_\alpha w^2+2\overline{\boldsymbol{W}}wr_\alpha-2\overline{\boldsymbol{W}}wG\,d\alpha+2\underline{\mathrm{err}}_2.\end{aligned}$$

We obtain

$$\begin{aligned}\frac{d}{dt}E^{(3)}_{\mathrm{lin}}(w,r)&=2\Re\int_{\mathbb{R}}[(g+\underline{a})\bar{w}-i\bar{R}_\alpha r_\alpha-2\overline{\boldsymbol{W}}w]G-i[\bar{r}_\alpha-(\bar{R}w)_\alpha]K\,d\alpha\\&\quad+c^2\Im\int_{\mathbb{R}}R|w|^2\,d\alpha+2\Re\int_{\mathbb{R}}\underline{a}\bar{w}\{\boldsymbol{P}[R\bar{w}_\alpha]-\boldsymbol{P}[\boldsymbol{W}\bar{r}_\alpha]\}\,d\alpha+\underline{\mathrm{err}}_1+2\underline{\mathrm{err}}_2.\end{aligned}$$

We introduce

$$\underline{\mathrm{err}}_3:=2\underline{\mathrm{err}}_2-2\Re\int_{\mathbb{R}}\underline{a}\bar{w}\{\boldsymbol{P}[\boldsymbol{W}\bar{r}_\alpha]-\boldsymbol{P}[R\bar{w}_\alpha]\}\,d\alpha,$$

which implies

$$\frac{d}{dt}E^{(3)}_{\mathrm{lin}}(w,r)=2\Re\int_{\mathbb{R}}[(g+\underline{a})\bar{w}-i\bar{R}_\alpha r_\alpha-2\overline{\boldsymbol{W}}w]G-i[\bar{r}-(\bar{R}w)]_\alpha K\,d\alpha+c^2\Im\int_{\mathbb{R}}R|w|^2\,d\alpha+\underline{\mathrm{err}}_1+\underline{\mathrm{err}}_3.$$

Given the bound (3-13) for $\underline{\mathrm{err}}_1$, the proof of (3-18) is concluded if we show that

$$|\underline{\mathrm{err}}_3|\lesssim\underline{A}\,\underline{B}\,E^{(2)}_{\mathrm{lin}}(w,r).\tag{3-22}$$

Due to the pointwise bound for $\underline{a}$, proved in Appendix A, and the Coifman–Meyer-type L^2-bound for $\boldsymbol{P}[\boldsymbol{W}\bar{r}_\alpha]-\boldsymbol{P}[R\bar{w}_\alpha]$ from [HIT16], it suffices to estimate $\underline{\mathrm{err}}_2$, which in turn reduces to estimating $\mathrm{err}_{2,1}$,

$$|\mathrm{err}_{2,1}|\lesssim\underline{A}\,\underline{B}\,E^{(2)}_{\mathrm{lin}}(w,r).$$

For the remainder of the proof we separately estimate several types of terms in $\mathrm{err}_{2,1}$:

Terms involving b_1. Here, we use the bounds for b_1 from Lemma A.2:

$$\|b_{1,\alpha}\|_{L^\infty}\lesssim A+A^2,\qquad \||D|^{1/2}b\|_{L^\infty}\lesssim A_{-1/2}+A_{-1/2}A.$$

We first collect all the terms that are contained in the first integral of $\mathrm{err}_{2,1}$, and which include b_1:

$$I_1:=\int_{\mathbb{R}}b_1\bar{R}_\alpha wr_\alpha+\bar{R}w\partial_\alpha\boldsymbol{P}[b_1r_\alpha]+\bar{R}r_\alpha\boldsymbol{P}[b_1w_\alpha]\,d\alpha.$$

After integrating by parts, we cancel two of the terms in I_1 and obtain

$$I_1=I_2+I_3,$$

where

$$I_2:=\int_{\mathbb{R}}\bar{R}_\alpha w\overline{\boldsymbol{P}}[b_1r_\alpha]\,d\alpha\quad\text{and}\quad I_3:=\int_{\mathbb{R}}\bar{R}r_\alpha\boldsymbol{P}[b_1w_\alpha]-\bar{R}w_\alpha\boldsymbol{P}[b_1r_\alpha]\,d\alpha.$$

The bounds for I_2 follow easily because it has a commutator structure:

$$I_2 := \int_{\mathbb{R}} \bar{R}_\alpha w \bar{\boldsymbol{P}}[b_1 r_\alpha]\, d\alpha = \int_{\mathbb{R}} \boldsymbol{P}[\bar{R}_\alpha w] \bar{\boldsymbol{P}}[b_1 r_\alpha]\, d\alpha = \int_{\mathbb{R}} [\boldsymbol{P}, w]\bar{R}_\alpha \cdot [\bar{\boldsymbol{P}}, b_1] r_\alpha \, d\alpha,$$

where we can estimate both factors in L^2 using Coifman–Meyer estimates,

$$\|[\boldsymbol{P}, w]\bar{R}_\alpha\|_{L^2} \lesssim \|w\|_{L^2}\, \|R_\alpha\|_{\mathrm{BMO}}, \qquad \|[\bar{\boldsymbol{P}}, b_1] r_\alpha\|_{L^2} \lesssim \||D|^{1/2} b_1\|_{\mathrm{BMO}}\, \|r\|_{\dot{H}^{1/2}}.$$

The bound for I_3 follows from Lemma 8.9 in Appendix B of [HIT16].

We next collect all the terms that are contained in the second integral appearing in the expression of $\mathrm{err}_{2,1}$, and which include b_1:

$$I_4 := \int_{\mathbb{R}} -b_1 \bar{\boldsymbol{W}}_\alpha w^2 - 2\bar{\boldsymbol{W}} w \boldsymbol{P}[b_1 w_\alpha]\, d\alpha.$$

As before, we integrate by parts and rewrite the expression for I_4 as

$$I_4 := \int_{\mathbb{R}} b_{1,\alpha} \bar{\boldsymbol{W}} w^2 + 2b_1 \bar{\boldsymbol{W}} w w_\alpha - 2\bar{\boldsymbol{W}} w \boldsymbol{P}[b_1 w_\alpha]\, d\alpha = \int_{\mathbb{R}} b_{1,\alpha} \bar{\boldsymbol{W}} w^2 + 2\bar{\boldsymbol{W}} w \bar{\boldsymbol{P}}[b_1 w_\alpha]\, d\alpha.$$

The first integral on the right-hand side is easy to bound since we know that $b_{1,\alpha}$ is in L^∞:

$$\left| \int_{\mathbb{R}} b_{1,\alpha} \bar{\boldsymbol{W}} w^2 \, d\alpha \right| \lesssim \|b_{1,\alpha}\|_{L^\infty}\, \|\boldsymbol{W}\|_{L^\infty}\, \|w\|^2_{L^2}.$$

For the last integral in I_4 we use the Coifman–Meyer-type estimate (established first in [HIT16]) to obtain

$$\left| \int_{\mathbb{R}} 2\bar{\boldsymbol{W}} w \bar{\boldsymbol{P}}[b_1 w_\alpha]\, d\alpha \right| \lesssim \|\boldsymbol{W}\|_{L^\infty}\, \|w\|_{L^2}\, \|[\bar{\boldsymbol{P}}, b_1] w_\alpha\|_{L^2} \lesssim \|b_{1,\alpha}\|_{L^\infty}\, \|\boldsymbol{W}\|_{L^\infty}\, \|w\|^2_{L^2}.$$

Thus,

$$|I_4| \lesssim \|b_{1,\alpha}\|_{L^\infty}\, \|\boldsymbol{W}\|_{L^\infty}\, \|w\|^2_{L^2}.$$

Quadrilinear terms bounded via $L^2 \cdot L^2$ pairing. Some of the remaining terms in $\mathrm{err}_{2,1}$ have straightforward bounds:

$$|I_7| := \left| \int_{\mathbb{R}} (1+\bar{\boldsymbol{W}})\bar{M}_1 w^2 \, d\alpha \right| \lesssim_A (1+A)A^2 \|w\|^2_{L^2},$$

$$|I_8| := \left| \int_{\mathbb{R}} \bar{\boldsymbol{W}}(\bar{\boldsymbol{W}} - \boldsymbol{W}) w^2 \, d\alpha \right| \lesssim_A A^2 \|w\|^2_{L^2},$$

but others require a little bit of work. This includes the following expressions:

$$I_5 := \int_{\mathbb{R}} \frac{\bar{R}\bar{\boldsymbol{W}} + R\bar{\boldsymbol{W}} + \bar{N}}{1+\bar{\boldsymbol{W}}} w r_\alpha \, d\alpha,$$

$$I_6 := \int_{\mathbb{R}} \bar{R} w \partial_\alpha \boldsymbol{P}\left[\frac{a_1}{1+\boldsymbol{W}} w\right] d\alpha = -\int_{\mathbb{R}} \partial_\alpha \bar{\boldsymbol{P}}[\bar{R}w] \boldsymbol{P}[a_1(1-Y)w]\, d\alpha.$$

To obtain the bound for I_6 we use the Cauchy–Schwarz inequality and Lemma 2.1 from [HIT16]

$$|I_6| \lesssim \|\partial_\alpha \bar{\boldsymbol{P}}[\bar{R}w]\|_{L^2}\, \|\boldsymbol{P}[a_1(1-Y)w]\|_{L^2} \lesssim \|R_\alpha\|_{\mathrm{BMO}}\, \|a_1\|_{L^\infty}\, (1+\|Y\|_{L^\infty})\, \|w\|^2_{L^2}.$$

Finally, the bound for I_5 is also a consequence of commutator estimates; to see this we rewrite it as

$$I_5 := \int_{\mathbb{R}} f w r_\alpha \, d\alpha = \int_{\mathbb{R}} \bar{\boldsymbol{P}}[f r_\alpha] w \, d\alpha, \quad \text{where } f := \frac{\bar{R}\bar{\boldsymbol{W}} + R\bar{\boldsymbol{W}} + \bar{N}}{1+\bar{\boldsymbol{W}}}.$$

Thus,

$$|I_5| = \left| \int_{\mathbb{R}} \bar{\boldsymbol{P}}[f r_\alpha] w \, d\alpha \right| \lesssim \|[\bar{\boldsymbol{P}}, f] r_\alpha\|_{L^2} \|w\|_{L^2} \lesssim \||D|^{1/2} f\|_{\mathrm{BMO}} \|r\|_{\dot{H}^{1/2}} \|w\|_{L^2},$$

where

$$\||D|^{1/2} f\|_{\mathrm{BMO}} \lesssim \underline{A}\,\underline{B}. \qquad \square$$

Proof of Proposition 3.5. To prove the bound in (3-19) it suffices to apply the estimate in (3-18) with

$$G = \boldsymbol{P}\underline{\mathcal{G}}, \quad K = \boldsymbol{P}\underline{\mathcal{K}}.$$

In fact, we only have to resume our work in finding the new terms introduced by the vorticity assumption, which are the components carrying the c when expanding the right-hand side of (3-5), $\boldsymbol{P}\mathcal{G}_1$ and $\boldsymbol{P}\mathcal{K}_1$:

$$\underline{\mathcal{G}} = \mathcal{G} - i\frac{c}{2}\mathcal{G}_1, \quad \underline{\mathcal{K}} = \mathcal{K} - i\frac{c}{2}\mathcal{K}_1.$$

The terms we want to single out are the c-cubic terms appearing in the cubic part of the energy $E^3(w, r)$, (3-8).

For this we need to recall the quadratic components of $\boldsymbol{P}\mathcal{G}_1$ and $\boldsymbol{P}\mathcal{K}_1$

$$\begin{aligned}
\boldsymbol{P}\mathcal{G}_1^{(2)}(w,r) &= \boldsymbol{P}[\boldsymbol{W}\bar{w}] + \boldsymbol{P}[W\bar{w}_\alpha] + \boldsymbol{P}[\bar{\boldsymbol{W}}w] - \boldsymbol{P}[\boldsymbol{W}w],\\
\boldsymbol{P}\mathcal{K}_1^{(2)}(w,r) &= \boldsymbol{P}[W\bar{r}_\alpha] - \boldsymbol{P}[R\bar{w}],
\end{aligned}$$

and the antiholomorphic components (which will not matter in this computation since we work with the projected equations)

$$\begin{aligned}
\bar{\boldsymbol{P}}\mathcal{G}_1^{(2)}(w,r) &= -\bar{\boldsymbol{P}}[\bar{\boldsymbol{W}}w] - \bar{\boldsymbol{P}}[\bar{W}w_\alpha] + \bar{\boldsymbol{P}}[\bar{\boldsymbol{W}}w],\\
\bar{\boldsymbol{P}}\mathcal{K}_1^{(2)}(w,r) &= \bar{\boldsymbol{P}}[\bar{W}r_\alpha] - \bar{\boldsymbol{P}}[\bar{R}w].
\end{aligned}$$

Thus, we need to bound the terms

$$\int_{\mathbb{R}} g w \boldsymbol{P}\mathcal{G}_1^{(2)}(w,r)\, d\alpha, \quad \int_{\mathbb{R}} \bar{r}_\alpha \boldsymbol{P}\mathcal{K}_1^{(2)}(w,r)\, d\alpha, \quad \int_{\mathbb{R}} c^2 \Im R|w|^2\, d\alpha.$$

These bounds are all obtained using Lemma 2.1 in [HIT16], for example, for the first integral we obtain

$$\left| \int_{\mathbb{R}} g w \{[\boldsymbol{P}, \boldsymbol{W}]\bar{w} + [\boldsymbol{P}, W]\bar{w}_\alpha - [\boldsymbol{P}, \boldsymbol{W}]w\}\, d\alpha \right| \lesssim A\|w\|_{L^2}^2.$$

For the second integral we need to move half of a derivative off of $\bar{r}_\alpha$:

$$\left| \int_{\mathbb{R}} \bar{r}_\alpha \{[\boldsymbol{P}, W]\bar{r}_\alpha - [\boldsymbol{P}, R]\bar{w}\}\, d\alpha \right| \lesssim B\|r\|_{\dot{H}^{1/2}}^2 + A\|r\|_{\dot{H}^{1/2}}\|w\|_{L^2}.$$

The last one, is trivial

$$\left| \int_{\mathbb{R}} R|w|^2\, d\alpha \right| \lesssim A_{-1/2}\|w\|_{L^2}^2. \qquad \square$$

4. Higher-order energy estimates

The main goal of this section is to prove energy bounds for the differentiated equations. These are the main ingredient for the lifespan part of Theorem 1, as well as for the cubic result in Theorem 2. Precisely, we will establish two types of energy bounds for $(\boldsymbol{W}, R)$ and their higher derivatives. The first one is a quadratic bound which applies independently of the size of the initial data; this yields the last part of Theorem 1. The cubic energy bound applies in the small-data case, and yields the cubic lifespan bound in our small-data result in Theorem 2. The large-data result is as follows:

Proposition 4.1. *For any $n \geq 0$ there exists an energy functional $E^{n,(2)}$ with the following properties whenever the conditions (1-25) and (1-26) hold uniformly*:

(i) *Norm equivalence*:

$$E^{n,(2)}(\boldsymbol{W}, R) \approx_{\underline{A}} \mathcal{E}_0(\partial^n \boldsymbol{W}, \partial^n R).$$

(ii) *Quadratic energy estimates for solutions to (1-19)*:

$$\frac{d}{dt} E^{n,(2)}(\boldsymbol{W}, R) \lesssim_{\underline{A}} \underline{B}\mathcal{E}_0(\partial^n \boldsymbol{W}, \partial^n R). \tag{4-1}$$

The small-data result is as follows:

Proposition 4.2. *For any $n \geq 0$ there exists an energy functional $E^{n,(3)}$ which has the following properties as long as $\underline{A} \ll 1$*:

(i) *Norm equivalence*:

$$E^{n,(3)}(\boldsymbol{W}, R) = (1 + O(\underline{A}))\mathcal{E}_0(\partial^n \boldsymbol{W}, \partial^n R) + O(c^4 \underline{A})\mathcal{E}_0(\partial^{n-1} \boldsymbol{W}, \partial^{n-1} R).$$

(ii) *Cubic energy estimates*:

$$\frac{d}{dt} E^{n,(3)}(\boldsymbol{W}, R) \lesssim_{\underline{A}} \underline{B}\,\underline{A}\big(\mathcal{E}_0(\partial^n \boldsymbol{W}, \partial^n R) + c^4 \mathcal{E}_0(\partial^{n-1} \boldsymbol{W}, \partial^{n-1} R)\big). \tag{4-2}$$

Here if $n = 0$ then $\mathcal{E}_0(\partial^{-1} \boldsymbol{W}, \partial^{-1} R)$ is naturally replaced by $\mathcal{E}(W, Q)$.

The case $n = 0$ of Proposition 4.1 corresponds to $\dot{\mathcal{H}}_0$-bounds for $(\boldsymbol{W}_\alpha, R)$. But these functions solve the linearized equation, so the desired bounds are a consequence of Proposition 3.1. Hence, we will begin with the proof of the cubic bounds for the $n = 0$ case.

Next, we compute the differentiated equations, first for $n = 1$ and then for $n \geq 2$, and show that, up to a certain class of bounded errors, suitable modifications of $(\boldsymbol{W}^{(n)}, R^{(n)})$ solve a linear system which is quite similar to the linearized equation. There are two reasons why we separate the case $n = 1$. First, this corresponds to $\dot{\mathcal{H}}_0$-bounds for $(\boldsymbol{W}_\alpha, R_\alpha)$, which play a special role as it is our threshold for local well-posedness. Secondly, there is a subtle difference in the choice of the modifications alluded to above, due to the fact that certain terms which are different for $n \geq 2$ coincide at $n = 1$.

At this point, the bounds for the linearized equation already yield the large-data result. What remains is to establish the cubic small-data result, which is where we implement our quasilinear modified energy method, using the normal form calculated in Section 2.

As part of the argument we need to express various nonlinear expressions in terms of their homogeneous expansions. To describe the decomposition of a nonlinear analytic expression F into homogeneous components we will use the notation $\Lambda^k F$ to denote the component of F of homogeneity k. We similarly introduce the operators $\Lambda^{\leq k}$ and $\Lambda^{\geq k}$. We carefully note that all of our multilinear expansions are with respect to the diagonal variables W and R, and not with respect to (W, Q).

Compared to [HIT16] here we lack scaling, but we can still introduce a notion of order of a multilinear expression. We begin with single terms, for which we assign orders as follows:

- The order of $W^{(k)}$ is $k-1$.
- The order of $R^{(k)}$ is $k-\frac{1}{2}$.
- The order of c is $\frac{1}{2}$.

For a multilinear form involving products of such terms we define the total order as the sum of the orders of all factors.

While not all expressions arising in the $(\mathbf{W}^{(n)}, R^{(n)})$ are multilinear in (W, R), they can be still viewed as multilinear in (W, R) and undifferentiated Y. Since Y scales like $\mathbf{W}$, it is natural to assign to it the homogeneity zero. According to this definition, all terms in the $\mathbf{W}^{(n)}$-equation have order $n+\frac{1}{2}$, and all terms in the R-equation have order n. Moving on to integral multilinear forms, all n-th energies have order $2n$, and their time derivatives have order $2n+\frac{1}{2}$.

A second useful bookkeeping device will be needed when we deal with integral multilinear forms. There it makes a difference how derivatives and also complex conjugations are distributed among factors. To account for this we define the leading order of a multilinear form to be the largest sum of the orders of two factors with opposite conjugations. Since we only allow nonnegative orders, for the n-th order energy this is at most $2n$. According to our definition, all the terms in our n-th order energy will have order $2n$, and all terms in its time derivative will have order $2n+\frac{1}{2}$. We remark that the half-integers in the definition of the orders impose a parity constraint in the terms associated to each power of c.

If all factors in a multilinear form have nonnegative orders, this imposes a constraint on the order of each factor. Unfortunately this does not appear to be the case here, as our multilinear forms will also contain factors of W and R, which have negative order. This is quite inconvenient. Fortunately, there is a simple way to avoid negative orders altogether. Precisely, we will never consider such factors alone, but in combination with c; thus, the allowed factors will be cR and c^2W, both of which have order 0. We carefully note last remark applies only partially in the special case $n=0$.

4A. ***The case $n=0$.*** The goal of this subsection is to obtain cubic energy estimates in $\dot{\mathcal{H}}_0$ for the system for diagonal variables $(\mathbf{W}, R)$. For convenience we recall the system here:

$$\begin{cases} \mathbf{W}_t + \underline{b}\mathbf{W}_\alpha + \dfrac{(1+\mathbf{W})R_\alpha}{1+\overline{\mathbf{W}}} = \underline{\mathcal{G}}_0, \\ R_t + \underline{b}R_\alpha + icR - i\dfrac{g\mathbf{W}-a}{1+\mathbf{W}} = \underline{\mathcal{K}}_0, \end{cases}$$

where

$$\underline{\mathcal{G}}_0 = (1+\mathbf{W})\underline{M} + i\frac{c}{2}\mathbf{W}(\mathbf{W}-\overline{\mathbf{W}}), \quad \underline{\mathcal{K}}_0 = i\frac{c}{2}\frac{R\mathbf{W}+\bar{R}\mathbf{W}+N}{1+\mathbf{W}}.$$

We want to be able to apply the quadratic and cubic bounds for the "modified" model (3-16), respectively. For that, we rewrite the above systems as

$$\begin{cases} \boldsymbol{W}_t + \underline{b}\boldsymbol{W}_\alpha + \dfrac{R_\alpha}{1+\overline{\boldsymbol{W}}} + \dfrac{R_\alpha}{1+\overline{\boldsymbol{W}}}\boldsymbol{W} = -\boldsymbol{P}[\bar{R}_\alpha \boldsymbol{W}] + \boldsymbol{P}[R\overline{\boldsymbol{W}}_\alpha] + \underline{G}_0, \\ R_t + \underline{b}R_\alpha + icR - i\dfrac{(g+\underline{a})\boldsymbol{W}}{1+\boldsymbol{W}} = -\boldsymbol{P}[R\bar{R}_\alpha] + \underline{K}_0, \end{cases} \tag{4-3}$$

where, the expressions for $\underline{G}_0$ and $\underline{K}_0$, for the purpose of this section, are

$$\begin{cases} \underline{G}_0 := \boldsymbol{W}(\boldsymbol{P}[R\overline{\boldsymbol{W}}_\alpha] - \boldsymbol{P}[\bar{R}_\alpha \boldsymbol{W}]) + i\dfrac{c}{2}(1+\boldsymbol{W})M_1 + i\dfrac{c}{2}\boldsymbol{W}(\boldsymbol{W} - \overline{\boldsymbol{W}}) \\ \qquad\quad + (1+\boldsymbol{W})\{\bar{\boldsymbol{P}}[\bar{R}Y_\alpha - R_\alpha \bar{Y}] - \boldsymbol{P}[R\partial_\alpha(\overline{\boldsymbol{W}}\bar{Y})] + \boldsymbol{P}[\bar{R}_\alpha \boldsymbol{W} Y]\}, \\ \underline{K}_0 := \bar{\boldsymbol{P}}[\bar{R}R_\alpha] - i\dfrac{c}{2}N. \end{cases}$$

In (4-3) we have identified the leading part of the equation. We want to interpret the terms $(\underline{G}_0, \underline{K}_0)$ on the right as mostly perturbative, but also pay attention to the holomorphic quadratic part, given by

$$\boldsymbol{P}G_0^{(2)} = i\frac{c}{2}(\boldsymbol{P}[W\bar{Y}]_\alpha + \boldsymbol{W}^2 - \boldsymbol{P}[\boldsymbol{W}\overline{\boldsymbol{W}}]), \quad \boldsymbol{P}K_0^{(2)} = i\frac{c}{2}\boldsymbol{P}[W\bar{R}_\alpha - \overline{\boldsymbol{W}}R]. \tag{4-4}$$

Precisely, we first claim that the quadratic and cubic parts of $(\underline{G}_0, \underline{K}_0)$ satisfy the bounds

$$\|(\underline{G}_0^{(2)}, \underline{K}_0^{(2)})\|_{\dot{\mathcal{H}}_0} \lesssim_A \underline{B}\boldsymbol{N}_0, \tag{4-5}$$

$$\|(\underline{G}_0^{(3)}, \underline{K}_0^{(3)})\|_{\dot{\mathcal{H}}_0} \lesssim_A A\underline{B}\boldsymbol{N}_0, \tag{4-6}$$

respectively, where

$$\boldsymbol{N}_0 = \|(\boldsymbol{W}, R)\|_{L^2\times \dot{H}^{1/2}}.$$

The bounds for the components of M_1 and N are discussed in Lemma A.2. The $\boldsymbol{W}$ prefactors in $\underline{G}_0$ are harmless, as they are bounded by A in L^∞. For the remaining terms it suffices to use Coifman–Meyer-type estimates discussed in Appendix A. For instance we have

$$\|\boldsymbol{P}[R\overline{\boldsymbol{W}}_\alpha]\|_{L^2} \lesssim \|[\boldsymbol{P}, R]\overline{\boldsymbol{W}}_\alpha\|_{L^2} \lesssim \||D|^{1/2}\boldsymbol{W}\|_{\mathrm{BMO}}\|R\|_{\dot{H}^{1/2}} \lesssim_A B\boldsymbol{N}_0,$$

and all other terms in $\underline{G}_0$ are similar. Finally, for the first term in $\underline{K}_0$ we have

$$\|\boldsymbol{P}[\bar{R}R_\alpha]\|_{\dot{H}^{1/2}} \lesssim \||D|^{1/2}R\|_{L^2}\|R_\alpha\|_{\mathrm{BMO}} \lesssim B\boldsymbol{N}_0.$$

We are now ready to look at the cubic energies. We start by constructing the cubic normal-form energy by selecting the quadratic and cubic terms from the corresponding linear energy for the normal-form variables. Precisely, we have

$$\mathcal{E}_0(\widetilde{W}_\alpha, \widetilde{Q}_\alpha) = \mathcal{E}_0(\boldsymbol{W}, Q_\alpha) + 2\int_{\mathbb{R}} \Re\overline{\boldsymbol{W}}\partial_\alpha W^{[2]} - 2\Im\bar{Q}_{\alpha\alpha}\partial_\alpha Q^{[2]}\, d\alpha + \text{quartic}.$$

In the first term we substitute $Q_\alpha = R(1+\boldsymbol{W})$. In the integral we use the expressions (2-8), integrate by parts to eliminate the $\partial^{-1}W$- and Q-factors, and then replace Q_α with R. Separating the outcome of these computations into a leading part and a lower-order part, we write it in the form

$$E^{(3)}_{\mathrm{NF}} := E^{(3)}_{\mathrm{NF,high}} + E^{(3)}_{\mathrm{NF,low}},$$

where

$$E^{3}_{\mathrm{NF,high}}(\boldsymbol{W}, R) := \int_{\mathbb{R}} (g + c\Re R)|\boldsymbol{W}|^2 + \Im(R\bar{R}_\alpha) + 2\Im(\bar{R}\boldsymbol{W}R_\alpha) - 2\Re(\overline{\boldsymbol{W}}\boldsymbol{W}^2)\, d\alpha,$$

and

$$\begin{aligned} E^{(3)}_{\mathrm{NF,low}}(\boldsymbol{W}, R) := & -c\int_{\mathbb{R}} 2\Re R\{|\boldsymbol{W}|^2 - \Im(\bar{R}_\alpha R)\} - \boldsymbol{W}\overline{W}R_\alpha + \boldsymbol{W}^2\bar{R}\, d\alpha \\ & - \frac{c^2}{g}\int_{\mathbb{R}} \tfrac{5}{2}\Im W\{|\boldsymbol{W}|^2 - \Im(\bar{R}_\alpha R)\} - \tfrac{1}{2}\overline{W}\boldsymbol{W}^2 - \tfrac{1}{2}\boldsymbol{W}\bar{R}^2\, d\alpha \\ & - \frac{3c^3}{2g}\Im\int_{\mathbb{R}} RW\overline{\boldsymbol{W}}\, d\alpha - \frac{3c^4}{2g}\Re\int_{\mathbb{R}} \boldsymbol{W}|W|^2\, d\alpha. \end{aligned}$$

For the leading-order part $E^{(3)}_{\mathrm{NF,high}}(\boldsymbol{W}, R)$ we consider the appropriate quasilinear correction

$$E^{(3)}_{\mathrm{high}}(\boldsymbol{W}, R) := \int_{\mathbb{R}} (g + \underline{a})|\boldsymbol{W}|^2 + \Im(R\bar{R}_\alpha) + 2\Im(\bar{R}\boldsymbol{W}R_\alpha) - 2\Re(\overline{\boldsymbol{W}}\boldsymbol{W}^2)\, d\alpha,$$

and the remainder $E^{3}_{\mathrm{NF,low}}(\boldsymbol{W}, R)$ remains unchanged. Hence, we define the quasilinear cubic energy

$$E^{0,(3)} := E^{3}_{\mathrm{high}} + E^{3}_{\mathrm{NF,low}}.$$

It remains to show that this energy has all the right properties in Proposition 4.2.

We begin with the energy equivalence. For the leading part this has already been done in the context of the linearized equation (see Proposition 3.4),

$$E^{(3)}_{\mathrm{high}}(\boldsymbol{W}, R) \approx (1 + O(A))E^{(2)}_{\mathrm{lin}}(\boldsymbol{W}, R).$$

So it remains to show that

$$E^{(3)}_{\mathrm{NF,low}}(\boldsymbol{W}, R) \approx O(\underline{A})(\mathcal{E}_0(\boldsymbol{W}, R) + c^4\mathcal{E}(W, Q)).$$

The bound is straightforward for all terms not containing R_α. So we now consider those. For the first one we have

$$\left|\int_{\mathbb{R}} R_\alpha |R|^2\, d\alpha\right| \lesssim \|R\|_{\dot{H}^{1/2}} \||D|^{1/2}(R\Re R)\|_{L^2} \lesssim \|R\|^2_{\dot{H}^{1/2}} \|R\|_{L^\infty} \lesssim A_{-1/2}\boldsymbol{N}_0^2,$$

which suffices. For the next term we have

$$\left|\int_{\mathbb{R}} R_\alpha \overline{W}\boldsymbol{W}\, d\alpha\right| \lesssim \|R\|_{\dot{H}^{1/2}} \||D|^{1/2}\boldsymbol{P}[W\overline{\boldsymbol{W}}]\|_{L^2} \lesssim \|R\|_{\dot{H}^{1/2}} \||D|^{1/2}W\|_{\mathrm{BMO}} \|\boldsymbol{W}\|_{L^2} \lesssim A_{-1/2}\boldsymbol{N}_0^2.$$

Finally,

$$\|R_\alpha \bar{R}W\|_{L^2} \lesssim \|R\|_{\dot{H}^{1/2}} \||D|^{1/2}(\bar{R}W)\|_{L^2} \lesssim (\||D|^{1/2}R\|_{\mathrm{BMO}} \|W\|_{L^2} + \|R\|_{L^\infty} \||D|^{1/2}W\|_{L^2}) \|R\|_{\dot{H}^{1/2}}.$$

Now we consider the time derivative of modified quasilinear energy $E^{0,(3)}$ in order to prove the bound (4-2). By construction we know that

$$\Lambda^{\leq 3} \frac{d}{dt} E^{0,(3)} = 0.$$

We note that this is an algebraic property which follows from the normal-form-based construction even though the normal form itself is unbounded. Therefore it remains to estimate

$$\Lambda^{\geq 4} \frac{d}{dt} E^3 = \Lambda^{\geq 4} \frac{d}{dt} E^3_{\mathrm{NF,high}} + \Lambda^{\geq 4} \frac{d}{dt} E^3_{\mathrm{NF,low}}.$$

Due to (4-5) and (4-6) the estimate for the first term on the right-hand side follows directly from the bound (3-18) in Proposition 3.4 for the leading part of the linearized equation. Hence, it remains to consider the last term.

$E^3_{\mathrm{NF,low}}(\boldsymbol{W}, R)$ is a trilinear form of order zero and leading order zero. We compute its time derivative using the relation

$$\frac{d}{dt}\int f_1 f_2 f_3 \, d\alpha = \int (\partial_t + \underline{b}\partial_\alpha) f_1 f_2 f_3 + f_1 (\partial_t + \underline{b}\partial_\alpha) f_2 f_3 + f_1 f_2 (\partial_t + \underline{b}\partial_\alpha) f_3 - \underline{b}_\alpha f_1 f_2 f_3 \, d\alpha. \quad (4\text{-}7)$$

Then its time derivative will be a multilinear form of order $\frac{1}{2}$, and also of leading order $\frac{1}{2}$. By inspection, we see that in this time derivative we can associate each W with a c^2-factor and each R with a c-factor, so that all each of the factors in all of the multilinear monomials have degree at least zero. Then, each multilinear monomial in $\Lambda^{\geq 4}\big(\frac{d}{dt} E^3_{\mathrm{NF,low}}\big)$ contains exactly one factor of order $\frac{1}{2}$, and the rest are all factors of order zero. The factor of order $\frac{1}{2}$ can be either R_α or c, and the factors of order zero could be $\boldsymbol{W}$, Y, $c^2 W$, or cR. We have two cases to consider:

(a) If the factor of order $\frac{1}{2}$ is c, then we simply bound two of the remaining factors in L^2 and the others in L^∞.

(b) If the factor of order $\frac{1}{2}$ is R_α, then we use the $\dot{H}^{1/2}$-bound for R, and we are left with estimating an expression of the form $|D|^{1/2}(f_1 \cdots f_k)$, where $k \geq 3$, and each f_k has order zero. For this we use Lemma A.1 to estimate

$$\| |D|^{1/2}(f_1 \cdots f_k) \|_{L^2} \lesssim \sum_j \| |D|^{1/2} f_j \|_{L^\infty} \Big\| \prod_{l \neq j} f_l \Big\|_{L^2},$$

and conclude as above.

4B. ***The differentiated equations for n = 1.*** We begin by differentiating (1-19) in order to obtain a system for $(\boldsymbol{W}_\alpha, R_\alpha)$:

$$\begin{cases} \boldsymbol{W}_{\alpha t} + \underline{b}\boldsymbol{W}_{\alpha\alpha} + \dfrac{[(1+\boldsymbol{W})R_\alpha]_\alpha}{1+\bar{\boldsymbol{W}}} = (\underline{M} - \underline{b}_\alpha)\boldsymbol{W}_\alpha + (1+\boldsymbol{W})(R_\alpha \bar{Y}_\alpha + \underline{M}_\alpha) + i\dfrac{c}{2}[\boldsymbol{W}(\boldsymbol{W} - \bar{\boldsymbol{W}})]_\alpha, \\ R_{\alpha t} + \underline{b}R_{\alpha\alpha} + icR_\alpha - i\left(\dfrac{(g+a)\boldsymbol{W}_\alpha}{(1+\boldsymbol{W})^2} - \dfrac{a_\alpha}{1+\boldsymbol{W}}\right) = -\underline{b}_\alpha R_\alpha + i\dfrac{c}{2}\left(\dfrac{\boldsymbol{W}(R+\bar{R}) + N}{1+\boldsymbol{W}}\right)_\alpha. \end{cases}$$

We can expand the last term in the second equation, putting together all terms which involve $\boldsymbol{W}_\alpha$. The reason for this is that the $\boldsymbol{W}_\alpha$-terms will be unbounded in a suitable sense, and need to be treated as part of the leading-order linear operator. This will lead us to the coefficient a_1, which will be moved to the left, completing the coupling coefficient a to $\underline{a}$. We have

$$\left(\frac{\boldsymbol{W}(R+\bar{R})+N}{1+\boldsymbol{W}}\right)_\alpha = \frac{\boldsymbol{W}_\alpha}{(1+\boldsymbol{W})^2}\{(R+\bar{R})-N\}+\frac{\boldsymbol{W}(R_\alpha+\bar{R}_\alpha)+N_\alpha}{1+\boldsymbol{W}}$$
$$=\frac{a_1\boldsymbol{W}_\alpha}{(1+\boldsymbol{W})^2}+\frac{\boldsymbol{W}(R_\alpha+\bar{R}_\alpha)+N_\alpha}{1+\boldsymbol{W}}.$$

Then our system becomes

$$\begin{cases} \boldsymbol{W}_{\alpha t}+\underline{b}\boldsymbol{W}_{\alpha\alpha}+\dfrac{[(1+\boldsymbol{W})R_\alpha]_\alpha}{1+\overline{\boldsymbol{W}}}=(\underline{M}-\underline{b}_\alpha)\boldsymbol{W}_\alpha+(1+\boldsymbol{W})(R_\alpha\bar{Y}_\alpha+\underline{M}_\alpha)+i\dfrac{c}{2}[\boldsymbol{W}(\boldsymbol{W}-\overline{\boldsymbol{W}})]_\alpha, \\ R_{t\alpha}+\underline{b}R_{\alpha\alpha}+icR_\alpha-i\dfrac{(g+\underline{a})\boldsymbol{W}_\alpha}{(1+\boldsymbol{W})^2}=-\underline{b}_\alpha R_\alpha-i\dfrac{a_\alpha}{1+\boldsymbol{W}}+i\dfrac{c}{2}\dfrac{\boldsymbol{W}(R_\alpha+\bar{R}_\alpha)+N_\alpha}{1+\boldsymbol{W}}. \end{cases}$$

In order to better compare this with the linearized system we introduce the modified variable $\boldsymbol{R}:=R_\alpha(1+\boldsymbol{W})$ to further obtain

$$\begin{cases} \boldsymbol{W}_{\alpha t}+\underline{b}\boldsymbol{W}_{\alpha\alpha}+\dfrac{\boldsymbol{R}_\alpha}{1+\overline{\boldsymbol{W}}}=(\underline{M}-\underline{b}_\alpha)\boldsymbol{W}_\alpha+\boldsymbol{R}\bar{Y}_\alpha+(1+\boldsymbol{W})\underline{M}_\alpha+i\dfrac{c}{2}[\boldsymbol{W}(\boldsymbol{W}-\overline{\boldsymbol{W}})]_\alpha, \\ \boldsymbol{R}_t+\underline{b}\boldsymbol{R}_\alpha+ic\boldsymbol{R}-i\dfrac{(g+\underline{a})\boldsymbol{W}_\alpha}{1+\boldsymbol{W}}=\left(\underline{M}-\underline{b}_\alpha-\dfrac{R_\alpha}{1+\overline{\boldsymbol{W}}}+i\dfrac{c}{2}(\boldsymbol{W}-\overline{\boldsymbol{W}})\right)\boldsymbol{R}-ia_\alpha+i\dfrac{c}{2}[\overline{\boldsymbol{W}}R_\alpha+\boldsymbol{W}\bar{R}_\alpha+N_\alpha]. \end{cases}$$

Expanding the $\underline{b}_\alpha$-terms via (1-18) this yields

$$\begin{cases} \boldsymbol{W}_{\alpha t}+\underline{b}\boldsymbol{W}_{\alpha\alpha}+\dfrac{\boldsymbol{R}_\alpha}{1+\overline{\boldsymbol{W}}}+\dfrac{R_\alpha}{1+\overline{\boldsymbol{W}}}\boldsymbol{W}_\alpha=\underline{G}_1, \\ \boldsymbol{R}_t+\underline{b}\boldsymbol{R}_\alpha+ic\boldsymbol{R}-i\dfrac{(g+a)\boldsymbol{W}_\alpha}{1+\boldsymbol{W}}=\underline{K}_1, \end{cases} \tag{4-8}$$

where on the right we have placed all terms which should be thought of as "bounded". Precisely, $(\underline{G}_1, \underline{K}_1)$ have the form

$$\underline{G}_1:=G_1-i\frac{c}{2}G_{1,1},\qquad \underline{K}_1:=K_1-i\frac{c}{2}K_{1,1},$$

where the leading components have the same form as in [HIT16],

$$\begin{cases} G_1=\boldsymbol{R}\bar{Y}_\alpha-\dfrac{\bar{R}_\alpha}{1+\boldsymbol{W}}\boldsymbol{W}_\alpha+2M\boldsymbol{W}_\alpha+(1+\boldsymbol{W})M_\alpha, \\ K_1=-2\left(\dfrac{\bar{R}_\alpha}{1+\boldsymbol{W}}+\dfrac{R_\alpha}{1+\overline{\boldsymbol{W}}}\right)\boldsymbol{R}+2M\boldsymbol{R}+(R_\alpha\bar{R}_\alpha-ia_\alpha), \end{cases}$$

while the extra terms containing the vorticity c are

$$\begin{cases} G_{1,1}=2M_1\boldsymbol{W}_\alpha+(1+\boldsymbol{W})M_{1,\alpha}-2\boldsymbol{W}_\alpha(\boldsymbol{W}-\overline{\boldsymbol{W}})-W(\boldsymbol{W}_\alpha-\overline{\boldsymbol{W}}_\alpha), \\ K_{1,1}=2M_1\boldsymbol{R}-2\boldsymbol{R}(\boldsymbol{W}-\overline{\boldsymbol{W}})-\boldsymbol{P}[2\boldsymbol{W}\bar{R}_\alpha+W\bar{R}_{\alpha\alpha}-\overline{\boldsymbol{W}}_\alpha R]-\overline{\boldsymbol{P}}[2\overline{\boldsymbol{W}}R_\alpha+\overline{W}R_{\alpha\alpha}-\boldsymbol{W}_\alpha\bar{R}]. \end{cases}$$

Here on the left we have again the leading part of the linearized equation. Following the same approach as in [HIT16], we will interpret the terms on the right as mostly perturbative, but also keep track of their quadratic part. Thus, for bookkeeping purposes, we introduce two types of error terms, denoted $\mathrm{err}(L^2)$

and $\text{err}(\dot{H}^{1/2})$, which correspond to the two equations. The bounds for these errors are in terms of the control variables $\underline{A}$, $\underline{B}$, as well as the L^2-type norm

$$\boldsymbol{N}_1 = \|(\boldsymbol{W}_\alpha, R_\alpha)\|_{L^2\times \dot{H}^{1/2}}.$$

In [HIT16], by $\text{err}(L^2)$ we denote terms G_1 which satisfy the estimates

$$\|PG_1\|_{L^2} \lesssim_A AB\boldsymbol{N}_1,$$

as well as either one of the following two:

$$\|\bar{P}G_1\|_{L^2} \lesssim_A B\boldsymbol{N}_1 \quad \text{or} \quad \|\bar{P}G_1\|_{\dot{H}^{-1/2}} \lesssim_A A\boldsymbol{N}_1.$$

Here, in order to manage the new c-terms, we also include in $\text{err}(L^2)$ the terms in $G_{1,1}$ for which

$$\|PG_{1,1}\|_{L^2} \lesssim_A A^2\boldsymbol{N}_1,$$

and either of the following two holds:

$$\|\bar{P}G_{1,1}\|_{L^2} \lesssim_A A\boldsymbol{N}_1 \quad \text{or} \quad \|\bar{P}G_{1,1}\|_{\dot{H}^{-1/2}} \lesssim_A A_{-1/2}\boldsymbol{N}_1.$$

Similarly, by $\text{err}(\dot{H}^{1/2})$ we have denoted in [HIT16] the terms K_1 which satisfy the estimates

$$\|PK_1\|_{\dot{H}^{1/2}} \lesssim_A AB\boldsymbol{N}_1, \quad \|PK_1\|_{L^2} \lesssim_A A^2\boldsymbol{N}_1,$$

and

$$\|\bar{P}K_1\|_{L^2} \lesssim_A A\boldsymbol{N}_1.$$

To that, for the c-terms we add expressions $K_{1,1}$ which satisfy the estimates

$$\|PK_{1,1}\|_{\dot{H}^{1/2}} \lesssim_A A^2\boldsymbol{N}_1, \quad \|PK_{2,1}\|_{L^2} \lesssim_A AA_{-1/2}\boldsymbol{N}_1,$$

and

$$\text{either} \quad \|\bar{P}K_{1,1}\|_{L^2} \lesssim_A A_{-1/2}\boldsymbol{N}_1 \quad \text{or} \quad \|\bar{P}K_{1,1}\|_{\dot{H}^{1/2}} \lesssim_A A\boldsymbol{N}_1.$$

We will rely on [HIT16] for the analysis of the expressions G_1 and K_1, and handle just the new entries, i.e., the new terms accompanied by the vorticity factor c.

We recall that the use of the more relaxed quadratic control on the antiholomorphic terms, as opposed to the cubic control on the holomorphic terms, is motivated by the fact that the equations will eventually get projected on the holomorphic space, so the antiholomorphic components will have less of an impact. A key property of the space of errors is contained in Lemma 3.3 in [HIT16], which we will refer to as *the multiplicative bounds lemma*. For convenience we recall it here:

Lemma 4.3. *Let Φ be a function which satisfies*

$$\|\Phi\|_{L^\infty} \lesssim A, \quad \||D|^{1/2}\Phi\|_{\text{BMO}} \lesssim B. \tag{4-9}$$

Then, we have the multiplicative bounds

$$\Phi \cdot \mathrm{err}(L^2) = \mathrm{err}(L^2), \qquad \Phi \cdot \mathrm{err}(\dot H^{1/2}) = \mathrm{err}(\dot H^{1/2}), \tag{4-10}$$

$$\Phi \cdot P\,\mathrm{err}(L^2) = A\,\mathrm{err}(L^2), \quad \Phi \cdot P\,\mathrm{err}(\dot H^{1/2}) = A\,\mathrm{err}(\dot H^{1/2}). \tag{4-11}$$

We now expand some of the terms in the above system. For this we will use the following bounds for M_1 (see Lemma A.2):

$$\|M_1\|_{L^\infty} \lesssim A^2, \quad \|M_1\|_{\dot H^1} \lesssim A\boldsymbol{N}_1. \tag{4-12}$$

First we note that

$$M_1 \boldsymbol{W}_\alpha = \mathrm{err}(L^2), \quad M_1 \boldsymbol{R} = \mathrm{err}(\dot H^{1/2}). \tag{4-13}$$

The first is straightforward in view of pointwise bound for M_1. For the second, by the multiplicative bounds lemma, we can replace $M_1\boldsymbol{R}$ by $M_1 R_\alpha$. After a Littlewood–Paley decomposition, the $\dot H^{1/2}$ estimate for $M_1 R_\alpha$ is a consequence of the pointwise bound in (4-12) for low-high and balanced interactions, and of the $\dot H^{1/2}$-bound in (4-12) combined with Lemma 8.1 from Appendix B in [HIT16], provided that we also move half of a derivative from R_α onto M_1 for the high-low interactions, arriving at

$$\|M_1 R_\alpha\|_{\dot H^{1/2}} \lesssim \|M_1\|_{L^\infty}\|R_\alpha\|_{\dot H^{1/2}} + \||D|^{1/2}R\|_{L^\infty}\|M_1\|_{\dot H^1} \lesssim A^2\boldsymbol{N}_1.$$

Next we consider $(1+\boldsymbol{W})M_{1,\alpha}$, for which we claim that

$$\begin{aligned} M_{1,\alpha} &= \boldsymbol{W}_\alpha \overline{\boldsymbol{W}} + 2\boldsymbol{W}\overline{\boldsymbol{W}}_\alpha + \boldsymbol{P}[W\boldsymbol{W}_{\alpha\alpha}] + \mathrm{err}(L^2),\\ \boldsymbol{P}[W\overline{\boldsymbol{W}}_{\alpha\alpha} + \overline{\boldsymbol{W}}_\alpha \boldsymbol{W}] &= A^{-1}\,\mathrm{err}(L^2). \end{aligned} \tag{4-14}$$

By Lemma 4.3, this shows that

$$(1+\boldsymbol{W})M_{1,\alpha} = \boldsymbol{W}_\alpha \overline{\boldsymbol{W}} + 2\boldsymbol{W}\overline{\boldsymbol{W}}_\alpha + \boldsymbol{P}[W\boldsymbol{W}_{\alpha\alpha}] + \mathrm{err}(L^2).$$

To prove (4-14) we discard cubic terms to rewrite

$$\begin{aligned} M_{1,\alpha} &= \boldsymbol{W}_\alpha \bar Y + 2\boldsymbol{W}\bar Y_\alpha + \boldsymbol{P}[W\bar Y_{\alpha\alpha}] - \overline{\boldsymbol{P}}[\boldsymbol{W}_\alpha \bar Y + 2\boldsymbol{W}\bar Y_\alpha + \overline{W}_\alpha Y + 2\overline{\boldsymbol{W}}Y_\alpha + \overline{W}Y_{\alpha\alpha}]\\ &= \boldsymbol{W}_\alpha \overline{\boldsymbol{W}} + 2\boldsymbol{W}\overline{\boldsymbol{W}}_\alpha + \boldsymbol{P}[W\overline{\boldsymbol{W}}_{\alpha\alpha}] - \boldsymbol{P}[f] - \overline{\boldsymbol{P}}[g] + \mathrm{err}(L^2), \end{aligned}$$

where

$$f = W(\overline{\boldsymbol{W}}\bar Y)_{\alpha\alpha}, \quad g = \boldsymbol{W}_\alpha \bar Y + 2\boldsymbol{W}\bar Y_\alpha + \overline{W}_\alpha Y + 2\overline{\boldsymbol{W}}Y_\alpha + \overline{W}Y_{\alpha\alpha}.$$

Finally, for f and g we have L^2-bounds

$$\|\boldsymbol{P}f\|_{L^2} \lesssim A^2\boldsymbol{N}_1, \quad \|\overline{\boldsymbol{P}}g\|_{L^2} \lesssim A\boldsymbol{N}_1,$$

which follow from a commutator-type bound

$$\|P[W\bar\Phi_{\alpha\alpha}]\|_{L^2} \lesssim \|W_\alpha\|_{\mathrm{BMO}}\|\Phi_\alpha\|_{L^2} \tag{4-15}$$

derived from Lemma 8.1 in [HIT16].

The last term in $K_{1,1}$ is antiholomorphic, and also easily placed in $\mathrm{err}(\dot H^{1/2})$ by Coifman–Meyer-type bounds.

Taking into account all of the above expansions, it follows that our system can be rewritten in the form

$$\begin{cases} \boldsymbol{W}_{\alpha t} + \underline{b}\boldsymbol{W}_{\alpha\alpha} + \dfrac{\boldsymbol{R}_\alpha}{1+\overline{\boldsymbol{W}}} + \dfrac{R_\alpha}{1+\overline{\boldsymbol{W}}}\boldsymbol{W}_\alpha = \operatorname{Princ}(G_1) - i\dfrac{c}{2}\operatorname{Princ}(G_{1,1}) + \operatorname{err}(L^2), \\ \boldsymbol{R}_t + \underline{b}\boldsymbol{R}_\alpha + ic\boldsymbol{R} - i\dfrac{(g+\underline{a})\boldsymbol{W}_\alpha}{1+\boldsymbol{W}} = \operatorname{Princ}(K_1) - i\dfrac{c}{2}\operatorname{Princ}(K_{1,1}) + \operatorname{err}(\dot{H}^{1/2}), \end{cases}$$

where $\operatorname{Princ}(G)$ refers to the terms in G that cannot be treated as error; they are quadratic and higher-order terms. We list their expressions below:

$$\begin{aligned} \operatorname{Princ}(G_1) &:= 2\boldsymbol{R}\bar{Y}_\alpha - \frac{2\bar{R}_\alpha \boldsymbol{W}_\alpha}{1+\boldsymbol{W}} + P[R\overline{\boldsymbol{W}}_{\alpha\alpha} - \bar{R}_{\alpha\alpha}\boldsymbol{W}], \\ \operatorname{Princ}(G_{1,1}) &:= -3\boldsymbol{W}\boldsymbol{W}_\alpha + 3\overline{\boldsymbol{W}}\boldsymbol{W}_\alpha + 3\boldsymbol{W}\overline{\boldsymbol{W}}_\alpha + P[W\overline{\boldsymbol{W}}_{\alpha\alpha}], \end{aligned} \tag{4-16}$$

and

$$\begin{aligned} \operatorname{Princ}(K_1) &:= -2\left(\frac{\bar{R}_\alpha}{1+\boldsymbol{W}} + \frac{R_\alpha}{1+\overline{\boldsymbol{W}}}\right)\boldsymbol{R} - P[\bar{R}_{\alpha\alpha}R], \\ \operatorname{Princ}(K_{1,1}) &:= -2\boldsymbol{R}(\boldsymbol{W} - \overline{\boldsymbol{W}}) - \boldsymbol{P}[2\boldsymbol{W}\bar{R}_\alpha + W\bar{R}_{\alpha\alpha} - \overline{\boldsymbol{W}}_\alpha R]. \end{aligned} \tag{4-17}$$

One might wish to compare this system with the linearized system which was studied before. However, of the terms on the right, i.e., those in $\operatorname{Princ}(G_1)$ and $\operatorname{Princ}(K_1)$, cannot be all bounded in $L^2 \times \dot{H}^{1/2}$, even after applying the projection operator $\boldsymbol{P}$. Precisely, the terms on the right which cannot be bounded directly in $L^2 \times \dot{H}^{1/2}$ are

$$-\frac{2\bar{R}_\alpha \boldsymbol{W}_\alpha}{1+\boldsymbol{W}} \quad \text{and} \quad -2\left(\frac{\bar{R}_\alpha}{1+\boldsymbol{W}} + \frac{R_\alpha}{1+\overline{\boldsymbol{W}}}\right)\boldsymbol{R}.$$

There are no such problematic terms in $\operatorname{Princ}(G_{1,1})$ or in $\operatorname{Princ}(K_{1,1})$.

As in [HIT16], we eliminate these terms by conjugation with respect to a real exponential weight $e^{2\phi}$, where $\phi = -2\Re\log(1+\boldsymbol{W})$. Then

$$\phi_\alpha = -2\Re\frac{\boldsymbol{W}_\alpha}{1+\boldsymbol{W}}, \quad (\partial_t + \underline{b}\partial_\alpha)\phi = 2\Re\frac{R_\alpha}{1+\overline{\boldsymbol{W}}} - 2\underline{M} - ic(\boldsymbol{W} - \overline{\boldsymbol{W}})\Re\frac{\boldsymbol{W}}{1+\boldsymbol{W}}.$$

We denote the weighted variables by

$$\mathfrak{w} = e^{2\phi}\boldsymbol{W}_\alpha, \quad \mathfrak{r} = e^{2\phi}\boldsymbol{R}.$$

Using (4-13) and Lemma 4.3 it follows that $\underline{M}w = \operatorname{err}(L^2)$ and $\underline{M}r = \operatorname{err}(\dot{H}^{1/2})$. Similarly we have

$$c\boldsymbol{W}_\alpha(\boldsymbol{W} - \overline{\boldsymbol{W}})\Re Y = \operatorname{err}(L^2), \quad c\boldsymbol{R}(\boldsymbol{W} - \overline{\boldsymbol{W}})\Re Y = \operatorname{err}(\dot{H}^{1/2}).$$

Then the only significant effect of the conjugation is to eliminate the above problem terms, and we are left with

$$\begin{cases} \mathfrak{w}_t + \underline{b}\mathfrak{w}_\alpha + \dfrac{\mathfrak{r}_\alpha}{1+\overline{\boldsymbol{W}}} + \dfrac{R_\alpha}{1+\overline{\boldsymbol{W}}}\mathfrak{w} = \boldsymbol{P}[R\overline{\boldsymbol{W}}_{\alpha\alpha} - \bar{R}_{\alpha\alpha}W] - i\dfrac{c}{2}\operatorname{Princ}(G_{1,1}) + \operatorname{err}(L^2), \\ \mathfrak{r}_t + \underline{b}\mathfrak{r}_\alpha + ic\mathfrak{r} - i\dfrac{(g+\underline{a})\mathfrak{w}}{1+\boldsymbol{W}} = -\boldsymbol{P}[\bar{R}_{\alpha\alpha}R] - i\dfrac{c}{2}\operatorname{Princ}(K_{1,1}) + \operatorname{err}(\dot{H}^{1/2}). \end{cases}$$

Finally, as in [HIT16], we can also harmlessly substitute $\mathfrak{w}$ and $\mathfrak{r}$ into the leading error terms on the right to rewrite the above equation as

$$\begin{cases} \mathfrak{w}_t + \underline{b}\mathfrak{w}_\alpha + \dfrac{\mathfrak{r}_\alpha}{1+\overline{W}} + \dfrac{R_\alpha}{1+\overline{W}}\mathfrak{w} = \boldsymbol{P}[R\bar{\mathfrak{w}}_\alpha - \bar{\mathfrak{r}}_\alpha W] - i\dfrac{c}{2}\operatorname{Princ}(G_{1,1}) + \operatorname{err}(L^2), \\ \mathfrak{r}_t + \underline{b}\mathfrak{r}_\alpha + ic\mathfrak{r} - i\dfrac{(g+\underline{a})\mathfrak{w}}{1+W} = -\boldsymbol{P}[\bar{\mathfrak{r}}_\alpha R] - i\dfrac{c}{2}\operatorname{Princ}(K_{1,1}) + \operatorname{err}(\dot{H}^{1/2}). \end{cases}$$

One downside to the conjugation is that the new variables $(\mathfrak{w}, \mathfrak{r})$ are no longer holomorphic. To remedy this we project onto the holomorphic space to write a system for the variables $(w, r) = (\boldsymbol{P}\mathfrak{w}, \boldsymbol{P}\mathfrak{r})$. At this point one may legitimately be concerned that restricting to the holomorphic part might remove a good portion of our variables. However, this is not the case, as one can verify the claim by reviewing the discussion in Lemma 3.4 in [HIT16]. Following again [HIT16], after estimating some Coifman–Meyer-type commutators, the system for (w, r) is

$$\begin{cases} w_t + \mathfrak{M}_{\underline{b}} w_\alpha + \boldsymbol{P}\left[\dfrac{r_\alpha}{1+\overline{W}}\right] + \boldsymbol{P}\left[\dfrac{R_\alpha}{1+\overline{W}}w\right] = G, \\ r_t + \mathfrak{M}_{\underline{b}} r_\alpha + icr - i\boldsymbol{P}\left[\dfrac{(g+\underline{a})w}{1+W}\right] = K, \end{cases} \tag{4-18}$$

where

$$\begin{cases} G := \boldsymbol{P}[R\bar{w}_\alpha - \bar{r}_\alpha W] - i\dfrac{c}{2}\boldsymbol{P}[\operatorname{Princ}(G_{1,1})] + \operatorname{err}(L^2), \\ K := -\boldsymbol{P}[\bar{r}_\alpha R] - i\dfrac{c}{2}\boldsymbol{P}[\operatorname{Princ}(K_{1,1})] + \operatorname{err}(\dot{H}^{1/2}). \end{cases}$$

This is our main system for the (slightly renormalized) differentiated variables $(\boldsymbol{W}_\alpha, R_\alpha)$. In order to use it we need to properly relate (w, r) to $(\boldsymbol{W}_\alpha, R_\alpha)$, and to estimate the terms in G and K. This is done in the following lemma.

Lemma 4.4. (a) *The energy of* (w, r) *above is equivalent to that of* $(\boldsymbol{W}_\alpha, R_\alpha)$,

$$\|(w, r)\|_{L^2\times\dot{H}^{1/2}} \approx_{\underline{A}} \|(\boldsymbol{W}_\alpha, R_\alpha)\|_{L^2\times\dot{H}^{1/2}} = N_1, \tag{4-19}$$

and the difference is estimated by

$$\|(w, r) - (\boldsymbol{W}_\alpha, R_\alpha)\|_{L^2\times\dot{H}^{1/2}} \lesssim_A \underline{A}N_1. \tag{4-20}$$

(b) *The error terms on the right in* (4-18) *are bounded*,

$$\|(\boldsymbol{P}[R\bar{w}_\alpha - \boldsymbol{W}\bar{r}_\alpha], \boldsymbol{P}[R\bar{r}_\alpha])\|_{L^2\times\dot{H}^{1/2}} \lesssim_A BN_1, \quad \|P[R\bar{r}_\alpha]\|_{L^2} \lesssim AN_1, \tag{4-21}$$

and

$$\|(\boldsymbol{P}\operatorname{Princ}(G_{1,1}), \boldsymbol{P}\operatorname{Princ}(K_{1,1}))\|_{L^2\times\dot{H}^{1/2}} \lesssim_A AN_1, \quad \|\boldsymbol{P}\operatorname{Princ}(K_{1,1})\|_{L^2} \lesssim_A A_{-1/2}N_1. \tag{4-22}$$

Proof. Part (a) and the first estimate in part (b) are from [HIT16], while part (c) follows either directly, for some terms, or by Coifman–Meyer estimates for the rest. ☐

Given the above lemma, the conclusion of Proposition 4.1 for $n=1$ follows from the energy estimates for the linearized equation, namely (3-8) in Proposition 3.1; further, if $n=1$ then we can take

$$E^{n,(2)}(\boldsymbol{W}, R) = E^{(2)}_{\mathrm{lin}}(w, r).$$

4C. *The differentiated equations for n ≥ 2.* The first step is to derive a set of equations for $(\boldsymbol{W}^{(n)}, R^{(n)})$. We start again with the differentiated equations (1-19) and differentiate n times.

Compared with the case $n=1$, we obtain many more terms. To separate them into leading order and lower order, we call lower-order terms any terms which do not involve $\boldsymbol{W}^{(n)}$, $R^{(n)}$ or derivatives thereof. In the computation below we take care to separate all the leading-order terms. Toward that end we define again the notion of *error term*. Unlike in the case $n=1$, here we also include lower-order quadratic terms into the error. As before, we describe the error bounds in terms of the parameters $\underline{A}$, $\underline{B}$ and

$$\boldsymbol{N}_n = \|(\boldsymbol{W}^{(n)}, R^{(n)})\|_{L^2\times \dot{H}^{1/2}}. \tag{4-23}$$

The acceptable errors in the $\boldsymbol{W}^{(n)}$-equation are denoted by $\mathrm{err}(L^2)$ and are of two types, $\mathrm{err}(L^2)^{[2]}$ and $\mathrm{err}(L^2)$. The lower-order quadratic holomorphic terms are placed in $\mathrm{err}(L^2)^{[2]}$, which is defined to be a linear combination of expressions of the form

$$\boldsymbol{P}[\boldsymbol{W}^{(j)} R^{(n+1-j)}], \quad \boldsymbol{P}[\overline{\boldsymbol{W}}^{(j)} R^{(n+1-j)}], \quad \boldsymbol{P}[\boldsymbol{W}^{(j)} \bar{R}^{(n+1-j)}], \quad 2\le j\le n-1,$$

as well as terms involving the vorticity c, namely

$$\begin{aligned} &c\boldsymbol{P}[\boldsymbol{W}^{(j)}\boldsymbol{W}^{(n-j)}], && c\boldsymbol{P}[\overline{\boldsymbol{W}}^{(j)}\boldsymbol{W}^{(n-j)}], && 1\le j\le n-1,\\ &c\boldsymbol{P}[R^{(j)}\bar{R}^{(n+1-j)}], && c\boldsymbol{P}[R^{(j)}R^{(n+1-j)}], && 2\le j\le n-1. \end{aligned}$$

By interpolation and Hölder's inequality, terms G in $\mathrm{err}(L^2)^{[2]}$ satisfy the bound

$$\|G\|_{L^2} \lesssim \underline{B}\boldsymbol{N}_n.$$

By $\mathrm{err}(L^2)^{[3]}$ we denote terms G which satisfy the same estimates as $\mathrm{err}(L^2)$ in the case $n=1$, but with $\boldsymbol{N}_1$ replaced by $\boldsymbol{N}_n$.

The acceptable errors in the $R^{(n)}$-equation are denoted by $\mathrm{err}(\dot{H}^{1/2})$ and are also of two types, $\mathrm{err}(\dot{H}^{1/2})^{[2]}$ and $\mathrm{err}(\dot{H}^{1/2})^{[3]}$. The first, $\mathrm{err}(\dot{H}^{1/2})^{[2]}$, consists of holomorphic quadratic lower-order terms of the form

$$\begin{aligned} &\boldsymbol{P}[R^{(j)}R^{(n+1-j)}], && \boldsymbol{P}[\bar{R}^{(j)}R^{(n+1-j)}], && 2\le j\le n-1,\\ &\boldsymbol{P}[\boldsymbol{W}^{(j)}\boldsymbol{W}^{(n+1-j)}], && \boldsymbol{P}[\overline{\boldsymbol{W}}^{(j)}\boldsymbol{W}^{(n-j)}], && 1\le j\le n-1, \end{aligned}$$

as well as the c-terms

$$c\boldsymbol{P}[\boldsymbol{W}^{(j)}R^{(n-j)}], \quad c\boldsymbol{P}[\overline{\boldsymbol{W}}^{(j)}R^{(n-j)}], \quad c\boldsymbol{P}[\boldsymbol{W}^{(j)}\bar{R}^{(n-j)}], \quad 1\le j\le n-1.$$

By interpolation and Hölder's inequality, terms K in $\mathrm{err}(\dot{H}^{1/2})^{[2]}$ satisfy the bound

$$\|K\|_{\dot{H}^{1/2}} \lesssim \underline{B}\boldsymbol{N}_n, \quad \|K\|_{L^2} \lesssim A\boldsymbol{N}_n.$$

By $\text{err}(\dot{H}^{1/2})^{[3]}$ we denote terms K which satisfy the same estimates as $\text{err}(\dot{H}^{1/2})$ in the case $n=1$, but with $\boldsymbol{N}_1$ replaced by $\boldsymbol{N}_n$.

We now proceed to differentiate n times the equation (1-19). Our task is simplified due to [HIT16], where this analysis has already been carried out for the terms without c. Hence, we only concentrate here on the c-terms.

In addition, we remark that, as all terms in the $\boldsymbol{W}^{(n)}$-equation have the same homogeneity, whenever all the Sobolev exponents are within the lower-order range, we are guaranteed to get the correct L^2 estimate after interpolation and Hölder's inequality. The same applies to all cubic terms with at most n derivatives on any single factor. The same observation applies to all the lower-order terms in the $R^{(n)}$-equation, as well as to all cubic terms containing $R^{(n)}$. However, the terms containing $\boldsymbol{W}^{(n)}$ are unbounded and belong to the principal part of the $R^{(n)}$-equation. Because of these considerations, the computation below is largely of an algebraic nature.

We begin with the $\boldsymbol{W}$-equation. For the b_1-term we have

$$\begin{aligned}\partial^n(b_1\boldsymbol{W}_\alpha) &= b_1\boldsymbol{W}_\alpha^{(n)} + nb_{1,\alpha}\boldsymbol{W}^{(n)} + b_1^{(n)}\boldsymbol{W}_\alpha + \text{err}(L^2)\\ &= b_1\boldsymbol{W}_\alpha^{(n)} + n(\boldsymbol{W}-\overline{\boldsymbol{W}})\boldsymbol{W}^{(n)} + \text{err}(L^2).\end{aligned}$$

Here at the last step we use the simple observation that the linear part of the expression $b_1^{(n)}\boldsymbol{W}_\alpha$ contributes only lower-order terms, and the rest contributes cubic terms, which can be estimated either directly or using Coifman–Meyer bounds.

Next we consider

$$\begin{aligned}\partial^n((1+\boldsymbol{W})M_1) &= (1+\boldsymbol{W})\partial^{n+1}(\boldsymbol{P}[W\bar{Y}] - \bar{\boldsymbol{P}}[\overline{W}Y]) + \text{err}(L^2)\\ &= \boldsymbol{W}^{(n)}\boldsymbol{W} + (n+1)\boldsymbol{W}\overline{\boldsymbol{W}}^{(n)} + \boldsymbol{P}[W\overline{\boldsymbol{W}}^{(n+1)}] + \text{err}(L^2).\end{aligned}$$

Here all cubic terms involving $\overline{\boldsymbol{W}}^{(n+1)}$ can be directly bounded using Coifman–Meyer estimates. Finally, for the last c-term in the $\boldsymbol{W}^{(n)}$-equation we have

$$\partial^n(\boldsymbol{W}(\boldsymbol{W}-\overline{\boldsymbol{W}})) = 2\boldsymbol{W}^{(n)}\boldsymbol{W} - \boldsymbol{W}^{(n)}\overline{\boldsymbol{W}} - \boldsymbol{W}\overline{\boldsymbol{W}}^{(n)} + \text{err}(L^2).$$

Next we turn our attention to the c-terms in the $R^{(n)}$-equation, We begin with

$$\begin{aligned}\partial^{n-1}(b_1R_\alpha) &= b_1R_\alpha^{(n)} + nb_{1,\alpha}R^{(n)} + b^{(n)}R_\alpha + \text{err}(\dot{H}^{1/2})\\ &= b_1R_\alpha^{(n)} + n(\boldsymbol{W}-\overline{\boldsymbol{W}})R^{(n)} + \text{err}(\dot{H}^{1/2}),\end{aligned}$$

where again the quadratic terms in $b^{(n)}R_\alpha$ are lower-order, and the cubic terms with $\overline{\boldsymbol{W}}^{(n)}$ are bounded as error terms via Coifman–Meyer estimates.

For the next term in the R-equation we write

$$\partial^n\frac{\boldsymbol{W}(R+\bar{R})}{1+\boldsymbol{W}} = \boldsymbol{W}(R^{(n)}+\bar{R}^{(n)}) + \frac{\boldsymbol{W}^{(n)}(R+\bar{R})}{1+\boldsymbol{W}} + \text{err}(\dot{H}^{1/2}),$$

where the cubic $R^{(n)}$-terms are estimated directly as error terms. Finally, we need to consider

$$\partial^n \frac{N}{1+\boldsymbol{W}} = \partial^n \left(\frac{\boldsymbol{P}[W\bar{R}_\alpha - \overline{\boldsymbol{W}}R] + \bar{\boldsymbol{P}}[\overline{W}R_\alpha - \boldsymbol{W}\bar{R}]}{1+\boldsymbol{W}} \right)$$
$$= -\frac{N\boldsymbol{W}^{(n)}}{(1+\boldsymbol{W})^2} + \boldsymbol{P}[W\bar{R}^{(n)}_\alpha + n\boldsymbol{W}\bar{R}^{(n)} - \overline{\boldsymbol{W}}^{(n)}R - \overline{\boldsymbol{W}}R^{(n)}] + \mathrm{err}(\dot{H}^{1/2}).$$

Combining the above computations for the c-terms in (1-19) with the prior computations for the non-c-terms in [HIT16] we obtain the differentiated system

$$\begin{cases} \boldsymbol{W}^{(n)}_t + \underline{b}\boldsymbol{W}^{(n)}_\alpha + \dfrac{((1+\boldsymbol{W})R^{(n)})_\alpha}{1+\overline{\boldsymbol{W}}} + \dfrac{R_\alpha}{1+\boldsymbol{W}}\boldsymbol{W}^{(n)} = \underline{G}_n + \mathrm{err}(L^2), \\ R^{(n)}_t + \underline{b}R^{(n)}_\alpha + icR^{(n)} - i\left(\dfrac{(g+\underline{a})\boldsymbol{W}^{(n)}}{(1+\boldsymbol{W})^2}\right) = \underline{K}_n + \mathrm{err}(\dot{H}^{1/2}), \end{cases}$$

where

$$\underline{G}_n = G_n - i\frac{c}{2}G_{n,1}, \qquad \underline{K}_n = K_n - i\frac{c}{2}K_{n,1}.$$

From [HIT16] we have

$$\begin{cases} G_n = -n\dfrac{\bar{R}_\alpha}{1+\boldsymbol{W}}\boldsymbol{W}^{(n-1)} - (n-1)\dfrac{R_\alpha}{1+\overline{\boldsymbol{W}}}\boldsymbol{W}^{(n-1)} + P[R\overline{\boldsymbol{W}}^{(n-1)}_\alpha - \boldsymbol{W}\bar{R}^{(n-1)}_\alpha] \\ \qquad + R^{(n-1)}(n\overline{\boldsymbol{W}}_\alpha - (n-1)\boldsymbol{W}_\alpha) + n(R_\alpha\overline{\boldsymbol{W}}^{(n-1)} - \boldsymbol{W}_\alpha\bar{R}^{(n-1)}), \\ K_n = -n\left(\dfrac{R_\alpha}{1+\overline{\boldsymbol{W}}} + \dfrac{\bar{R}_\alpha}{1+\boldsymbol{W}}\right)R^{(n-1)} - (P[R\bar{R}^{(n-1)}_\alpha] - nR_\alpha\bar{R}^{(n-1)}), \end{cases}$$

and from the above computations,

$$\begin{cases} G_{n,1} = (n+2)[\boldsymbol{W}\overline{\boldsymbol{W}}^{(n)} - (\boldsymbol{W}-\overline{\boldsymbol{W}})\boldsymbol{W}^{(n)}] + \boldsymbol{P}[W\overline{\boldsymbol{W}}^{(n+1)}], \\ K_{n,1} = -(n+1)[(\boldsymbol{W}-\overline{\boldsymbol{W}})R^{(n)} + \boldsymbol{W}\bar{R}^{(n)}] + \boldsymbol{P}[R\overline{\boldsymbol{W}}^{(n)} - W\bar{R}^{(n)}_\alpha]. \end{cases}$$

To bring this equation to a form closer to the linearized equation we follow the lead of [HIT16] and perform several algebraic holomorphic substitutions for the $R^{(n)}$-variable, beginning with

$$\boldsymbol{R} = (1+\boldsymbol{W})R^{(n)},$$

and followed by

$$\widetilde{\boldsymbol{R}} = \boldsymbol{R} - R_\alpha\boldsymbol{W}^{(n-1)} + (2n+1)\boldsymbol{W}_\alpha R^{(n-1)}.$$

Finally, we conclude with the exponential conjugation by $e^{n\phi}$ where we use the same ϕ as before, namely $\phi = -2\Re\log(1+\boldsymbol{W})$, in order to eliminate the unbounded terms on the right. At the end we obtain an equation for ($\mathfrak{w} := e^{n\phi}\boldsymbol{W}^{(n-1)}$, $\mathfrak{r} := e^{n\phi}\widetilde{\boldsymbol{R}}$) where the leading part is exactly as in the linearized equation. We do not repeat this computation, as it primarily affects the part of the equation without c, which is fully described in [HIT16]. All the additional c-contributions are either cubic and bounded or quadratic and lower-order, so they are placed in the error. The outcome is the equation

$$\begin{cases} \mathfrak{w}_t + \underline{b}\mathfrak{w}_\alpha + \dfrac{\mathfrak{r}_\alpha}{1+\overline{\mathbf{W}}} + \dfrac{R_\alpha}{1+\mathbf{W}}\mathfrak{w} = \mathbf{P}[R\overline{\mathbf{W}}^{(n)}_\alpha - \mathbf{W}\bar{R}^{(n)}_\alpha] + (n+1)(R_\alpha\bar{\mathfrak{w}} - \mathbf{W}_\alpha\bar{\mathfrak{r}}) - i\,\dfrac{c}{2}G_{n,1} + \mathrm{err}(L^2), \\ \mathfrak{r}_t + \underline{b}\mathfrak{r}_\alpha + ic\mathfrak{r} - i\left(\dfrac{(g+\underline{a})\mathfrak{w}}{1+\mathbf{W}}\right) = -\mathbf{P}[R\bar{R}^{(n)}_\alpha] - (n+1)R_\alpha\bar{\mathfrak{r}} - i\,\dfrac{c}{2}K_{n,1} + \mathrm{err}(\dot{H}^{1/2}). \end{cases}$$

As $(\mathfrak{w}, \mathfrak{r})$ are no longer holomorphic, we project and work with the projected variables $(w, r) = (\mathbf{P}\mathfrak{w}, \mathbf{P}\mathfrak{r})$. After some additional commutator estimates, which are identical to those in the $n = 1$ case, we finally obtain

$$\begin{cases} w_t + \mathfrak{M}_{\underline{b}} w_\alpha + \mathbf{P}\left[\dfrac{r_\alpha}{1+\overline{\mathbf{W}}}\right] + \mathbf{P}\left[\dfrac{R_\alpha w}{1+\mathbf{W}}\right] = \mathbf{P}[R\bar{w}_\alpha - \mathbf{W}\bar{r}_\alpha] + (n+1)\mathbf{P}[R_\alpha\bar{w} - \mathbf{W}_\alpha\bar{r}] \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad -i\,\dfrac{c}{2}\mathbf{P}G_{n,1} + \mathrm{err}(L^2), \\ r_t + \mathfrak{M}_{\underline{b}} r_\alpha + icr - i\mathbf{P}\left[\dfrac{(g+\underline{a})w}{1+\mathbf{W}}\right] = -\mathbf{P}[R\bar{r}_\alpha] - (n+1)\mathbf{P}[R_\alpha\bar{r}] - i\,\dfrac{c}{2}\mathbf{P}K_{n,1} + \mathrm{err}(\dot{H}^{1/2}). \end{cases} \tag{4-24}$$

This is our main system for the (slightly modified) differentiated variables $(\mathbf{W}^{(n)}, R^{(n)})$. In order to use it we again need to relate (w, r) to $(\mathbf{W}^{(n)}, R^{(n)})$, and also to estimate the terms in $G_{n,1}$ and $K_{n,1}$. This is done in the following.

Lemma 4.5. (a) *The energy of* (w, r) *above is equivalent to that of* $(\mathbf{W}^{(n)}, R^{(n)})$,

$$\|(w, r)\|_{L^2\times\dot{H}^{1/2}} \approx_{\underline{A}} \|(\mathbf{W}^{(n)}, R^{(n)})\|_{L^2\times\dot{H}^{1/2}} = \mathbf{N}_n, \tag{4-25}$$

and the difference is estimated by

$$\|(w, r) - (\mathbf{W}^{(n)}, R^{(n)})\|_{L^2\times\dot{H}^{1/2}} \lesssim_A \underline{A}\mathbf{N}_n. \tag{4-26}$$

(b) *The error terms on the right in* (4-18) *are bounded*,

$$\|(\mathbf{P}[R\bar{w}_\alpha - \mathbf{W}\bar{r}_\alpha], \mathbf{P}[R\bar{r}_\alpha])\|_{L^2\times\dot{H}^{1/2}} \lesssim_A B\mathbf{N}_n, \quad \|\mathbf{P}[R\bar{r}_\alpha]\|_{L^2} \lesssim A\mathbf{N}_n, \tag{4-27}$$

$$\|(\mathbf{P}[R_\alpha\bar{w} - \mathbf{W}_\alpha\bar{r}], \mathbf{P}[R_\alpha\bar{r}])\|_{L^2\times\dot{H}^{1/2}} \lesssim_A B\mathbf{N}_n, \quad \|\mathbf{P}[R_\alpha\bar{r}]\|_{L^2} \lesssim_A A\mathbf{N}_n, \tag{4-28}$$

and

$$\|(\mathbf{P}G_{1,n}, \mathbf{P}K_{1,n})\|_{L^2\times\dot{H}^{1/2}} \lesssim_A A, \quad \|\mathbf{P}K_{1,n}\|_{L^2} \lesssim_A A_{-1/2}\mathbf{N}_n. \tag{4-29}$$

Proof. Part (a) and the first two bounds in part (b) are from [HIT16], while the last bound in part (b) follows either directly, for some terms, or by Coifman–Meyer estimates, for the rest. □

Given the above Lemma 4.4, the $n \geq 2$ case of the result in Proposition 4.1 is a direct consequence of our quadratic estimates for the linearized equation in Proposition 3.1.

Comparing the linear (w, r)-equation obtained above for the case $n \geq 2$ to the corresponding equation previously obtained for the case $n = 1$, we note that here we have two extra terms, namely the ones estimated in (4-28). On one hand this is good, because the bound (4-28) is no longer true if $n = 1$. But on the other hand, it means that we will no longer be able to use directly the cubic energy $E^{(3)}_{\mathrm{lin}}(w, r)$

introduced in the context of the linearized equation as the full high-frequency part of our cubic energy. To account for these extra terms, we will add a further correction to the energy $E^{(3)}_{\text{lin}}(w,r)$, and define

$$E^{n,(3)}_{\text{high}}(w,r) := \int (1+a)|w|^2 + \Im(\bar{r}r_\alpha) + 2n\Im(R_\alpha\bar{w}\bar{r}) + 2(\Im[\bar{R}wr_\alpha] - \Re[\bar{W}_\alpha w^2])\,d\alpha. \tag{4-30}$$

Then we will prove a counterpart to the small-data (nearly) cubic energy estimates in Proposition 4.2. Precisely, we claim that the evolution of this energy is governed by the following.

Lemma 4.6. *Let (w,r) be defined as above. Then*:

(a) *Assuming that $A \ll 1$, we have*

$$E^{n,(3)}_{\text{high}}(w,r) \approx \mathcal{E}_0(w,r) \approx \boldsymbol{N}_n. \tag{4-31}$$

(b) *The solutions (w,r) of (4-24) satisfy*

$$\begin{aligned}\frac{d}{dt}E^{n,(3)}_{\text{high}}(w,r) = 2\int -\Re\Big(\bar{w}\cdot\Big(i\frac{c}{2}G_{n,1} - \text{err}(L^2)^{[2]}\Big)\Big) + \Im\Big(\bar{r}_\alpha\cdot\Big(i\frac{c}{2}K_{n,1} - \text{err}(\dot{H}^{1/2})^{[2]}\Big)\Big)\,d\alpha \\ + \int \frac{c^2}{2}\Im R|w|^2\,d\alpha + O_A(A\underline{B}\boldsymbol{N}_n).\end{aligned} \tag{4-32}$$

Further, the same relation holds if $(\bar{w},\bar{r})$ on the right are replaced by $(\bar{\boldsymbol{W}}^{(n)}, \bar{R}^{(n)})$.

We note that in the proof of this lemma we crucially use the fact that (w,r) are the ones associated to the differentiated equation, and not arbitrary functions in the energy space. This is also the reason why this lemma is presented here, rather than in Section 3.

Proof. Part (a) was already proved in [HIT16], so we proceed to part (b). Here, we begin with the cubic linearized energy, $E^{(3)}_{\text{lin}}$. According to the bound (3-18) in Proposition 3.4, we have

$$\begin{aligned}\frac{d}{dt}E^{(3)}_{\text{lin}}(w,r) = \int 2\Re\Big(\Big(n\boldsymbol{P}[R_\alpha\bar{w} - \boldsymbol{W}_\alpha\bar{r}] - i\frac{c}{2}G_{n,1} + P\,\text{err}(L^2)\Big)\cdot(\bar{w} - \bar{\boldsymbol{P}}[\bar{R}r_\alpha] - \bar{\boldsymbol{P}}[\bar{W}_\alpha w])\Big) \\ - 2\Im\Big(\Big(-n\boldsymbol{P}[R_\alpha\bar{r}] - i\frac{c}{2}K_{n,1} + \boldsymbol{P}\,\text{err}(\dot{H}^{1/2})\Big)\cdot(\bar{r}_\alpha + \bar{\boldsymbol{P}}[\bar{R}w]_\alpha)\Big)\,d\alpha \\ + O_A(\underline{A}\underline{B}\|(w,r)\|^2_{L^2\times\dot{H}^{1/2}}).\end{aligned}$$

By the Coifman–Meyer-type estimates in Lemma A.1 the following bound holds:

$$\|\bar{P}[\bar{R}r_\alpha]\|_{L^2} + \|\bar{P}[\bar{W}_\alpha w]\|_{L^2} + \|\bar{P}[\bar{R}w]\|_{\dot{H}^{1/2}} \lesssim A\|(w,r)\|_{L^2\times\dot{H}^{1/2}}. \tag{4-33}$$

Combining this with (4-21) and with the bounds for $G_{n,1}$, $K_{n,1}$ and the error terms we get

$$\begin{aligned}\frac{d}{dt}E^{(3)}_{\text{lin}}(w,r) \le \int 2\Re\Big(\Big(n\boldsymbol{P}[R_\alpha\bar{w} - \boldsymbol{W}_\alpha\bar{r}] - i\frac{c}{2}G_{n,1} + \boldsymbol{P}\,\text{err}(L^2)^{[2]}\Big)\cdot\bar{w}\Big) \\ - 2\Im\Big(\Big(-n\boldsymbol{P}[R_\alpha\bar{r}] - i\frac{c}{2}K_{n,1} + \boldsymbol{P}\,\text{err}(\dot{H}^{1/2})^{[2]}\Big)\cdot\bar{r}_\alpha\Big)\,d\alpha \\ + O_A(\underline{A}\underline{B})\|(w,r)\|^2_{L^2\times\dot{H}^{1/2}},\end{aligned}$$

where the output from all error terms which are cubic-and-higher is all included in the last term on the right-hand side.

It remains to consider the contribution of the extra term in $E^{n,(3)}_{\text{high}}$ and show that

$$\frac{d}{dt}\int \Im(R_\alpha \bar{w}\bar{r})\,d\alpha = \int \Re((R_\alpha \bar{w} - \boldsymbol{W}_\alpha \bar{r})\bar{w}) + \Im(R_\alpha \bar{r}\bar{r}_\alpha)\,d\alpha + O_A(\underline{A}\underline{B})\|(w,r)\|^2_{L^2\times \dot{H}^{1/2}}. \tag{4-34}$$

Denote by G_n and K_n, respectively, the two right-hand sides in (4-24). By the definition of error terms and by (4-21)–(4-22) they satisfy the bounds

$$\|(G_n, K_n)\|_{L^2\times \dot{H}^{1/2}} \lesssim_A \underline{B}\boldsymbol{N}_n, \quad \|K_n\|_{L^2} \lesssim_A \underline{A}\boldsymbol{N}_n.$$

Then their contribution in the above time derivative is estimated as

$$\left|\int \Im(R_\alpha \bar{\boldsymbol{P}}\bar{G}_n\bar{r} + R_\alpha \bar{w}\bar{\boldsymbol{P}}\bar{K}_n)\,d\alpha\right| = \left|\int \Im(R_\alpha \bar{\boldsymbol{P}}\bar{F}_n\bar{r} + P[R_\alpha \bar{w}]\bar{\boldsymbol{P}}\bar{K}_n)\,d\alpha\right| \lesssim_A \underline{A}\underline{B}\boldsymbol{N}_n$$

by using Hölder's inequality for the first term and the Coifman–Meyer commutator estimate in Lemma A.1 for the second.

We now consider the contributions due to the terms on the left of (4-24). These were already estimated in [HIT16] when $c = 0$, so we only need to estimate the c-terms, i.e., the contributions of b_1 and a_1.

The contributions of the b_1-terms are collected together in the real part of the expression

$$\begin{aligned} I &= \frac{c}{2}\int \partial_\alpha(b_1 R_\alpha)\bar{w}\bar{r} + R_\alpha \bar{\boldsymbol{P}}(b_1\bar{w}_\alpha)\bar{r} + R_\alpha \bar{w}\bar{\boldsymbol{P}}(b_1\bar{r}_\alpha)\,d\alpha \\ &= \frac{c}{2}\int R_\alpha\big([b_1, \boldsymbol{P}](\bar{w}_\alpha)\bar{r} + \bar{w}[b_1, \boldsymbol{P}](\bar{r}_\alpha)\big)\,d\alpha. \end{aligned}$$

Since $\|b_{1,\alpha}\|_{\text{BMO}} \lesssim A$, we can bound this using Lemma A.1, and then use Hölder's inequality for all terms.

The contribution of a_1 is

$$\frac{c}{2}\Re\int R_\alpha \bar{w} P\left[\frac{a_1\bar{w}}{1+\boldsymbol{W}}\right]d\alpha,$$

for which it suffices to use the pointwise bound for a_1 in Proposition A.4 and the classical Coifman–Meyer bound.

Finally, we consider the remaining contribution of the c-terms in the time derivative of R_α, for which we use the equations (1-19)

$$\frac{c}{2}\Re\int \bar{w}\bar{r}\partial_\alpha \frac{R\boldsymbol{W} + \bar{R}\boldsymbol{W} + N}{1+\boldsymbol{W}}\,d\alpha.$$

But this is a quartic expression for which we only need Sobolev embeddings and interpolation. □

4D. ***Cubic small-data energy estimates: $n \geq 1$.*** In this section we construct an n-th order energy functional $E^{n,(3)}$ with cubic estimates, for $n \geq 1$. While in essence the argument does not depend on n, there are nevertheless a few small differences between the cases $n = 1$ and $n \geq 2$. These differences will be pointed out at the appropriate places.

One ingredient in the construction of $E^{n,(3)}$ is the high-frequency (nearly) cubic energy $E^{n,(3)}_{\text{high}}$ in Lemma 4.6 ($n \geq 2$), and $E^{(3)}_{\text{lin}}$ in Proposition 3.4 ($n = 1$). However, this does not suffice, as the right-hand side of the energy relation (4-32) still contains lower-order cubic terms. To remedy this, here we use normal forms in order to add a lower-order correction to $E^{n,(3)}_{\text{high}}$, which removes the above-mentioned cubic terms. For this we follow the method introduced in [HIT16], though the computations here are significantly more involved.

We first provide an outline of the argument, and then return to each step in detail. The three main steps are as follows:

(1) *Construct the normal-form energy*. The normal-form variables $\widetilde{W}$, $\widetilde{Q}$ solve a system of the form

$$\begin{cases} \widetilde{W}_t + \widetilde{Q}_\alpha = \text{cubic and higher}, \\ \widetilde{Q}_t - i\widetilde{W} = \text{cubic and higher}. \end{cases} \tag{4-35}$$

Thus, the associated linear energies satisfy

$$\frac{d}{dt}\mathcal{E}_0(\widetilde{W}^{(n+1)}, \widetilde{Q}^{(n+1)}) = \text{quartic and higher}.$$

Here the quartic part of the left-hand side is highly unbounded, and it is not uniquely determined, as one can add arbitrary cubic terms to the normal form. Fortunately, it is also irrelevant for the above relation.

Thus, we eliminate it, and we define the normal-form energy as

$$E^n_{\text{NF}} = (\text{quadratic + cubic})[\mathcal{E}_0(\widetilde{W}^{(n+1)}, \widetilde{Q}^{(n+1)})].$$

Here, we carefully choose the quadratic and cubic terms in the expansion of $\mathcal{E}_0(\widetilde{W}^{(n)}, \widetilde{Q}^{(n)})$ in terms of the diagonal variables $(\boldsymbol{W}, R)$, rather than (W, Q).

The expression E^n_{NF} has only quadratic and cubic terms in $(\boldsymbol{W}, R)$, and still satisfies

$$\frac{d}{dt}E^n_{\text{NF}} = \text{unbounded quartic and higher}.$$

Its disadvantage, due to the fact that our problem is quasilinear, is that the terms on the right are highly unbounded.

(2) *Construct the quasilinear modified energy*. Here we construct *the quasilinear modified energy* $E^{n,(3)}$ using the normal-form energy E^n_{NF} and the high frequency modified energy $E^{n,(3)}_{\text{high}}(w, r)$. The first one has quartic estimates, which are unbounded. The second one has good (quartic) high-frequency estimates, but with cubic lower-order errors,

$$\frac{d}{dt}E^{n,(3)}_{\text{high}}(w, r) = \text{bounded quartic and higher} + \text{lower-order cubic}.$$

We seek to combine the two into a quasilinear modified energy $E^{n,(3)}$ which satisfies

$$\frac{d}{dt}E^{n,(3)} = \text{bounded quartic and higher}.$$

To achieve this, it is natural to separate E^n_{NF} into a leading part and a lower-order part,

$$E^n_{\text{NF}} = E^n_{\text{NF,high}} + E^n_{\text{NF,low}},$$

so that the leading part coincides with the high-frequency energy up to quartic terms,

$$E^{n,(3)}_{\text{high}} = E^n_{\text{NF,high}} + \text{quartic and higher},$$

and then define

$$E^{n,(3)} = E^{n,(3)}_{\text{high}} + E^n_{\text{NF,low}}.$$

This was the argument in [HIT16]; here it is slightly more complicated, as additional terms in $E^n_{\text{NF,low}}$ need to be treated as leading-order, and thus corrected in a quasilinear fashion.

(3) *Show that $E^{n,(3)}$ is a good quasilinear cubic energy*. That is, prove that the bound (4-2) holds. Here, by construction, we know that

$$\frac{d}{dt}E^{n,(3)} = \text{quartic and higher};$$

therefore we can write

$$\frac{d}{dt}E^{n,(3)} = \text{quartic and higher}\Big(\frac{d}{dt}E^{n,(3)}_{\text{high}}(w,r)\Big) + \text{quartic and higher}\Big(\frac{d}{dt}E^n_{\text{NF,low}}\Big).$$

The favorable bound for the first part is already a direct consequence of Lemma 4.6, so it remains to use the equations in order to bound the second term. But this is technically simple, and the argument is more a matter of bookkeeping.

Now we proceed to implement the above strategy.

Step 1. The first step described above is implemented in the following proposition:

Proposition 4.7. *There exists a modified normal-form energy $E^n_{\text{NF}} = E^n_{\text{NF}}(W, R)$ with the following properties*:

(A) *E^n_{NF} has the form*

$$E^n_{\text{NF}} = E^n_{\text{NF,high}} + E^{n,c}_{\text{NF,high}} + E^n_{\text{NF,low}},$$

where the three components are defined as follows:

$$\begin{aligned} E^n_{\text{NF,high}} = &\int (1-4(n+1)\Re \boldsymbol{W})(g|\boldsymbol{W}^{(n)}|^2 + \Im[\bar{R}^{(n+1)}R^{(n)}])\,d\alpha, \\ &+2\int \Im[\bar{R}\boldsymbol{W}^{(n)}R^{(n+1)}] - g\Re[\overline{\boldsymbol{W}}(\boldsymbol{W}^{(n)})^2] + (n+1)\Im[R_\alpha \overline{\boldsymbol{W}}^{(n)}\bar{R}^{(n)}]\,d\alpha \\ &+\int c\Re R|\boldsymbol{W}^{(n)}|^2 + 2\Im(\boldsymbol{W}\bar{R}^{(n+1)}R^{(n)})\,d\alpha, \end{aligned} \tag{4-36}$$

where the last term in the second integral is dropped if $n=1$, and

$$E^{n,c}_{\text{NF,high}} = -\int \big[c(2n+3)\Re R + c^2\big(2n+\tfrac{5}{2}\big)\Im W\big](|\boldsymbol{W}^{(n)}|^2 - i\bar{R}^{(n+1)}R^{(n)})\,d\alpha. \tag{4-37}$$

Finally, $E^n_{\text{NF,low}}$ is a trilinear integral form of order $2n$, with the following restrictions:

(i) *The leading order for the terms without c is at most $2n-1$.*

(ii) *The leading order for the terms with c is at most $2n-\frac{1}{2}$.*

(iii) *The highest power of c is 4.*

(iv) *At most one factor of the form c^2W or cR is present.*

(B) *The energy functional E^n_{NF} is cubically accurate*,

$$\Lambda^{\leq 3}\frac{d}{dt}E^n_{\mathrm{NF}}=0. \tag{4-38}$$

(C) *Its components satisfy the following estimates*:

$$\begin{aligned} E^n_{\mathrm{NF,high}}&=[1+O(A)]\mathcal{E}_0(\boldsymbol{W}^{(n)},R^{(n)}),\\ E^{n,c}_{\mathrm{NF,high}}&=O(\underline{A})\mathcal{E}_0(\boldsymbol{W}^{(n)},R^{(n)}),\\ E^n_{\mathrm{NF,low}}&=O(\underline{A})(\mathcal{E}_0(\boldsymbol{W}^{(n)},R^{(n)})+c^4\mathcal{E}_0(\boldsymbol{W}^{(n-1)},R^{(n-1)})). \end{aligned} \tag{4-39}$$

Proof. Here we use the algebraic expression for the normal-form transformation, which is given below. We call lower-order terms any terms that can be included in $E^n_{\mathrm{NF,low}}$. Up to quartic terms we seek to have, at least formally, the relation

$$\begin{aligned} E^n_{\mathrm{NF}}&=\|(\widetilde{W}^{(n+1)},\widetilde{Q}^{(n+1)})\|^2_{L^2\times\dot H^{1/2}}+\text{quartic}\\ &=\mathcal{E}_0(W^{(n+1)},Q^{(n+1)})-2\Re\int \overline{W}^{(n+1)}\partial^{n+1}W^{[2]}-i\,\bar Q^{(n+1)}\partial^{n+1}Q^{[2]}\,d\alpha+\text{quartic}. \end{aligned} \tag{4-40}$$

In the first term we introduce the diagonal variable R by using $Q_\alpha=R(1+\boldsymbol{W})$, which allows us to write it as

$$I_{\text{main}}=\mathcal{E}_0(\boldsymbol{W}^{(n)},R^{(n)})+2\Im\int \bar R^{(n+1)}\partial^n(R\boldsymbol{W})\,d\alpha+\text{quartic}.$$

Differentiating we obtain its principal part, containing all terms with leading order above $2n-1$:

$$I_{\text{main,high}}=\mathcal{E}_0(\boldsymbol{W}^{(n)},R^{(n)})+2\Im\int \boldsymbol{W}\bar R^{(n+1)}R^{(n)}+R\bar R^{(n+1)}\boldsymbol{W}^{(n)}-(n+2)R_\alpha\bar R^{(n)}\boldsymbol{W}^{(n)}\,d\alpha.$$

Here the last term on the right no longer appears if $n=1$, which accounts for the similar modification in the proposition.

In the integral in (4-40), on the other hand, we have two concerns. First, we want to eliminate its unbounded part of low frequency, i.e., all terms with inverse derivatives of W, as well as all terms with undifferentiated Q; we remark that once this is done, the switch to diagonal variables is achieved simply by replacing Q_α with R. Secondly, we want to keep track of its highest-order terms, i.e., those terms which are at least energy strength (i.e., $\boldsymbol{W}^{(n)}\overline{\boldsymbol{W}}^{(n)}$, or $\bar R^{(n+1)}R^{(n)}$, and also $\bar R^{(n)}\boldsymbol{W}^{(n)}$ for terms without c). First we recall the expression for the normal form, see (2-8):

$$\begin{aligned} W^{[2]}&=-(W+\overline{W})W_\alpha-\frac{c}{2g}[(Q+\bar Q)W_\alpha+(W+\overline{W})Q_\alpha]\\ &\quad+i\frac{c^2}{2g}\big[(\partial^{-1}W-\partial^{-1}\overline{W})W_\alpha+W^2+\tfrac12|W|^2\big]-\frac{c^2}{4g^2}(Q+\bar Q)Q_\alpha\\ &\quad+i\frac{c^3}{4g^2}[(Q+\bar Q)W+(\partial^{-1}W-\partial^{-1}\overline{W})Q_\alpha]+\frac{c^4}{4g^2}(\partial^{-1}W-\partial^{-1}\overline{W})W, \end{aligned}$$

$$\begin{aligned} Q^{[2]}&=-(W+\overline{W})Q_\alpha-\frac{c}{2g}(Q+\bar Q)Q_\alpha+i\frac{c}{4}(W^2+2|W|^2)\\ &\quad+i\frac{c^2}{2g}\big[(\partial^{-1}W-\partial^{-1}\overline{W})Q_\alpha+\tfrac12(Q+\bar Q)W\big]+\frac{c^3}{4g}(\partial^{-1}W-\partial^{-1}\overline{W})W. \end{aligned}$$

Next, we successively consider the terms in the integral in (4-40) organized by the power of c.

Terms with c^0. These are

$$I_0 = 2\Re \int -\overline{\boldsymbol{W}}^{(n)}\partial^{n+1}[(W+\overline{W})W_\alpha] + i\,\bar{Q}^{(n+2)}\partial^{n+1}[(W+\overline{W})Q_\alpha]\,d\alpha.$$

Here there are no low-frequency issues. Its high-frequency part, on the other hand, has the form

$$\begin{aligned} I_{0,\mathrm{high}} = 2\Re \int &-(2n+2)\Re \boldsymbol{W}(|\boldsymbol{W}^{(n)}|^2 - i\,\bar{Q}^{(n+2)}Q^{(n+1)}) \\ &+ i\,Q_\alpha \bar{Q}^{(n+2)}(W^{(n+1)}+\overline{W}^{(n+1)}) - (n+2)Q_\alpha \bar{Q}^{(n+1)}\boldsymbol{W}^{(n)}\,d\alpha. \end{aligned}$$

Terms with c. These are

$$\begin{aligned} I_1 = c\Re \int &-\overline{\boldsymbol{W}}^{(n)}\partial^{n+1}(\boldsymbol{W}(Q+\bar{Q}) + Q_\alpha(W+\overline{W})) \\ &+ \bar{Q}^{(n+2)}\partial^{n+1}\big(\tfrac{1}{2}W^2+|W|^2\big) + \frac{i}{g}\bar{Q}^{(n+2)}\partial^{n+1}(QQ_\alpha + \bar{Q}Q_\alpha)\,d\alpha. \end{aligned}$$

We first verify that the terms with undifferentiated Q disappear when integrating by parts. These are

$$\begin{aligned} c\Re \int -2\Re Q\,\overline{\boldsymbol{W}}^{(n)}\boldsymbol{W}^{(n+1)}\,d\alpha &= c\int \Re Q_\alpha |W^{(n)}|^2\,d\alpha, \\ \frac{c}{g}\Re \int i\Re Q\,\bar{Q}^{(n+2)}Q^{(n+2)}\,d\alpha &= 0. \end{aligned}$$

Secondly, we compute the high-frequency component. Taking into account the above integration by parts, this is

$$I_{1,\mathrm{high}} = c\int -(2n+2)|\boldsymbol{W}^{(n)}|^2\Re Q_\alpha + \frac{i}{g}(2n+3)\Re Q_\alpha \bar{Q}^{(n+2)}Q^{(n+1)}\,d\alpha.$$

Terms with c^2. These are

$$\begin{aligned} I_2 = c^2\Re \int &i\,\overline{\boldsymbol{W}}^{(n)}\partial^{n+1}\big(\boldsymbol{W}(\partial^{-1}W - \partial^{-1}\overline{W}) + W^2 + \tfrac{1}{2}|W|^2\big) - \frac{1}{2g}\overline{\boldsymbol{W}}^{(n)}\partial^{n+1}(QQ_\alpha + \bar{Q}Q_\alpha) \\ &+ \frac{1}{g}\bar{Q}^{(n+2)}\partial^{n+1}(Q_\alpha(\partial^{-1}W - \partial^{-1}\overline{W})) + \frac{1}{2g}\bar{Q}^{(n+2)}\partial^{n+1}(WQ + W\bar{Q})\,d\alpha. \end{aligned}$$

For the low-frequency analysis we compute

$$c^2\Re \int -i\,\overline{\boldsymbol{W}}^{(n)}\boldsymbol{W}^{(n+1)}(\partial^{-1}W - \partial^{-1}\overline{W})\,d\alpha = -c^2\int |\boldsymbol{W}^{(n)}|^2\Im W\,d\alpha,$$

while the remaining $\partial^{-1}W$-terms and the Q-terms directly cancel.

Now we look at the high-frequency part. This is

$$I_{2,\mathrm{high}} = c^2\int -\big(2n+\tfrac{5}{2}\big)\Im W(|\boldsymbol{W}^{(n)}|^2 - i\,\bar{Q}^{(n+2)}Q^{(n+1)}) - (2n+4)(Q_\alpha+\bar{Q}_\alpha)\overline{\boldsymbol{W}}^{(n)}Q^{(n+1)}\,d\alpha.$$

Terms with c^3. These are

$$I_3 = -\frac{c^3}{2g}\Im \int \overline{\boldsymbol{W}}^{(n)}\partial^{n+1}[W(Q+\bar{Q}) + (\partial^{-1}W - \partial^{-1}\overline{W})Q_\alpha] - \bar{Q}^{(n+2)}\partial^{n+1}[(\partial^{-1}W - \partial^{-1}\overline{W})W]\,d\alpha.$$

For the low-frequency analysis we compute

$$\Re \int i\overline{\boldsymbol{W}}^{(n)}(\partial^{-1}W-\partial^{-1}\overline{W})Q^{(n+2)}-i\,\bar{Q}^{(n+2)}(\partial^{-1}W-\partial^{-1}\overline{W})\boldsymbol{W}^{(n)}\,d\alpha=0,$$

$$\Re \int i\overline{\boldsymbol{W}}^{(n)}\boldsymbol{W}^{(n)}(Q+\bar{Q})\,d\alpha=0,$$

while for the high-frequency analysis, taking the above into account, we are left with

$$I_{3,\mathrm{high}}=\frac{c^3}{2g}\int i(2n+4)(W-\overline{W})\overline{\boldsymbol{W}}^{(n)}Q^{(n+1)}\,d\alpha.$$

Terms with c^4. This is

$$I_4=\frac{c^4}{2g^2}\Re\int \boldsymbol{W}^{(n)}\partial^{(n+1)}[(\partial^{-1}W-\partial^{-1}\overline{W})W]\,d\alpha,$$

which exhibits cancellation at low frequency and has no high-frequency component.

The normal-form energy E^n_{NF} is obtained by replacing Q_α with R in the all the above terms, after eliminating the Q and $\partial^{-1}W$. The expressions of $E^n_{\mathrm{NF,high}}$ and $E^{n,c}_{\mathrm{NF,high}}$ are simply obtained by adding up all the high-frequency contributions above.

Since E^n_{NF} is obtained from the normal form, the relation (4-38) immediately follows. We note that establishing this property is a purely algebraic computation, which does not require the direct use of the normal form (which is unbounded).

Step 2. We begin with the quasilinear energy $E^{n,(3)}_{\mathrm{high}}(w,r)$, which was defined in (4-30) if $n\geq 2$, and is set to be equal to $E^{(3)}_{\mathrm{lin}}(w,r)$ if $n=1$. We first compare it with the cubic energy $E^n_{\mathrm{NF,high}}$. Precisely, we have

$$\Lambda^{\leq 3}E^{(3)}_{\mathrm{lin}}(w,r)=\Lambda^{\leq 3}E^n_{\mathrm{NF,high}},\quad n=1, \tag{4-41}$$

$$\Lambda^{\leq 3}E^{n,(3)}_{\mathrm{high}}(w,r)=\Lambda^{\leq 3}E^n_{\mathrm{NF,high}},\quad n\geq 2. \tag{4-42}$$

This computation was already done in [HIT16] and is not repeated here. The only difference is in the c-term in $E^n_{\mathrm{NF,high}}$, which arises due to the a_1 in $E^{(3)}_{\mathrm{lin}}(w,r)$ and $E^{n,(3)}_{\mathrm{high}}(w,r)$.

Next, we consider the term $E^{n,c}_{\mathrm{NF,high}}$. This contains the leading-order energy density, so we have to treat it in a quasilinear manner. Precisely, we replace it by

$$E^{n,(3)}_{\mathrm{high},c}(w,r)=-\int\left[c(2n+3)\Re R+c^2\left(2n+\tfrac{5}{2}\right)\Im W\right]((g+\underline{a})|w|^2-i\bar{r}_\alpha r)\,d\alpha,$$

which admits a straightforward bound

$$|E^{n,(3)}_{\mathrm{high},c}(w,r)|\lesssim cA_{-1/2}\boldsymbol{N}_n+c^2A_{-1}\boldsymbol{N}_n. \tag{4-43}$$

Now we are in a position to define our quasilinear modified cubic energy

$$E^{n,(3)}(\boldsymbol{W},R)=E^{n,(3)}_{\mathrm{high}}(w,r)+E^{n,(3)}_{\mathrm{high},c}(w,r)+E^n_{\mathrm{NF,low}}(\boldsymbol{W},R). \tag{4-44}$$

By construction this satisfies

$$\Lambda^{<3} E^{n,(3)}(W, R) = \Lambda^{<3} E^n_{\mathrm{NF}}(W, R). \tag{4-45}$$

Step 3. Now we proceed to prove Proposition 4.2. The norm equivalence is already known from (4-31) and (4-39), so we still need the energy estimate. For this we use (4-45) and (4-38) to see that the cubic terms vanish,

$$\Lambda^{\leq 3} \frac{d}{dt} E^{n,(3)} = \Lambda^{\leq 3} \frac{d}{dt} E^n_{\mathrm{NF}} = 0.$$

Hence, it remains to look at the quartic-and-higher terms. For this we can split the terms,

$$\Lambda^{\geq 4} \frac{d}{dt} E^{n,(3)} = \Lambda^{\geq 4} \frac{d}{dt} E^{n,(3)}(w, r) + \Lambda^{\geq 4} \frac{d}{dt} E^{n,(3)}_{\mathrm{high},c}(w, r) + \Lambda^{\geq 4} \frac{d}{dt} E^n_{\mathrm{NF,low}}.$$

The bound for the first term on the right is a direct consequence of Proposition 3.1 for $n = 1$, and Lemma 4.6 for $n \geq 2$. From here on, the cases $n = 1$ and $n \geq 2$ are identical.

For the middle term $\Lambda^{\geq 4} \frac{d}{dt} E^{n,(3)}_{\mathrm{high},c}(w, r)$ we will use Lemma 3.2, with either $f = W$ or $f = R$, and (w, r) as in this section. We claim that this yields

$$\left| \frac{d}{dt} E^{n,(3)}_{\mathrm{high},c}(w, r) \right| \lesssim_{\underline{A}} \underline{B} N_n,$$

which by homogeneity considerations yields

$$\left| \Lambda^{\geq 4} \frac{d}{dt} E^{n,(3)}_{\mathrm{high},c}(w, r) \right| \lesssim_{\underline{A}} \underline{A}\underline{B} N_n,$$

as desired.

(a) If $f = W$ then the bounds

$$\|f\|_{L^\infty} \lesssim_A A_{-1}, \quad \|D^{1/2} f\|_{L^\infty} \lesssim A_{-1/2}$$

allow us to estimate the difference I in the Lemma 3.2. The remaining terms

$$c^2 \int \Lambda^{\geq 2}(\partial_t + \underline{b}\partial_\alpha) W((g + \underline{a})|w|^2 - i\bar{r}_\alpha r)\, d\alpha \quad \text{and} \quad c^2 \int f((g + \underline{a})\bar{w} F - i\bar{r}_\alpha G)\, d\alpha$$

are c-times forms which are at least quartic, with order $2n + \frac{1}{2}$ and leading order $2n$ (this is because we have exactly one R-factor). Hence we can bound them using Hölder's inequality and interpolation by $cA\underline{A}N_n$.

(b) If $f = R$ then the bounds

$$\|f\|_{L^\infty} \lesssim_A A_{-1/2}, \quad \|D^{1/2} f\|_{L^\infty} \lesssim A$$

allow us to estimate the difference D in the lemma. The remaining terms are exactly as above, still bounded by $cA\underline{A}N_n$.

Finally, for the lower-order terms $\Lambda^{\geq 4} \frac{d}{dt} E^n_{\mathrm{NF,low}}$ it suffices to have the following property:

Lemma 4.8. *For any term I in $E^n_{\mathrm{NF,low}}$ we have*

$$\left| \Lambda^{\geq 4} \frac{d}{dt} I \right| \lesssim \underline{A}\underline{B}(N_n + c^4 N_{n-1}). \tag{4-46}$$

Proof. The terms without c in this computation were already estimated in [HIT16], so we need to consider only c-terms. A simple order analysis for these terms leads to the following properties:

(a) Their order is $2n+\frac{1}{2}$.

(b) Their leading order is at most $2n$.

(c) The highest power of c is 4.

Here part (b) is a obtained by using the relation (4-7) to compute time derivatives. This guarantees that the leading order does not increase by a full unit.

We also take a closer look at the c^2W- and cR-factors that can arise. In the trilinear form $E^n_{\mathrm{NF,low}}$ they can arise just once, so the question is what happens if we differentiate with respect to time. If W and R are differentiated in time then at most we obtain another W-factor, respectively an R- or cW-factor, so this pattern is preserved. If instead we differentiate a higher-order $W^{(k)}$- or $R^{(k)}$-factor using the equations (1-19), we can further produce undifferentiated W- and R-factors, as follows:

(i) Arising from cM_1 and its derivatives. There cW will appear in combinations of the form $\boldsymbol{P}[W\bar{Y}^{(k)}]$ and its conjugate. However, the frequency localization enforced by $\boldsymbol{P}$ guarantees that W is the higher-frequency factor, which, for estimates, ensures that we can freely move derivatives from $\bar{Y}$ to W.

(ii) Arising from derivatives of cN, where, for the same reasons as above, we can disregard the W-factors, but we can still produce a cR-factor.

(iii) Arising from b_1 and its derivatives. In the case of derivatives of b_1 the discussion concerning W-factors is again identical to case (i), so we can neglect those. Hence the only potential W-factor can arise from undifferentiated b_1. However, these are avoided due to the use of (4-7), where the transport is fully included in the time differentiation and the undifferentiated advection coefficient $\underline{b}$ never appears.

To summarize the above discussion, in the time derivative of $E^n_{\mathrm{NF,low}}$ we can have at most one c^2W-factor and at most two c^2W, cR-factors. A further simplification arises from the constraint (b) above. Precisely, there we can integrate by parts to rebalance derivatives in such a way that either

(a) both leading order terms have order at most n, or

(b) we have exactly the product $R^{(n+1)}\bar{R}^{(n)}$ as part of our multilinear expression.

In case (a), our estimates follow directly by Hölder's inequality and interpolation. In case (b), the remaining factors must have order zero, except for one which has order $\frac{1}{2}$. There we consider two scenarios.

(b1) If the factors $R^{(n+1)}$ and $\bar{R}^{(n)}$ are not frequency balanced, then another factor has the highest frequency and we can rebalance derivatives and use again Hölder's inequality and interpolation.

(b2) If the factors $R^{(n+1)}$ and $\bar{R}^{(n)}$ are frequency balanced, then all we need is to bound the remaining factors in L^∞. The only factor of order $\frac{1}{2}$ which we do not control in L^∞ is R_α, so Hölder's inequality and interpolation work unless our multilinear expression exactly contains

$$R^{(n+1)}\bar{R}^{(n)}R_\alpha,$$

and possibly other zero-order factors. Backtracking, the only way to produce such a term is by differentiating a cubic expression which has a leading order exactly $2n-\frac{1}{2}$ as well as an R_α-factor. Then this expression must be exactly

$$\boldsymbol{W}^{(n)}\bar{R}^{(n)}R_\alpha,$$

which cannot contain any c-factors so it is not within our purview here. □

Thus we have completed the proof of Proposition 4.7. □

5. Proofs of the main theorems

The results in Theorems 1 and 2 are a consequence of the estimates for the linearized equation in Section 3, and of the energy estimates for the higher Sobolev norms of the solutions in Section 4. The arguments here closely match their counterparts in our previous gravity waves paper [HIT16]. Because of this, we only provide an outline here, and refer the reader to [HIT16] for more details.

Outline of proof of Theorem 1. The steps of the proof are as follows:

Step 1: Existence of regular solutions. A standard approach here is to obtain solutions as limits of solutions to frequency truncated equations. As always, this truncation needs to happen in a symmetric way, so that uniform energy estimates survive. In addition, a specific difficulty we encounter in water wave evolutions is the fact that the evolution we are trying to truncate is degenerate hyperbolic, a condition which might not survive in a naive direct truncation procedure. In [HIT16] we bypass this difficulty by solving directly the differentiated system for the diagonal variables $(\boldsymbol{W}, R)$. In our case this is not entirely possible, as the system for $(\boldsymbol{W}, R)$ is not fully self-contained. Precisely, it contains W as a part of $\underline{b}$. Fortunately, this does not cause extra difficulties, as we can make the frequency truncation consistent with differentiation in the W-equation, and thus be able to freely use W in the $(\boldsymbol{W}, R)$-system. We note that formally $\boldsymbol{W}$ also appears undifferentiated in M_1 and N, but the commutator structure of those expressions implies that they actually can be rewritten (and estimated) in terms of $\boldsymbol{W}$ instead.

Precisely, our main truncated system for $(\boldsymbol{W}, R)$ has the form

$$\begin{cases} \boldsymbol{W}_t+\boldsymbol{P}_{<n}\underline{b}_n\boldsymbol{P}_{<n}\boldsymbol{W}_\alpha+\boldsymbol{P}_{<n}\dfrac{(1+\boldsymbol{P}_{<n}\boldsymbol{W})\boldsymbol{P}_{<n}R_\alpha}{1+\overline{\boldsymbol{P}_{<n}\boldsymbol{W}}}=\boldsymbol{P}_{<n}G(\boldsymbol{W}_{<n}, R_{<n}), \\ R_t+\boldsymbol{P}_{<n}\underline{b}_n\boldsymbol{P}_{<n}R_\alpha+ic\boldsymbol{P}_{<n}R-i\boldsymbol{P}_{<n}\dfrac{g\boldsymbol{P}_{<n}\boldsymbol{W}-a_n}{1+\boldsymbol{P}_{<n}\boldsymbol{W}}=\boldsymbol{P}_{<n}K(\boldsymbol{W}_{<n}, R_{<n}), \end{cases} \tag{5-1}$$

where G and K represent the right-hand-side terms in (1-19), and

$$\underline{b}_n=\underline{b}(W_{<n}, R_{<n}), \quad a_n=a(R_{<n}).$$

Here n is a dyadic-frequency parameter, and the multiplier $\boldsymbol{P}_{<n}$ selects the frequencies less than 2^n. The expression for $\boldsymbol{W}_t$ in the first equation above retains the structure from the original equations, i.e., it is an exact derivative. Thus we can recover the undifferentiated variable W in a way consistent with the $\boldsymbol{W}$-equation above, simply by integrating in time the relation

$$W_t+\boldsymbol{P}_{<n}[\underline{F_n}(1+\boldsymbol{P}_{<n}\boldsymbol{W})]+i\frac{c}{2}\boldsymbol{P}_{<n}W=0.$$

The above set of equations is consistent, and for each dyadic frequency scale n it generates an ODE in the space $\mathcal{H}_k$, $k \geq 2$ for $(\boldsymbol{W}, R)$, with the additional property that W in L^2. A priori, the lifespan for this system depends on n. However, the same type of estimates used for the linearized system yield uniform bounds and a uniform lifespan depending only on the Sobolev norm of the data, and not on n. Then the regular solutions are obtained as weak limits of these approximate solutions as $n \to \infty$.

<u>Step 2</u>: Uniqueness of regular solutions ($(\boldsymbol{W}, R) \in \dot{\mathcal{H}}_2$). This is achieved in a more standard fashion. Given two solutions (W_1, Q_1) and (W_2, Q_2) we subtract the equations and do energy estimates for the difference $(\boldsymbol{W}_1 - \boldsymbol{W}_2, R_1 - R_2)$ in $\dot{\mathcal{H}}_0$, as well as simpler integrated bounds for $W_1 - W_2$ in L^2. Then close the argument and prove uniqueness via Gronwall's inequality.

<u>Step 3</u>: $\dot{\mathcal{H}}_1$-bounds for $(\boldsymbol{W}, R)$. Here we use the uniform bounds for the $\dot{\mathcal{H}}_2$-norm of $(\boldsymbol{W}, R)$ which depend only on the $\dot{\mathcal{H}}_1$-norm of $(\boldsymbol{W}, R)$ (via the control parameters $\underline{A}$ and $\underline{B}$), as in Proposition 4.1, to conclude that the regular solutions can be continued up to a time which depends only on the $\dot{\mathcal{H}}_1$-norm of the data.

<u>Step 4</u>: Construction of rough solutions, $(\boldsymbol{W}, R) \in \dot{\mathcal{H}}_1$. We regularize the data, $(W_{<k}(0), Q_{<k}(0))$, by truncating at frequencies $< 2^k$. The corresponding solutions will be regular, with a uniform lifespan bound. Thinking of k as a continuous parameter, we associate the k-derivative of the solutions $(W_{<k}, Q_{<k})$ to solutions for the linearized equation around $(W_{<k}, Q_{<k})$. Using both the energy bounds for the solutions $(\boldsymbol{W}_{<k}, R_{<k})$ in $\dot{\mathcal{H}}_0$ and $\dot{\mathcal{H}}_1$, and the $\dot{\mathcal{H}}_0$-bounds for the linearized equation, we show that these solutions inherit not only uniform bounds in $\dot{\mathcal{H}}_1$, but also uniform frequency envelope bounds in terms of the initial data frequency envelope. This suffices in order to prove strong convergence of $(\boldsymbol{W}_{<k}, R_{<k})$ in $\dot{\mathcal{H}}_1$, and of $(W_{<k}, Q_{<k})$ in $\dot{\mathcal{H}}_0$.

<u>Step 5</u>: Weak Lipschitz dependence for rough solutions. Here we show that for rough solutions, i.e., $(W, Q) \in \dot{\mathcal{H}}_0$ and $(\boldsymbol{W}, R) \in \dot{\mathcal{H}}_1$, we have Lipschitz dependence on the initial data in the $\dot{\mathcal{H}}_0$ topology for both $(W, Q) \in \dot{\mathcal{H}}_0$ and $(\boldsymbol{W}, R)$. This is a direct consequence of the $\dot{\mathcal{H}}_0$-bounds for the linearized equation.

<u>Step 6</u>: Continuous dependence on the data for rough solutions. This is a standard consequence of the frequency envelope bounds in Step 4 and the weak Lipschitz bounds in Step 5.

<u>Step 7</u>: Continuation of solutions. Here we show that the solutions extend with uniform bounds for as long as our control norms $\underline{A}$, $\underline{B}$ remain bounded. This is a consequence of the above local well-posedness in $\dot{\mathcal{H}}_1$ and the energy estimates in Proposition 4.1. □

Outline of proof of Theorem 2. This is an easy consequence of the $\dot{\mathcal{H}}_1$ well-posedness and the $\dot{\mathcal{H}}_1$ uniform bounds in Proposition 4.2, as the control norms $\underline{A}$ and $\underline{B}$ can be estimated in terms if the $\dot{\mathcal{H}}_1$-norm of $(\boldsymbol{W}, R)$ and the $\dot{\mathcal{H}}_0$-norm of (W, Q). □

Appendix A: Water-waves-related estimates

In this section we gather a number of bilinear and multilinear estimates which are used throughout the paper. Some are from [HIT16] and are just recalled here. The rest are connected to the new structure induced by the vorticity. We begin with some commutator bounds from [HIT16].

Lemma A.1. *The following commutator estimates hold*:

$$\||D|^s[\boldsymbol{P}, R]|D|^\sigma w\|_{L^2} \lesssim \||D|^{\sigma+s} R\|_{\mathrm{BMO}} \|w\|_{L^2}, \quad \sigma \geq 0,\ s \geq 0, \tag{A-1}$$

$$\||D|^s[\boldsymbol{P}, R]|D|^\sigma w\|_{L^2} \lesssim \||D|^{\sigma+s} R\|_{L^2} \|w\|_{\mathrm{BMO}}, \quad \sigma > 0,\ s \geq 0. \tag{A-2}$$

We remark that later this is applied to functions which are holomorphic/antiholomorphic, but that no such assumption is made above. Next, we have several bilinear estimates:

Lemma A.2. *The functions N and M_1 satisfy the pointwise bounds*

$$\|N\|_{L^\infty} \lesssim_A AA_{-1/2}, \quad \|M_1\|_{L^\infty} \lesssim_A A^2, \tag{A-3}$$

as well as the Sobolev bounds

$$\|N\|_{\dot H^{n+1/2}} \lesssim_A A\boldsymbol{N}_n, \quad \|M_1\|_{\dot H^n} \lesssim_A A\boldsymbol{N}_n.$$

Proof. We begin with the bounds for N, where N was defined in (1-13) as

$$N = \boldsymbol{P}[W\bar R_\alpha - \overline{\boldsymbol{W}} R] + \bar{\boldsymbol{P}}[\bar W R_\alpha - \boldsymbol{W}\bar R].$$

For the pointwise bound we claim that

$$\|N\|_{L^\infty} \lesssim \|W\|_{B^{3/4,\infty}_2} \|R\|_{B^{1/4,\infty}_2}. \tag{A-4}$$

This suffices since each of the right-hand-side factors is bounded by $\sqrt{AA_{-1/2}}$ by interpolation. To achieve this we observe that we can alternatively write N in the form

$$N = \bar{\boldsymbol{P}}[\bar W R_\alpha - \boldsymbol{W}\bar R] + \boldsymbol{P}[W\bar R_\alpha - \overline{\boldsymbol{W}} R] = \partial_\alpha(\bar{\boldsymbol{P}}[\bar W R] + \boldsymbol{P}[W\bar R]) - (\overline{\boldsymbol{W}} R + \boldsymbol{W}\bar R).$$

We apply a bilinear Littlewood–Paley decomposition and use the first expression above for the high-low interactions, and the second for the high-high interactions, to write $N = N_1 + N_2$, where

$$N_1 = \sum_k [\bar W_k R_{<k,\alpha} - W_{<k,\alpha}\bar R_k] + [W_k \bar R_{<k,\alpha} - \bar W_{<k,\alpha} R_k],$$

$$N_2 = \sum_k \partial_\alpha(\bar{\boldsymbol{P}}[\bar W_k R_k] + \boldsymbol{P}[W_k\bar R_k]) - (\bar W_{k,\alpha} R_k + W_{k,\alpha}\bar R_k).$$

We estimate the terms in N_1 separately; we show the argument for the first term:

$$\Big\|\sum_k \bar W_k R_{<k,\alpha}\Big\|_{L^\infty} \lesssim \sum_{j\leq k} 2^{(3/4)(j-k)} \|W_j\|_{L^\infty} \|R_k\|_{L^\infty} \lesssim \|W\|_{B^{3/4,\infty}_2} \|R\|_{B^{1/4,\infty}_2}.$$

For the first term in N_2 we note that the multiplier $\partial_\alpha \boldsymbol{P}_{<k+4}\boldsymbol{P}$ has an $O(2^k)$ L^∞-bound. Hence, we can estimate

$$\|N_2\|_{L^\infty} \lesssim \sum_k 2^k \|W_k\|_{L^\infty} \|R_k\|_{L^\infty} \lesssim \|W\|_{B^{3/4,\infty}_2} \|R\|_{B^{1/4,\infty}_2}.$$

For simplicity, we only prove the $\dot{H}^{1/2}$-bound for N. The rest can be done in a very similar way. To obtain the bound we apply a Coifman–Meyer-type commutator estimate:

$$\|N\|_{\dot{H}^{1/2}} \lesssim \||D|^{1/2}[\bar{\boldsymbol{P}}, \bar{W}]R_\alpha\|_{L^2} + \||D|^{1/2}[\bar{\boldsymbol{P}}, \bar{R}]\boldsymbol{W}\|_{L^2} + \||D|^{1/2}[\boldsymbol{P}, W]\bar{R}_\alpha\|_{L^2} + \||D|^{1/2}[\boldsymbol{P}, \boldsymbol{W}]R\|_{L^2}.$$

Each commutator can be bounded by $\lesssim AN_0$. For pointwise and Sobolev bounds of M_1 we refer to Lemma 2.8 (see the Appendix in [HIT16]); the exact same approach applies in the current case. □

Essential in the article are the bounds for a, which we established in [HIT16]:

Proposition A.3. *Assume that $R \in \dot{H}^{1/2} \cap \dot{H}^{3/2}$. Then the real frequency-shift a is nonnegative and satisfies the* BMO *bound*

$$\|a\|_{\mathrm{BMO}} \lesssim \|R\|^2_{\mathrm{BMO}^{1/2}}, \tag{A-5}$$

and the uniform bound

$$\|a\|_{L^\infty} \lesssim \|R\|^2_{\dot{B}^{1/2}_{\infty,2}}. \tag{A-6}$$

Moreover,

$$\||D|^{1/2}a\|_{\mathrm{BMO}} \lesssim \|R_\alpha\|_{\mathrm{BMO}}\, \||D|^{1/2}R\|_{L^\infty}, \qquad \|a\|_{B_2^{1/2,\infty}} \lesssim \|R_\alpha\|_{B_2^{1/2,\infty}}\, \||D|^{1/2}R\|_{L^\infty}, \tag{A-7}$$

$$\|(\partial_t + b\partial_\alpha)a\|_{L^\infty} \lesssim AB, \tag{A-8}$$

$$\|a\|_{\dot{H}^s} \lesssim \|R\|_{\dot{H}^{s+1/2}}\, \|R\|_{\mathrm{BMO}^{1/2}}, \quad s > 0. \tag{A-9}$$

Here we need to supplement this with bounds for a_1. One notable difference between the two is that a_1 has a linear component, whereas a is purely quadratic. For various estimates we need to separate the two components of a_1:

Proposition A.4. *Assume that $R \in \dot{H}^{1/2} \cap \dot{H}^{3/2} \cap L^\infty$. Then*

$$\|a_1\|_{L^\infty} \lesssim_A A_{-1/2}(1 + A).$$

Moreover, *the following estimate holds*:

$$\|(\partial_t + \underline{b}\partial_\alpha)a_1 + 2g\Im \boldsymbol{W} - 2c\Im R\|_{L^\infty} \lesssim AB + cA^2.$$

Proof. We first recall the expression for a_1:

$$a_1 = R + \bar{R} - N.$$

Using the equation for R, we only need to prove the pointwise bound for $(\partial_t + \underline{b}\partial_\alpha)N$. We begin with the following computation:

$$(\partial_t + \underline{b}\partial_\alpha)a_1 + 2g\Im \boldsymbol{W} - 2c\Im R = -2a\Im Y - c\Im \frac{R\boldsymbol{W} + \bar{R}\boldsymbol{W} + N}{1 + \boldsymbol{W}} - (\partial_t + \underline{b}\partial_\alpha)N.$$

Based on the previously established pointwise bounds for a and N we can estimate all but the last term. This is considerably more delicate. However, as seen in the proof of Lemma A.2, the expression for N exhibits exactly the same cancellation structure for "high × high → low" interactions as we have in the bilinear expression for a. Hence, the argument in the proof of (A-8) in [HIT16] immediately adapts here. □

Appendix B: Holomorphic coordinates

Our goal here is to introduce the holomorphic coordinates and the holomorphic set of variables (W, Q) describing the free surface, respectively the holomorphic velocity potential restricted to the free surface, and to derive the evolution equations for (W, Q).

We proceed as in [Hunter, Ifrim and Tataru 2016; Ifrim and Tataru 2017]. Let $\mathbb{H}$ be the lower half-space, with complex coordinates denoted by $\alpha + i\beta$. Let $\mathcal{F} : \mathbb{H} \to \Omega(t)$ to be the conformal transformation that maps the α-axis into $\Gamma(t)$, which is unique up to α-translations. The x- and y-coordinates are given by

$$x = x(\alpha, \beta, t), \qquad y = y(\alpha, \beta, t),$$

and satisfy the Cauchy–Riemann equations

$$x_\alpha = y_\beta, \qquad x_\beta = -y_\alpha.$$

We fix the conformal map by assuming that

$$x(\alpha, \beta, t) + iy(\alpha, \beta, t) - \alpha + i\beta \to 0, \qquad \alpha, \beta \to \infty.$$

To switch between the two sets of coordinates we will use

$$\begin{aligned} \partial_\alpha &= x_\alpha \partial_x + y_\alpha \partial_y, & \partial_\beta &= x_\beta \partial_x + y_\beta \partial_y, \\ \partial_x &= \frac{1}{j}(x_\alpha \partial_\alpha + x_\beta \partial_\beta), & \partial_y &= \frac{1}{j}(y_\alpha \partial_\alpha + y_\beta \partial_\beta), \end{aligned}$$

where

$$j = x_\alpha y_\beta - x_\beta y_\alpha = x_\alpha^2 + x_\beta^2 = y_\alpha^2 + y_\beta^2.$$

We also need a similar relation for the time derivative. Assume we have

$$\mathbb{H} \xrightarrow{\mathcal{F}} \Omega(t) \xrightarrow{f} \mathbb{C}$$

for an arbitrary function f. Let

$$g(\alpha, \beta, t) = f(x(\alpha, \beta, t), y(\alpha, \beta, t), t).$$

Then

$$g_t = f_t + x_t f_x + f_y y_t.$$

So

$$\begin{aligned} f_t &= g_t - x_t f_x - f_y y_t \\ &= g_t - x_t \frac{1}{j}(x_\alpha g_\alpha + x_\beta g_\beta) - y_t \frac{1}{j}(y_\alpha g_\alpha + y_\beta g_\beta) \\ &= g_t - \frac{1}{j}(x_t x_\alpha + y_t y_\alpha) g_\alpha - \frac{1}{j}(x_t x_\beta + y_t y_\beta) g_\beta. \end{aligned} \tag{B-1}$$

Define ψ to be the (harmonic) composition of $\mathcal{F}$ to the velocity potential φ,

$$\psi = \varphi \circ \mathcal{F} : \mathbb{H} \to \mathbb{R}, \qquad \psi(\alpha, \beta, t) = \varphi(x(\alpha, \beta, t), y(\alpha, \beta, t), t).$$

The velocity components (u, v) can now be expressed in terms of the velocity potential ψ by

$$u = \frac{1}{j}(x_\alpha \psi_\alpha + x_\beta \psi_\beta) + cy, \quad v = \frac{1}{j}(y_\alpha \psi_\alpha + y_\beta \psi_\beta). \tag{B-2}$$

It follows that

$$u^2 + v^2 = \varphi_x^2 + \varphi_y^2 = \frac{1}{j}(\psi_\alpha^2 + \psi_{\beta^2}).$$

Boundary values. Setting $\beta = 0$ gives the boundary values of the holomorphic functions defined in the lower half-plane. In particular, we introduce the notation

$$x(\alpha, 0, t) =: X(\alpha, t), \quad y(\alpha, 0, t) =: Y(\alpha, t),$$

so that $\alpha \to X + iY$ parametrizes $\Gamma(t)$. The function $(x - \alpha) + i(y - \beta)$ is holomorphic in $\mathbb{H}$ and decays at infinity, which implies that on the boundary we have

$$\begin{cases} Y = H(X - \alpha), \\ X = \alpha + HY. \end{cases}$$

Also set

$$z(\alpha, \beta, t) = x(\alpha, \beta, t) + iy(\alpha, \beta, t), \quad Z(\alpha, t) = z(\alpha, 0, t).$$

Then our "holomorphic" variable W, which describes the surface $\Gamma(t)$, will be

$$W = Z - \alpha.$$

As z is holomorphic in $\mathbb{H}$, so is $1/z_\alpha - 1$; further, it decays at infinity. Its boundary values on the real axis are given by

$$\frac{1}{Z_\alpha - 1} = \left(\frac{X_\alpha}{J} - 1\right) - i\frac{Y_\alpha}{J},$$

which leads to the relations

$$\begin{cases} \dfrac{X_\alpha}{J} - 1 = -H\left[\dfrac{Y_\alpha}{J}\right], \\ \dfrac{Y_\alpha}{J} = H\left[\dfrac{X_\alpha}{J} - 1\right]. \end{cases} \tag{B-3}$$

We also introduce the notation $\Psi(\alpha, t) := \psi(\alpha, \beta, t)$ for the real velocity potential restricted to $\Gamma(t)$ and expressed in holomorphic coordinates, and at the same time define $\Theta(\alpha, t)$ by

$$\begin{cases} \Psi = H\Theta, \\ \Theta = -H\Psi. \end{cases}$$

Up to a constant, this is the trace of the stream function θ on the free boundary. Since ψ is harmonic in the lower half-plane, we have

$$\psi_\beta|_{\beta=0} - H\Psi_\alpha = -\Theta_\alpha.$$

Our holomorphic velocity potential will be the function

$$Q = \Psi + i\Theta.$$

Further we need to focus on the two boundary conditions: kinematic and dynamic.

The kinematic boundary condition. The kinematic boundary condition states that the normal component of the velocity of the boundary is equal to the normal component of the fluid velocity, meaning that

$$(X_t, Y_t)\cdot(-Y_\alpha, X_\alpha) = (u, v)\cdot(-Y_\alpha, X_\alpha),$$

where $(-Y_\alpha, X_\alpha)$ is a normal to $\Gamma(t)$. Expanding the expression above and using (B-2) we can re-express the kinematic boundary condition in holomorphic coordinates:

$$X_\alpha Y_t - Y_\alpha X_t = H\Psi_\alpha - cYY_\alpha = -\Theta_\alpha - cYY_\alpha. \tag{B-4}$$

The goal now is to obtain a second equations for X_t and Y_t, and then solve for an explicit form of those boundary values. Divide (B-4) by J,

$$\left(\frac{X_\alpha}{J} - 1\right)Y_t - \frac{Y_\alpha}{J}X_t + Y_t = -\frac{\Theta_\alpha}{J} - c\frac{YY_\alpha}{J},$$

and use (B-3) to obtain

$$-H\left[\frac{Y_\alpha}{J}\right]Y_t - \frac{Y_\alpha}{J}H[Y_t] + Y_t = -\frac{\Theta_\alpha}{J} - c\frac{YY_\alpha}{J},$$

which further simplifies to

$$\frac{Y_\alpha}{J}Y_t - H\left[\frac{Y_\alpha}{J}\right]H[Y_t] + X_t = -H\left[\frac{\Theta_\alpha}{J}\right] - cH\left[\frac{YY_\alpha}{J}\right].$$

Thus, a second equation for (X_t, Y_t) is

$$X_\alpha X_t + Y_\alpha Y_t = -JH\left[\frac{\Theta_\alpha}{J}\right] - cJH\left[\frac{YY_\alpha}{J}\right]. \tag{B-5}$$

From (B-4) and (B-5) we have

$$\begin{cases} X_t = -H\left[\frac{\Theta_\alpha}{J}\right]X_\alpha - cH\left[\frac{YY_\alpha}{J}\right]X_\alpha + \frac{\Theta_\alpha}{J}Y_\alpha + c\frac{YY_\alpha}{J}Y_\alpha, \\ Y_t = -H\left[\frac{\Theta_\alpha}{J}\right]Y_\alpha - cH\left[\frac{YY_\alpha}{J}\right]Y_\alpha - \frac{\Theta_\alpha}{J}X_\alpha - c\frac{YY_\alpha}{J}X_\alpha. \end{cases} \tag{B-6}$$

The dynamic boundary condition. We have already determined the spatial form of the dynamic boundary condition in (1-4). From (B-1) and the kinematic boundary conditions (B-4)–(B-5) it follows that, on the boundary, φ_t is given as

$$\varphi_t|_{\beta=0} = \Psi_t - \frac{1}{J}(X_\alpha X_t + Y_\alpha Y_t)\Psi_\alpha + \frac{1}{J}(X_\alpha Y_t - Y_\alpha X_t)\Theta_\alpha.$$

Substituting X_t and Y_t from (B-6) yields

$$\varphi_t|_{\beta=0} = \Psi_t + H\left[\frac{\Theta_\alpha}{J}\right]\Psi_\alpha + cH\left[\frac{YY_\alpha}{J}\right]\Psi_\alpha - \frac{1}{J}\Theta_\alpha^2 - \frac{1}{J}cYY_\alpha\Theta_\alpha. \tag{B-7}$$

We express all the terms on the right in (1-4) in terms of the traces on the boundary of the corresponding functions. Doing so, after some simplifications we arrive at the following equation:

$$\Psi_t + H\left[\frac{\Theta_\alpha}{J}\right]\Psi_\alpha - \frac{\Theta_\alpha^2}{J} + \frac{1}{2J}(\Psi_\alpha^2 + \Theta_\alpha^2) + gY + cH\left[\frac{YY_\alpha}{J}\right]\Psi_\alpha - c\Theta + c\frac{Y}{J}X_\alpha\Psi_\alpha = 0. \tag{B-8}$$

The real form of the equations. The equations (B-6) and (B-8) provide us with a system describing the evolution of the free boundary and the velocity potential restricted to the free boundary, as follows:

$$\begin{cases} Y_t = -H\left[\dfrac{\Theta_\alpha}{J}\right]Y_\alpha - cH\left[\dfrac{YY_\alpha}{J}\right]Y_\alpha - \dfrac{\Theta_\alpha}{J}X_\alpha - c\dfrac{YY_\alpha}{J}X_\alpha, \\ \Psi_t = -H\left[\dfrac{\Theta_\alpha}{J}\right]\Psi_\alpha + \dfrac{\Theta_\alpha^2}{J} - \dfrac{1}{2J}(\Psi_\alpha^2 + \Theta_\alpha^2) - gY - cH\left[\dfrac{YY_\alpha}{J}\right]\Psi_\alpha + c\Theta - c\dfrac{Y}{J}X_\alpha\Psi_\alpha. \end{cases}$$

Here X and Θ are dependent variables.

The Hamiltonian associated to the system is

$$\mathcal{E}(Y, \Psi) = \frac{1}{2}\int \left\{ \Psi|\partial_\alpha|\Psi + gY^2X_\alpha + c\Psi_\alpha Y^2 + \frac{1}{3}c^2Y^3X_\alpha \right\} d\alpha,$$

where $|\partial_\alpha| = H\partial_\alpha$. Thus

$$\begin{cases} \dfrac{\delta\mathcal{E}}{\delta Y} = gYX_\alpha + g\tfrac{1}{2}|\partial_\alpha|(Y^2) + c\Psi_\alpha Y + \tfrac{1}{2}c^2Y^2X_\alpha + \tfrac{1}{6}c^2|\partial_\alpha|(Y^3), \\ \dfrac{\delta\mathcal{E}}{\delta\Psi} = |\partial_\alpha|\Psi - cYY_\alpha. \end{cases}$$

We can write the above equations for (Y, Ψ) in a skew-symmetric form

$$\begin{pmatrix} Y \\ \Psi \end{pmatrix}_t = \begin{pmatrix} 0 & \boldsymbol{A} \\ -\boldsymbol{A}^* & \boldsymbol{B} \end{pmatrix}\begin{pmatrix} \delta\mathcal{E}/\delta Y \\ \delta\mathcal{E}/\delta\Psi \end{pmatrix}, \tag{B-9}$$

where

$$\boldsymbol{A} = \frac{X_\alpha}{J} + Y_\alpha H\frac{1}{J}, \qquad \boldsymbol{B} = \Psi_\alpha H\frac{1}{J} + \frac{1}{J}H\Psi_\alpha - c\partial_\alpha^{-1},$$
$$\boldsymbol{A}^* = \frac{X_\alpha}{J} - \frac{1}{J}HY_\alpha, \qquad \boldsymbol{B}^* = -\boldsymbol{B}.$$

This form immediately implies that $\mathcal{E}$ is conserved along the flow. There are several Hamiltonian symmetries, which correspond to conservation laws of the system in accordance with Noether's principle. The horizontal translation invariance

$$Y(\alpha, t) \to Y(\alpha + a, t), \qquad \Psi(\alpha, t) \to \Psi(\alpha + a, t)$$

is generated by the functional $\mathcal{P}$ which we will derive below. This functional corresponds to total horizontal momentum, and it is given by the system

$$\begin{pmatrix} Y \\ \Psi \end{pmatrix}_\alpha = \begin{pmatrix} 0 & \boldsymbol{A} \\ -\boldsymbol{A}^* & \boldsymbol{B} \end{pmatrix}\begin{pmatrix} \delta\mathcal{P}/\delta Y \\ \delta\mathcal{P}/\delta\Psi \end{pmatrix}, \tag{B-10}$$

where

$$\mathcal{P}(Y,\Psi)=\int\left\{\Psi Y_\alpha-\frac{c}{2}Y^2X_\alpha\right\}d\alpha.$$

The variational derivatives of $\mathcal{P}$ are

$$\begin{aligned}\delta\mathcal{P}/\delta Y&=-\Psi_\alpha-cYX_\alpha-\frac{c}{2}|\partial_\alpha|(Y^2),\\ \delta\mathcal{P}/\delta\Psi&=Y_\alpha.\end{aligned}$$

Thus, our conserved energies are

$$\begin{aligned}\mathcal{E}(Y,\Psi)&=\frac{1}{2}\int\left\{\Psi|\partial_\alpha|\Psi+gY^2X_\alpha+c\Psi_\alpha Y^2+\frac{c^2}{3}Y^3X_\alpha\right\}d\alpha,\\ \mathcal{P}(Y,\Psi)&=\int\left\{\Psi Y_\alpha-\frac{c}{2}Y^2X_\alpha\right\}d\alpha,\end{aligned}\tag{B-11}$$

where $|\partial_\alpha|=H\partial_\alpha$.

The complex form of the equations. Recall that our holomorphic variables are

$$Z=X+iY,\qquad Q=\Psi+i\Theta.$$

We also introduce the notation

$$F=H\left[\frac{\Theta_\alpha}{J}\right]+i\frac{\Theta_\alpha}{J}=\boldsymbol{P}\left[\frac{Q_\alpha-\bar{Q}_\alpha}{J}\right],\qquad J=|Z_\alpha|^2,\tag{B-12}$$

noting that F is also the boundary value of a function which is holomorphic in the lower half-plane.

Using these notations, a straightforward computation allows us to re-express the kinematic boundary conditions (B-6) in a single equation for the motion of the boundary

$$Z_t+FZ_\alpha+2ic\boldsymbol{P}\left[\frac{YY_\alpha}{J}\right]Z_\alpha=0,$$

which is equivalent to

$$Z_t+FZ_\alpha-i\frac{c}{4}\boldsymbol{P}\left[\frac{\partial_\alpha(Z-\bar{Z})^2}{J}\right]Z_\alpha=0.\tag{B-13}$$

Next we rewrite the dynamic boundary condition. For this we apply the operator $2\boldsymbol{P}=\boldsymbol{I}-iH$ in the equation (B-8). We do not repeat the computations in [HIT16] for the case $c=0$. Instead, we use them directly to obtain, from (B-8), the equation

$$Q_t-ig(Z-\alpha)+FQ_\alpha+icQ+\boldsymbol{P}\left[\frac{|Q_\alpha|^2}{J}\right]+cK=0,\tag{B-14}$$

where

$$K=(\boldsymbol{I}-iH)\left[H\left[\frac{YY_\alpha}{J}\right]\Psi_\alpha+\frac{Y}{J}X_\alpha\Psi_\alpha\right].$$

To simplify K we use the relation $H(fg - HfHg) = fHg + Hfg$ to rewrite it as

$$\begin{aligned}
K &= H\left[\frac{YY_\alpha}{J}\right]\Psi_\alpha - iH\left[\frac{YY_\alpha}{J}\Theta_\alpha\right] + i\frac{YY_\alpha}{J}\Psi_\alpha + iH\left[\frac{YY_\alpha}{J}\right]\Theta_\alpha + (\boldsymbol{I} - iH)\left[\frac{Y}{J}X_\alpha\Psi_\alpha\right] \\
&= i(\boldsymbol{I} - iH)\left[\frac{YY_\alpha}{J}\right]Q_\alpha + (\boldsymbol{I} - iH)\left[\frac{Y}{J}(X_\alpha\Psi_\alpha + Y_\alpha\Theta_\alpha)\right] \\
&= \boldsymbol{P}\left[\frac{Y}{\bar{Z}_\alpha} - \frac{Y}{Z_\alpha}\right]Q_\alpha + \boldsymbol{P}\left[\frac{YQ_\alpha}{Z_\alpha} + \frac{Y\bar{Q}_\alpha}{\bar{Z}_\alpha}\right].
\end{aligned}$$

Further, we write $Y = -i((Z-\alpha) - (\bar{Z} - \alpha))$, eliminate the projected antiholomorphic terms and remove the projection in front of holomorphic terms to obtain

$$K = -\frac{i}{2}\boldsymbol{P}\left[\frac{Z-\alpha}{\bar{Z}_\alpha} + \frac{\bar{Z}-\alpha}{Z_\alpha}\right]Q_\alpha - \frac{i}{2}\boldsymbol{P}\left[\frac{(Z-\alpha)\bar{Q}_\alpha}{\bar{Z}_\alpha} - \frac{(\bar{Z}-\alpha)Q_\alpha}{Z_\alpha}\right]. \tag{B-15}$$

Thus our set of holomorphic equations consists of (B-13) and (B-14)–(B-15).

For the last step we use the relation $Z = W + \alpha$ to replace Z by W. For the last expression in (B-13) we have

$$\begin{aligned}
\boldsymbol{P}\left[\frac{\partial_\alpha(Z-\bar{Z})^2}{J}\right] &= 2\boldsymbol{P}\left[\frac{Z-\alpha}{\bar{Z}_\alpha} - \frac{Z-\alpha}{Z_\alpha} + \frac{\bar{Z}-\alpha}{Z_\alpha}\right] \\
&= 2\boldsymbol{P}\left[\frac{W}{1+\bar{W}_\alpha} - \frac{W}{1+W_\alpha} + \frac{\bar{W}}{1+W_\alpha}\right] \\
&= 2F_1 - \frac{2W}{1+W_\alpha},
\end{aligned}$$

where

$$F_1 = \boldsymbol{P}\left[\frac{W}{1+\bar{W}_\alpha} + \frac{\bar{W}}{1+W_\alpha}\right].$$

For K on the other hand we have

$$K = -\frac{i}{2}F_1Q_\alpha - \frac{i}{2}\boldsymbol{P}\left[\frac{W\bar{Q}_\alpha}{1+\bar{W}_\alpha} - \frac{\bar{W}Q_\alpha}{1+W_\alpha}\right].$$

Hence, setting

$$\underline{F} = F - i\frac{c}{2}F_1,$$

our equations become

$$\begin{cases}
W_t + \underline{F}(W_\alpha + 1) + i\dfrac{c}{2}W = 0, \\
Q_t + \underline{F}Q_\alpha + icQ - igW + \boldsymbol{P}\left[\dfrac{|Q_\alpha|^2}{J}\right] + i\dfrac{c}{2}\left\{\boldsymbol{P}\left[\dfrac{W\bar{Q}_\alpha}{1+\bar{W}_\alpha}\right] - \boldsymbol{P}\left[\dfrac{\bar{W}Q_\alpha}{1+W_\alpha}\right]\right\} = 0.
\end{cases} \tag{B-16}$$

We can also re-express the Hamiltonian and horizontal momentum in terms of the holomorphic variables (W, Q). This gives

$$\mathcal{E}(W,Q) = \Re\int g|W|^2(1+W_\alpha) - iQ\bar{Q}_\alpha + cQ_\alpha(\Im W)^2 - \frac{c^3}{2i}|W|^2W(1+W_\alpha)\,d\alpha. \tag{B-17}$$

A second conserved quantity is the horizontal momentum,

$$\mathcal{P} = \int \left\{ \frac{1}{i}(\bar{Q} W_\alpha - Q \overline{W}_\alpha) - c|W|^2 + \frac{c}{2}(W^2 \overline{W}_\alpha + \overline{W}^2 W_\alpha) \right\} d\alpha. \tag{B-18}$$

References

[Alazard and Delort 2015] T. Alazard and J.-M. Delort, "Global solutions and asymptotic behavior for two dimensional gravity water waves", *Ann. Sci. Éc. Norm. Supér.* (4) **48**:5 (2015), 1149–1238. MR Zbl

[Alazard, Burq and Zuily 2011] T. Alazard, N. Burq, and C. Zuily, "On the water-wave equations with surface tension", *Duke Math. J.* **158**:3 (2011), 413–499. MR Zbl

[Alazard, Burq and Zuily 2014] T. Alazard, N. Burq, and C. Zuily, "On the Cauchy problem for gravity water waves", *Invent. Math.* **198**:1 (2014), 71–163. MR Zbl

[Ashton and Fokas 2011] A. C. L. Ashton and A. S. Fokas, "A non-local formulation of rotational water waves", *J. Fluid Mech.* **689** (2011), 129–148. MR Zbl

[Christodoulou and Lindblad 2000] D. Christodoulou and H. Lindblad, "On the motion of the free surface of a liquid", *Comm. Pure Appl. Math.* **53**:12 (2000), 1536–1602. MR Zbl

[Coifman and Meyer 1976] R. R. Coifman and Y. Meyer, "Le double commutateur", Publication mathématique d'Orsay 180, 1976, available at https://tinyurl.com/Coifman180. Séminaire Analyse Harmonique d'Orsay.

[Constantin and Varvaruca 2011] A. Constantin and E. Varvaruca, "Steady periodic water waves with constant vorticity: regularity and local bifurcation", *Arch. Ration. Mech. Anal.* **199**:1 (2011), 33–67. MR Zbl

[Constantin, Kalimeris and Scherzer 2015] A. Constantin, K. Kalimeris, and O. Scherzer, "Approximations of steady periodic water waves in flows with constant vorticity", *Nonlinear Anal. Real World Appl.* **25** (2015), 276–306. MR Zbl

[Coutand and Shkoller 2007] D. Coutand and S. Shkoller, "Well-posedness of the free-surface incompressible Euler equations with or without surface tension", *J. Amer. Math. Soc.* **20**:3 (2007), 829–930. MR Zbl

[Dyachenko, Kuznetsov, Spector and Zakharov 1996] A. I. Dyachenko, E. A. Kuznetsov, M. D. Spector, and V. E. Zakharov, "Analytical description of the free surface dynamics of an ideal fluid (canonical formalism and conformal mapping)", *Phys. Lett. A* **221**:1-2 (1996), 73–79.

[Ehrnström 2008] M. Ehrnström, "A new formulation of the water wave problem for Stokes waves of constant vorticity", *J. Math. Anal. Appl.* **339**:1 (2008), 636–643. MR Zbl

[Germain and Masmoudi 2014] P. Germain and N. Masmoudi, "Global existence for the Euler–Maxwell system", *Ann. Sci. Éc. Norm. Supér.* (4) **47**:3 (2014), 469–503. MR Zbl

[Hunter and Ifrim 2012] J. K. Hunter and M. Ifrim, "Enhanced life span of smooth solutions of a Burgers–Hilbert equation", *SIAM J. Math. Anal.* **44**:3 (2012), 2039–2052. MR Zbl

[Hunter, Ifrim and Tataru 2016] J. K. Hunter, M. Ifrim, and D. Tataru, "Two dimensional water waves in holomorphic coordinates", *Comm. Math. Phys.* **346**:2 (2016), 483–552. MR Zbl

[Hunter, Ifrim, Tataru and Wong 2015] J. K. Hunter, M. Ifrim, D. Tataru, and T. K. Wong, "Long time solutions for a Burgers–Hilbert equation via a modified energy method", *Proc. Amer. Math. Soc.* **143**:8 (2015), 3407–3412. MR Zbl

[Ifrim and Tataru 2016] M. Ifrim and D. Tataru, "Two dimensional water waves in holomorphic coordinates, II: Global solutions", *Bull. Soc. Math. France* **144**:2 (2016), 369–394. MR Zbl

[Ifrim and Tataru 2017] M. Ifrim and D. Tataru, "The lifespan of small data solutions in two dimensional capillary water waves", *Arch. Ration. Mech. Anal.* **225**:3 (2017), 1279–1346. MR Zbl

[Ionescu and Pusateri 2015] A. D. Ionescu and F. Pusateri, "Global solutions for the gravity water waves system in 2d", *Invent. Math.* **199**:3 (2015), 653–804. MR Zbl

[Kozlov, Kuznetsov and Lokharu 2014] V. Kozlov, N. Kuznetsov, and E. Lokharu, "Steady water waves with vorticity: an analysis of the dispersion equation", *J. Fluid Mech.* **751** (2014), art. id. R3. MR

[Lannes 2005] D. Lannes, "Well-posedness of the water-waves equations", *J. Amer. Math. Soc.* **18**:3 (2005), 605–654. MR Zbl

[Levi-Civita 1925] T. Levi-Civita, "Détermination rigoureuse des ondes permanentes d'ampleur finie", *Math. Ann.* **93**:1 (1925), 264–314. MR Zbl

[Lindblad 2005] H. Lindblad, "Well-posedness for the motion of an incompressible liquid with free surface boundary", *Ann. of Math.* (2) **162**:1 (2005), 109–194. MR Zbl

[Ovsjannikov 1974] L. V. Ovsjannikov, "To the shallow water theory foundation", *Arch. Mech.* (*Arch. Mech. Stos.*) **26** (1974), 407–422. MR

[Shatah 1985] J. Shatah, "Normal forms and quadratic nonlinear Klein–Gordon equations", *Comm. Pure Appl. Math.* **38**:5 (1985), 685–696. MR Zbl

[Shatah and Zeng 2008] J. Shatah and C. Zeng, "Geometry and a priori estimates for free boundary problems of the Euler equation", *Comm. Pure Appl. Math.* **61**:5 (2008), 698–744. MR Zbl

[Strauss 2010] W. A. Strauss, "Steady water waves", *Bull. Amer. Math. Soc.* (*N.S.*) **47**:4 (2010), 671–694. MR Zbl

[Taylor 1950] G. Taylor, "The instability of liquid surfaces when accelerated in a direction perpendicular to their planes, I", *Proc. Roy. Soc. London. Ser. A.* **201** (1950), 192–196. MR Zbl

[Wahlén 2007] E. Wahlén, "A Hamiltonian formulation of water waves with constant vorticity", *Lett. Math. Phys.* **79**:3 (2007), 303–315. MR Zbl

[Wu 1997] S. Wu, "Well-posedness in Sobolev spaces of the full water wave problem in 2-D", *Invent. Math.* **130**:1 (1997), 39–72. MR Zbl

[Wu 2009] S. Wu, "Almost global wellposedness of the 2-D full water wave problem", *Invent. Math.* **177**:1 (2009), 45–135. MR Zbl

[Zhang and Zhang 2008] P. Zhang and Z. Zhang, "On the free boundary problem of three-dimensional incompressible Euler equations", *Comm. Pure Appl. Math.* **61**:7 (2008), 877–940. MR Zbl

Received 17 Feb 2017. Revised 22 May 2018. Accepted 14 Jul 2018.

MIHAELA IFRIM: ifrim@math.berkeley.edu
Department of Mathematics, University of California at Berkeley, Berkeley, CA, United States

DANIEL TATARU: tataru@math.berkeley.edu
Department of Mathematics, University of California at Berkeley, Berkeley, CA, United States

ANALYSIS AND PDE
Vol. 12, No. 4, 2019
dx.doi.org/10.2140/apde.2019.12.969

ABSOLUTE CONTINUITY AND α-NUMBERS ON THE REAL LINE

TUOMAS ORPONEN

Let μ, ν be Radon measures on $\mathbb{R}$, with μ nonatomic and ν doubling, and write $\mu = \mu_a + \mu_s$ for the Lebesgue decomposition of μ relative to ν. For an interval $I \subset \mathbb{R}$, define $\alpha_{\mu,\nu}(I) := \mathbb{W}_1(\mu_I, \nu_I)$, the Wasserstein distance of normalised blow-ups of μ and ν restricted to I. Let $\mathcal{S}_\nu$ be the square function

$$\mathcal{S}_\nu^2(\mu) = \sum_{I \in \mathcal{D}} \alpha_{\mu,\nu}^2(I)\chi_I,$$

where $\mathcal{D}$ is the family of dyadic intervals of side-length at most 1. I prove that $\mathcal{S}_\nu(\mu)$ is finite μ_a almost everywhere and infinite μ_s almost everywhere. I also prove a version of the result for a nondyadic variant of the square function $\mathcal{S}_\nu(\mu)$. The results answer the simplest "$n = d = 1$" case of a problem of J. Azzam, G. David and T. Toro.

1. Introduction

Wasserstein distance and α-numbers. In this paper, μ and ν are nonzero Radon measures on $\mathbb{R}$. The measure ν is generally assumed to be either *dyadically doubling* or *globally doubling*. Dyadically doubling means that

$$\nu(\hat{I}) \leq C\nu(I), \qquad I \in \mathcal{D}, \tag{1.1}$$

where $\mathcal{D}$ is the standard family of dyadic intervals, and $\hat{I}$ is the *parent* of I, that is, the smallest interval in $\mathcal{D}$ strictly containing I. Globally doubling means that $\nu(B(x, 2r)) \leq C\nu(B(x, r))$ for $x \in \mathbb{R}$ and $r > 0$; in particular, this implies $\operatorname{spt} \nu = \mathbb{R}$. The main example for ν is the Lebesgue measure $\mathcal{L}$, and the proofs in this particular case would differ little from the ones presented below. No a priori homogeneity is assumed of μ.

The research was supported by the grants 274512 and 309365 of the Finnish Academy.
MSC2010: 42A99.
Keywords: Wasserstein distance, α-numbers, doubling measures.

Definition 1.2 (Wasserstein distance). I will use the following definition of the (first) Wasserstein distance: given two Radon measures ν_1, ν_2 on [0, 1], set

$$\mathbb{W}_1(\nu_1, \nu_2) := \sup_{\psi} \left| \int \psi \, d\nu_1 - \int \psi \, d\nu_2 \right|,$$

where the sup is taken over all 1-Lipschitz functions $\psi\colon \mathbb{R} \to \mathbb{R}$ which are supported on [0, 1]. Such functions will be called *test functions*. A slightly different — and also quite common — definition would allow the sup to run over all 1-Lipschitz functions $\psi\colon [0, 1] \to \mathbb{R}$. To illustrate the difference, let $\nu_1 = \delta_0$ and $\nu_2 = \delta_1$. Then $\mathbb{W}_1(\nu_1, \nu_2) = 0$, but the alternative definition, say $\widetilde{\mathbb{W}}_1$, would give $\widetilde{\mathbb{W}}_1(\nu_1, \nu_2) = 1$. The main reason for using $\mathbb{W}_1$ instead of $\widetilde{\mathbb{W}}_1$ in this paper is to comply with the definitions in [Azzam et al. 2016; 2017].

As in the paper [Azzam et al. 2016] of J. Azzam, G. David and T. Toro, I make the following definition:

Definition 1.3 (α-numbers). Assume that $I \subset \mathbb{R}$ is an interval. Define

$$\alpha_{\mu,\nu}(I) := \mathbb{W}_1(\mu_I, \nu_I),$$

where μ_I and ν_I are normalised blow-ups of μ and ν restricted to I. More precisely, let T_I be the increasing affine mapping taking $\bar{I}$ to [0, 1], and define

$$\mu_I := \frac{T_{I\sharp}(\mu|_I)}{\mu(I)} \quad \text{and} \quad \nu_I := \frac{T_{I\sharp}(\nu|_I)}{\nu(I)}.$$

If $\mu(I) = 0$ (or $\nu(I) = 0$), define $\mu_I \equiv 0$ (or $\nu_I \equiv 0$).

The quantity defined above is somewhat awkward to work with, as it lacks (see Example 5.2) the following desirable stability property: if $I, J \subset \mathbb{R}$ are intervals of comparable length, and $I \subset J$, then $\alpha_{\mu,\nu}(I) \lesssim \alpha_{\mu,\nu}(J)$. Chiefly for this reason, I also need to consider the following "smooth" α-numbers; the definition below is essentially the same as the one given in [Azzam et al. 2017, Section 5]:

Definition 1.4 (smooth α-numbers). Let $\varphi := \operatorname{dist}(\,\cdot\,, \mathbb{R} \setminus (0, 1))$. For an interval $I \subset \mathbb{R}$, define $\alpha_{s,\mu,\nu}(I) := \mathbb{W}_1(\mu_{\varphi,I}, \nu_{\varphi,I})$, where

$$\mu_{\varphi,I} := \frac{T_{I\sharp}(\mu|_I)}{\mu(\varphi_I)} \quad \text{and} \quad \nu_{\varphi,I} := \frac{T_{I\sharp}(\nu|_I)}{\nu(\varphi_I)}.$$

Here T_I is the map from Definition 1.3, $\varphi_I = \varphi \circ T_I$, and $\mu(\varphi_I) = \int \varphi_I \, d\mu$. If $\mu(\varphi_I) = 0$ (or $\nu(\varphi_I) = 0$), set $\mu_{\varphi,I} \equiv 0$ (or $\nu_{\varphi,I} \equiv 0$).

The only difference between the numbers $\alpha_{\mu,\nu}(I)$ and $\alpha_{s,\mu,\nu}(I)$ is in the normalisation of the measures μ_I, φ_I and $\mu_{\varphi,I}, \nu_{\varphi,I}$: if I is closed, the measures μ_I, ν_I are probability measures on [0, 1], while $\mu_{\varphi,I}([0, 1]) = \mu(I)/\mu(\varphi_I)$. The numbers $\alpha_{s,\mu,\nu}(I)$ enjoy the stability property alluded to above. Moreover, if either μ or ν is a doubling, one has $\alpha_{s,\mu,\nu}(I) \lesssim \alpha_{\mu,\nu}(I)$. These facts are contained in Proposition 5.4 (or see [Azzam et al. 2017, Section 5]).

Remark 1.5. The α-numbers were first introduced by X. Tolsa [2009], where he used them to characterise the uniform rectifiability of Ahlfors–David regular measures in $\mathbb{R}^d$. Tolsa's original definition of the

α-numbers has a different, asymmetric, normalisation compared to either $\alpha_{\mu,\nu}$ or $\alpha_{s,\mu,\nu}$ above; see [Tolsa 2009, p. 394].

Main results. Before explaining the results in [Azzam et al. 2016], and their connection to the current manuscript, I emphasise that that paper treats "n-dimensional" measures in $\mathbb{R}^d$ for any $1 \le n \le d$. For the current paper, only the case $n = d = 1$ is relevant. So, to avoid digressing too much, I need to state the results of [Azzam et al. 2016] in far smaller generality than they deserve.

With this proviso in mind, the main results of [Azzam et al. 2016] imply the following: if μ is a doubling measure on $\mathbb{R}$, and the numbers $\alpha_{\mu,\mathcal{L}}$ satisfy a Carleson condition of the form

$$\int_{B(x,2r)} \int_0^{2r} \alpha_{\mu,\mathcal{L}}(B(y,t)) \, \frac{dt \, d\mu y}{t} \le C\mu(B(x,r)), \tag{1.6}$$

then μ, or at least a large part of μ, is absolutely continuous with respect to $\mathcal{L}$, with quantitative upper and lower bounds on the density. As the authors of [Azzam et al. 2016] point out, the main shortcoming of their result is that condition (1.6) imposes a hypothesis on the first powers of the numbers $\alpha_{\mu,\mathcal{L}}$, whereas evidence suggests (see in that paper Remark 1.6.1, the discussion after Theorem 1.7, and Example 4.6) that the correct power should be 2. More support for this belief comes from the following "converse" result [Tolsa 2015, Lemma 2.2]: if μ is a finite Borel measure on $\mathbb{R}$ then

$$\int_0^{\infty} \tilde{\alpha}^2_{\mu,\mathcal{L}}(x,r) \, \frac{dr}{r} < \infty \quad \text{for } \mathcal{L} \text{ a.e. } x \in \mathbb{R}. \tag{1.7}$$

In particular, if $\mu \ll \mathcal{L}$, then (1.7) holds for μ almost every $x \in \mathbb{R}$. I should again mention that this is only the easiest $n = d = 1$ case of Tolsa's result. Here $\tilde{\alpha}_{\mu,\mathcal{L}}$ is a variant of the α-number (in fact the one discussed in Remark 1.5).

The purpose of this paper is to address the problem of Azzam, David and Toro in the simplest case $n = d = 1$. I show that control for the second powers of the $\alpha_{\mu,\mathcal{L}}$-numbers does guarantee absolute continuity with respect to Lebesgue measure. In fact, the doubling assumption on μ can be dropped, the Carleson condition (1.6) can be relaxed considerably, and the results remain valid, if $\mathcal{L}$ is replaced by any doubling measure ν. The results below also contain the "converse" statement, analogous to (1.7).

I prove two variants of the main result: one dyadic, and the other nondyadic. Here is the dyadic version:

Theorem 1.8. *Let $\mathcal{D}$ be the family of dyadic subintervals of $[0,1)$, and let μ, ν be Borel probability measures on $[0,1)$. Assume that μ does not charge the boundaries of intervals in $\mathcal{D}$, and ν is dyadically doubling. Write $\mu = \mu_a + \mu_s$ for the Lebesgue decomposition of μ relative to ν, where $\mu_a \ll \nu$ and $\mu_s \perp \nu$. Finally, let $\mathcal{S}_{\mathcal{D},\nu}(\mu)$ be the square function*

$$\mathcal{S}^2_{\mathcal{D},\nu}(\mu) = \sum_{I \in \mathcal{D}} \alpha^2_{\mu,\nu}(I)\chi_I.$$

Then,

(a) *$\mathcal{S}_\nu(\mu)$ is finite μ_a almost surely, and*

(b) *$\mathcal{S}_\nu(\mu)$ is infinite μ_s almost surely.*

In particular,

$$\sum_{I \in \mathcal{D}} \alpha_{\mu,\nu}^2(I)\mu(I) < \infty \quad \Longrightarrow \quad \mu \ll \nu.$$

Heuristically, this corresponds to assuming (1.6) at the scale $r = 1$, but I could not find a way to *reduce* the continuous problem to the dyadic one; on the other hand, a reduction in the other direction does not appear straightforward either, so perhaps one needs to treat the cases separately. A caveat of the dyadic set-up is the "nonatomicity" hypothesis on μ. It cannot be dispensed with: for instance, if $\mu = \delta_x$ for any $x \in [0, 1)$, which only belongs to the interiors of finitely many dyadic intervals, then $\mathcal{S}_{\mathcal{D},\mathcal{L}}(\mu)$ is uniformly bounded (for instance $\mathcal{S}_{\mathcal{D},\mathcal{L}}(\delta_0) \equiv 0$), but $\mu \perp \mathcal{L}$.

Here is the nondyadic version of the main theorem:

Theorem 1.9. *Assume that μ, ν are Radon measures, and ν is globally doubling. Write $\mu = \mu_a + \mu_s$, as in Theorem 1.8. Let $\mathcal{S}_\nu$ be the square function*

$$\mathcal{S}_\nu^2(\mu)(x) = \int_0^1 \alpha_{s,\mu,\nu}^2(B(x,r))\,\frac{dr}{r}, \quad x \in \mathbb{R},$$

defined via the smooth α-numbers $\alpha_{s,\mu,\nu}$. Then,

(a) *$\mathcal{S}_\nu(\mu)$ is finite μ_a almost surely, and*

(b) *$\mathcal{S}_\nu(\mu)$ is infinite μ_s almost surely.*

Recall that $\alpha_{s,\mu,\nu}(B(x,r)) \lesssim \alpha_{\mu,\nu}(B(x,r))$ whenever ν is doubling, such as $\nu = \mathcal{L}$; see Proposition 5.4. So, Theorem 1.9 has the following corollary:

Corollary 1.10. *Assume that μ, ν are Radon measures, and ν is globally doubling. If*

$$\int_0^1 \alpha_{\mu,\nu}^2(B(x,t))\,\frac{dt}{t} < \infty \tag{1.11}$$

for μ almost every $x \in \mathbb{R}$, then $\mu \ll \nu$.

The following question remains open:

Question 1. *In the setting of Theorem 1.9 and Corollary 1.10, is the square function in (1.11) finite μ_a almost everywhere?*

The difficulties arise from the nonstability of the numbers $\alpha_{\mu,\nu}$. See [Azzam et al. 2017, Section 5], and in particular Lemma 5.3 of that paper, for related discussion.

Assuming the full Carleson condition (1.6), and that μ is globally doubling, the authors of [Azzam et al. 2016] prove something more quantitative than $\mu \ll \mathcal{L}$; see in particular Theorem 1.9 of that paper. The same ought to be true for the second powers of the α-numbers, and indeed the following result can be easily deduced with the method of the current paper:

Theorem 1.12. *Assume that μ, ν are Borel probability measures on* $[0, 1)$, *both dyadically doubling, and assume that the Carleson condition*

$$\sum_{I \subset J} \alpha_{\mu,\nu}^2(I)\mu(I) \le C\mu(J), \quad J \in \mathcal{D}, \tag{1.13}$$

holds for some $C \geq 1$. *Then* μ *belongs to* $A_\infty^{\mathcal{D}}(\nu)$, *the dyadic* A_∞ *class relative to* ν. *Similarly, if* μ, ν *are Radon measures on* $\mathbb{R}$, *both globally doubling, and the Carleson condition* (1.6) *holds for the second powers* $\alpha^2_{\mu,\nu}(B(y,t))$, *then* $\mu \in A_\infty(\nu)$.

The a priori doubling assumptions cannot be omitted (that is, they are not implied by the Carleson condition): Just consider $\mu = 2\chi_{[0,1/2)}\, d\mathcal{L}$. It is clear that the Carleson condition (1.13) holds for the numbers $\alpha^2_{\mu,\mathcal{L}}(I)$, but nevertheless $\mu \notin A_\infty^{\mathcal{D}}(\mathcal{L}|_{[0,1]})$.

Outline of the paper, and the main steps of the proofs. The main substance of the paper is proving the dyadic result, Theorem 1.8, and in particular part (b). This work takes up Sections 2-4. The proof of part (a) is simpler, and closely follows a previous argument of Tolsa — namely the one used to prove (1.7). The details (both in the dyadic and continuous settings) are given in Section 6. Modifications required to prove part (b) of the "continuous" Theorem 1.9 are outlined in Section 5.

The proof of Theorem 1.8(b) has three main steps. First, the numbers $\alpha_{\mu,\nu}(I)$ are used to control something analyst-friendlier, namely the dyadic variants

$$\Delta_{\mu,\nu}(I) = \left| \frac{\mu(I_-)}{\mu(I)} - \frac{\nu(I_-)}{\nu(I)} \right|. \tag{1.14}$$

Here I_- stands for the left half of I. This would be simple, if $\chi_{[0,1/2)}$ happened to be one of the admissible test functions ψ in the definition of $\mathbb{W}_1$. It is not, however, and in fact there seems to be no direct (and sufficiently efficient) way to control $\Delta_{\mu,\nu}(I)$ by $\alpha_{\mu,\nu}(I)$, or even $\alpha_{\mu,\nu}(3I)$. However, it turns out that the numbers are equivalent at the level of certain Carleson sums over trees; proving this statement is the main content of Section 2.

The numbers $\Delta_{\mu,\nu}(I)$ are well-known quantities: they are the (absolute values of the) coefficients in an orthogonal representation of μ in terms of ν-adapted Haar functions, and it is known that they can be used to characterise A_∞. The following theorem is due to S. Buckley [1993]:

Theorem 1.15 [Buckley 1993, Theorem 2.2(iii)]. *Let* μ, ν *be dyadically doubling Borel probability measures on* [0, 1]. *Then* $\mu \in A_\infty^{\mathcal{D}}(\nu)$ *if and only if*

$$\sum_{I \subset J} \Delta^2_{\mu,\nu}(I)\mu(I) \leq C\mu(J), \qquad J \in \mathcal{D}. \tag{1.16}$$

The result in [Buckley 1993] is only stated for $\nu = \mathcal{L}|_{[0,1]}$, but the proof works in the greater generality. Note the similarity between the Carleson conditions (1.16) and (1.13): The dyadic part of Theorem 1.12 is, in fact, nothing but a corollary of Buckley's result, assuming that one knows how to control the numbers $\Delta_{\mu,\nu}(I)$ by the numbers $\alpha_{\mu,\nu}(I)$ at the level of Carleson sums; consequently, the short proof of this half of Theorem 1.12 can be found in Section 2. The continuous version is discussed briefly in Remark 5.19.

Buckley's result is not applicable for Theorem 1.8: the measure μ is not dyadically doubling, and the information available is much weaker than the Carleson condition (1.13). Handling these issues constitutes the remaining two steps in the proof: all dyadic intervals are split into trees, where μ is "tree-doubling" (Section 4), and the absolute continuity of μ with respect to ν is studied in each tree separately (Section 3).

2. Comparison of α-numbers and Δ-numbers

In this section, μ and ν are Borel probability measures on $[0, 1)$, μ does not charge the boundaries of dyadic intervals, and ν is dyadically doubling inside $[0, 1)$:

$$\nu(\hat{I}) \leq D_\nu \nu(I), \quad I \in \mathcal{D} \setminus \{[0, 1)\}.$$

This implies, in particular, that $\nu(I) > 0$ for all $I \in \mathcal{D}$ with $I \subset [0, 1)$. The main task of the section is to bound the numbers $\Delta_{\mu,\nu}(I)$ by the numbers $\alpha_{\mu,\nu}(I)$, where $\Delta_{\mu,\nu}(I)$ is the quantity

$$\Delta_{\mu,\nu}(I) = \left|\frac{\mu(I_-)}{\mu(I)} - \frac{\nu(I_-)}{\nu(I)}\right| = \left|\int \chi_{(0,1/2)}\, d\mu_I - \int \chi_{(0,1/2)}\, d\nu_I\right|.$$

The task would be trivial if $\chi_{(0,1/2)}$ were a 1-Lipschitz function vanishing at the boundary of $[0, 1]$. It is not: in fact, the difference between $\Delta_{\nu_1,\nu_2}(I)$ and $\alpha_{\nu_1,\nu_2}(I)$ can be rather large for a given interval I.

Example 2.1. If $\nu_1 = \delta_{1/2-1/n}$ and $\nu_2 = \delta_{1/2+1/n}$, then $\Delta_{\nu_1,\nu_2}([0, 1)) = 1$, but $\alpha_{\nu_1,\nu_2}([0, 1)) \lesssim 1/n$. These measures do not satisfy the assumptions of the section, so consider also the following example. Let $\mu = f\, d\mathcal{L}$, where f takes the value 1 everywhere, except in the 2^{-n}-neighbourhood of $\frac{1}{2}$. Let $f \equiv \frac{1}{2}$ on the interval $\left[\frac{1}{2} - 2^{-n}, \frac{1}{2}\right]$, and $f \equiv \frac{3}{2}$ on the interval $\left(\frac{1}{2}, \frac{1}{2} + 2^{-n}\right]$. Then μ is a dyadically 4-doubling probability measure on $[0, 1]$, $\Delta_{\mu,\mathcal{L}}([0, 1)) \sim 2^{-n}$, and $\alpha_{\mu,\mathcal{L}}([0, 1)) \sim 2^{-2n}$.

Fortunately, "pointwise" estimates between $\Delta_{\mu,\nu}(I)$ and $\alpha_{\mu,\nu}(I)$ are not really needed in this paper, and it turns out that certain sums of these numbers are comparable, up to a manageable error. To state such results, I need to introduce some terminology. A family $\mathcal{C} \subset \mathcal{D}$ of dyadic intervals is called *coherent* if the implication

$$Q, R \in \mathcal{C} \text{ and } Q \subset P \subset R \quad \Longrightarrow \quad P \in \mathcal{C}$$

holds for all $Q, P, R \in \mathcal{D}$.

Definition 2.2 (trees, leaves, boundary). A *tree* $\mathcal{T} \subset \mathcal{D}$ is any coherent family of dyadic intervals with a unique largest interval, $\mathrm{Top}(\mathcal{T}) \in \mathcal{T}$, and with the property that

$$\mathrm{card}(\mathrm{ch}(I) \cap \mathcal{T}) \in \{0, 2\}, \quad I \in \mathcal{T}.$$

For the tree $\mathcal{T}$, define the set family $\mathrm{Leaves}(\mathcal{T})$ to consist of the minimal intervals in $\mathcal{T}$, in other words those $I \in \mathcal{T}$ with $\mathrm{card}(\mathrm{ch}(I) \cap \mathcal{T}) = 0$. Abusing notation, I often write $\mathrm{Leaves}(\mathcal{T})$ also for the set $\bigcup\{I : I \in \mathrm{Leaves}(\mathcal{T})\}$. Finally, define the *boundary* of the tree $\partial\mathcal{T}$ by

$$\partial T := \mathrm{Top}(\mathcal{T}) \setminus \mathrm{Leaves}(\mathcal{T}).$$

Then $x \in \partial\mathcal{T}$ if and only if $x \in \mathrm{Top}(\mathcal{T})$ and all intervals $I \in \mathcal{D}$ with $x \in I \subset \mathrm{Top}(\mathcal{T})$ are contained $\mathcal{T}$.

Definition 2.3 (($\mathcal{T}$, D)-doubling measures). A Borel probability measure μ on $[0, 1]$ is called $(\mathcal{T}, D)$-doubling if

$$\mu(\hat{I}) \leq D\mu(I), \quad I \in \mathcal{T} \setminus \mathrm{Top}(\mathcal{T}).$$

Here is the main result of this section:

Proposition 2.4. *Let μ, ν be measures satisfying the assumptions of the section, and let $\mathcal{T} \subset \mathcal{D}$ be a tree. Moreover, assume that μ is $(\mathcal{T}, D)$-doubling for some constant $D \geq 1$. Then*

$$\sum_{I \in \mathcal{T}} \Delta^2_{\mu,\nu}(I)\mu(I) \lesssim_{D_\nu, D} \sum_{I \in \mathcal{T} \setminus \mathrm{Leaves}(\mathcal{T})} \alpha^2_{\mu,\nu}(I)\mu(I) + \mu(\mathrm{Top}(\mathcal{T})).$$

The "dyadic part" of Theorem 1.12 is an immediate corollary:

Proof of Theorem 1.12, dyadic part. By hypothesis, both measures μ and ν are $(\mathcal{D}, C)$-doubling. Hence, by the Carleson condition (1.13), and Proposition 2.4 applied to the trees $\mathcal{T}_J := \{I \in \mathcal{D} : I \subset J\}$, one has

$$\sum_{I \subset J} \Delta^2_{\mu,\nu}(I)\mu(I) \lesssim_C \sum_{I \subset J} \alpha^2_{\mu,\nu}(I)\mu(I) + \mu(J) \lesssim \mu(J).$$

This is precisely the condition in Buckley's result, Theorem 1.15, so $\mu \in A^{\mathcal{D}}_\infty(\nu)$. □

I then begin the proof of Proposition 2.4. It would, in fact, suffice to assume that ν is also just $(\mathcal{T}, D_\nu)$-doubling, but checking this would result in some unnecessary book-keeping below. The proof is based on the observation that $\chi_{(0,1/2)}$ can be written as a series of Lipschitz functions, each supported on subintervals of $[0, 1]$. This motivates the following considerations.

Assume that

$$\Psi := \Psi_0 := \sum_{j \geq 0} \psi_j$$

is a bounded function such that each $\psi_j : \mathbb{R} \to [0, \infty)$ is an L_j-Lipschitz function supported on some interval $I_j \in \mathcal{D}_j$. Assume moreover that the intervals I_j are nested: $[0, 1) \supset I_1 \supset I_2 \supset \cdots$. Then, as a first step in proving Proposition 2.4, I claim that

$$\left| \int \Psi \, d\mu - \int \Psi \, d\nu \right| \\ \leq \sum_{k=0}^{N} \frac{L_k}{2^k} \alpha_{\mu,\nu}(I_k)\mu(I_k) + \sum_{k=0}^{N} \left(\frac{1}{\nu(I_{k+1})} \int \Psi_{k+1} \, d\nu \right) \Delta_{\mu,\nu}(I_k)\mu(I_k) + 2\|\Psi\|_\infty \mu(I_{N+1}) \quad (2.5)$$

for any $N \in \{0, 1, \ldots, \infty\}$, where

$$\Psi_k := \sum_{j \geq k} \psi_j, \quad m \geq 0.$$

For $N = \infty$, the symbol "I_{N+1}" should be interpreted as the intersection of all the intervals I_j. I will first verify that, for any $m \geq 0$,

$$\left| \frac{1}{\mu(I_m)} \int \Psi_m \, d\mu - \frac{1}{\nu(I_m)} \int \Psi_m \, d\nu \right| \leq \frac{L_m}{2^m} \alpha_{\mu,\nu}(I_m) + \left(\frac{1}{\nu(I_{m+1})} \int \Psi_{m+1} \, d\nu \right) \Delta_{\mu,\nu}(I_m) \\ + \frac{\mu(I_{m+1})}{\mu(I_m)} \left| \frac{1}{\mu(I_{m+1})} \int \Psi_{m+1} \, d\mu - \frac{1}{\nu(I_{m+1})} \int \Psi_{m+1} \, d\nu \right|, \quad (2.6)$$

from which it will be easy to derive (2.5). If $\mu(I_m) = 0$, the corresponding term should be interpreted as "0" (recall that $\nu(I_m)$ is never zero by the doubling assumption). The proof of (2.6) is straightforward.

First, note that since $\psi_m \colon \mathbb{R} \to \mathbb{R}$ is an L_m-Lipschitz function supported on I_m, and $|I_m| = 2^{-m}$, one has

$$\left|\frac{1}{\mu(I_m)} \int \psi_m \, d\mu - \frac{1}{\nu(I_m)} \int \psi_m \, d\nu\right| = \left|\int \psi_m \circ T_{I_m}^{-1} \, d\mu_{I_m} - \int \psi_m \circ T_{I_m}^{-1} \, d\nu_{I_m}\right| \leq \frac{L_m}{2^m} \alpha_{\mu,\nu}(I_m).$$

(The mappings T_I are familiar from Definition 1.3). This gives rise to the first term in (2.6). What remains is bounded by

$$\begin{aligned}\left|\frac{1}{\mu(I_m)} \int \Psi_{m+1} \, d\mu - \frac{1}{\nu(I_m)} \int \Psi_{m+1} \, d\nu\right| &\leq \frac{\mu(I_{m+1})}{\mu(I_m)} \left|\frac{1}{\mu(I_{m+1})} \int \Psi_{m+1} \, d\mu - \frac{1}{\nu(I_{m+1})} \int \Psi_{m+1} \, d\nu\right| \\ &\quad + \left(\frac{1}{\nu(I_{m+1})} \int \Psi_{m+1} \, d\nu\right) \left|\frac{\mu(I_{m+1})}{\mu(I_m)} - \frac{\nu(I_{m+1})}{\nu(I_m)}\right|.\end{aligned}$$

This is (2.6), observing that

$$\Delta_{\mu,\nu}(I_m) = \left|\frac{\mu(I_{m+1})}{\mu(I_m)} - \frac{\nu(I_{m+1})}{\nu(I_m)}\right|,$$

since either $I_{m+1} = (I_m)_+$ or $I_{m+1} = (I_m)_-$, and both possibilities give the same number $\Delta_{\mu,\nu}(I_m)$. Finally, (2.5) is obtained by repeated application of (2.6). By induction, one can check that N iterations of (2.6) (starting from $m = 0$, and recalling that μ, ν are probability measures on $[0, 1)$) leads to

$$\begin{aligned}\left|\int \Psi \, d\mu - \int \Psi \, d\nu\right| &\leq \sum_{k=0}^{N} \frac{L_k}{2^k} \alpha_{\mu,\nu}(I_k) \mu(I_k) + \sum_{k=0}^{N} \left(\frac{1}{\nu(I_{k+1})} \int \Psi_{k+1} \, d\nu\right) \Delta_{\mu,\nu}(I_k) \mu(I_k) \\ &\quad + \mu(I_{N+1}) \left|\frac{1}{\mu(I_{N+1})} \int \Psi_{N+1} \, d\mu - \frac{1}{\nu(I_{N+1})} \int \Psi_{N+1} \, d\nu\right|. \end{aligned} \tag{2.7}$$

This gives (2.5) immediately, observing that $\|\Psi_{N+1}\|_\infty \leq \|\Psi\|_\infty$.

Now, it is time to specify the functions ψ_j. I first define a hands-on Whitney decomposition for $\left(0, \frac{1}{2}\right)$. Pick a small parameter $\tau > 0$, to be specified later, and let $U_0 := \left[\tau, \frac{1}{2} - \tau\right)$. Then, set $U_{-k} := [\tau 2^{-k}, \tau 2^{-k+1})$ and $U_k := \frac{1}{2} - U_{-k}$ for $k \geq 1$. Let $\{\psi_k\}_{k \in \mathbb{Z}}$ be a partition of unity subordinate to slightly enlarged versions of the sets U_k, $k \in \mathbb{Z}$. By this, I first mean that each ψ_k is nonnegative and L_k-Lipschitz with

$$L_k \leq \frac{C 2^{|k|}}{\tau}. \tag{2.8}$$

Second, the supports of the functions ψ_k should satisfy $\psi_0 \subset \left[\frac{1}{2}\tau, \frac{1}{2} - \frac{1}{2}\tau\right)$,

$$\operatorname{spt} \psi_{-k} \subset \left[\left(\tfrac{1}{2}\tau\right) 2^{-k}, 2\tau 2^{-k+1}\right) \subset (0, 2\tau 2^{-k+1}) \quad \text{and} \quad \psi_k \subset \tfrac{1}{2} - (0, 2\tau 2^{-k+1})$$

for $k \geq 1$. Third,

$$\sum_{k \in \mathbb{Z}} \psi_k = \chi_{(0,1/2)}.$$

Let $\Psi^- := \sum_{k>0} \psi_{-k} + \frac{1}{2}\psi_0$ and $\Psi^+ := \sum_{k>0} \psi_k + \frac{1}{2}\psi_0$. Then

$$\Delta_{\mu,\nu}([0, 1)) \leq \left|\int \Psi^- \, d\mu - \int \Psi^- \, d\nu\right| + \left|\int \Psi^+ \, d\mu - \int \Psi^+ \, d\nu\right|. \tag{2.9}$$

This is the only place in the paper where the assumption of μ not charging the boundaries of dyadic intervals is used (however, the estimate (2.9) will eventually be applied to all the measures μ_I, $I \in \mathcal{D}$, so the full strength of the hypothesis is needed). The function Ψ^- is precisely of the form treated above with $I_j := [0, 2^{-j})$, since clearly $\operatorname{spt} \psi_{-k} \subset I_k$. Applying the inequality (2.5) with any $N_1 \in \{0, 1, \ldots, \infty\}$ yields

$$\left|\int \Psi^- \, d\mu - \int \Psi^- \, d\nu\right| \\ \leq \sum_{k=0}^{N_1} \frac{L_{-k}}{2^k} \alpha_{\mu,\nu}(I_k)\mu(I_k) + \sum_{k=0}^{N_1} \left(\frac{1}{\nu(I_{k+1})} \int \Psi^-_{k+1} \, d\nu \right) \Delta_{\mu,\nu}(I_k)\mu(I_k) + 2\mu(I_{N_1+1}). \quad (2.10)$$

Next, observe that each function Ψ^-_{k+1}, $k \geq 0$, is bounded by 1 and vanishes outside

$$\bigcup_{j=k+1}^{\infty} \operatorname{spt} \psi_{-k} \subset (0, 2\tau 2^{-k}).$$

It follows that

$$\frac{1}{\nu(I_{k+1})} \int \Psi^-_{k+1} \, d\nu \leq \frac{\nu((0, 2\tau 2^{-k}))}{\nu(I_{k+1})} = o_{D_\nu}(\tau),$$

where the implicit constants only depend on the dyadic doubling constant D_ν of ν. In the sequel, I assume that τ is so small that $o_{D_\nu}(\tau) \leq \kappa$, where $\kappa > 0$ is another small constant, which will eventually depend on the $(\mathcal{T}, D)$-doubling constant D for μ. Recalling also (2.8), the estimate (2.10) then becomes

$$\left|\int \Psi^- \, d\mu - \int \Psi^- \, d\nu\right| \leq \frac{C}{\tau} \sum_{k=0}^{N_1} \alpha_{\mu,\nu}(I_k)\mu(I_k) + \kappa \sum_{k=0}^{N_1} \Delta_{\mu,\nu}(I_k)\mu(I_k) + 2\mu(I_{N_1+1}). \quad (2.11)$$

The last term simply vanishes if $N_1 = \infty$, because $\mu(\{0\}) = 0$. A heuristic point to observe is that the left-hand side is roughly $\Delta_{\mu,\nu}([0, 1])$; the right-hand side also contains the same term, but multiplied by a small constant $\kappa > 0$. This gain is “paid for” by the large constant C/τ.

Next, the estimate is replicated for Ψ^+. This time, the inequality (2.5) is applied to the sequence $\tilde{I}_0 = [0, 1)$, $\tilde{I}_1 = \left[0, \frac{1}{2}\right)$, $\tilde{I}_2 = (\tilde{I}_1)_+$, and in general $\tilde{I}_{k+1} = (\tilde{I}_k)_+$ for $k \geq 1$ (here J_+ is the right half of J). Then, if τ is small enough, it is again clear that $\operatorname{spt} \psi_k \subset \tilde{I}_k$. Thus, by inequality (2.5),

$$\left|\int \Psi^+ \, d\mu - \int \Psi^+ \, d\nu\right| \\ \leq \sum_{k=0}^{N_2} \frac{L_k}{2^k} \alpha_{\mu,\nu}(\tilde{I}_k)\mu(\tilde{I}_k) + \sum_{k=0}^{N_2} \left(\frac{1}{\nu(\tilde{I}_{k+1})} \int \Psi^+_{k+1} \, d\nu \right) \Delta_{\mu,\nu}(\tilde{I}_k)\mu(\tilde{I}_k) + 2\mu(\tilde{I}_{N_2+1}) \quad (2.12)$$

for any $N_2 \geq 0$. As before, the term $\mu(\tilde{I}_{N_2})$ vanishes for $N_2 = \infty$ $\left(\text{because } \mu\left(\left\{\frac{1}{2}\right\}\right) = 0\right)$, and one can ensure

$$\frac{1}{\nu(\tilde{I}_{k+1})} \int \Psi^+_{k+1} \, d\nu \leq \kappa$$

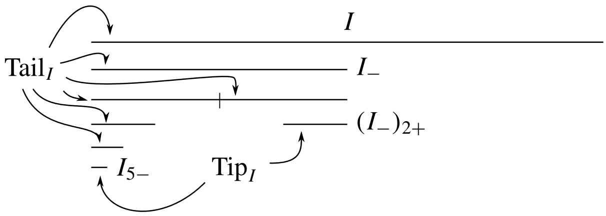

Figure 1. An example of $\mathrm{Tail}_I(4, 1)$ and Tip_I.

by choosing $\tau = \tau(D_\nu) > 0$ small enough. Consequently (recalling (2.9)), (2.11) and (2.12) together imply

$$\Delta_{\mu,\nu}([0, 1)) \leq \frac{C}{\tau} \sum_{I \in \mathrm{Tail}} \alpha_{\mu,\nu}(I)\mu(I) + \kappa \sum_{I \in \mathrm{Tail}} \Delta_{\mu,\nu}(I)\mu(I) + 2\mu(I_{N_1+1}) + 2\mu(\tilde{I}_{N_2+1}). \tag{2.13}$$

Here Tail is the collection of all the intervals $I_0, \dots, I_{N_1}$ and $\tilde{I}_0, \dots, \tilde{I}_{N_2}$. The intervals $[0, 1)$ and $\left[0, \frac{1}{2}\right)$ arise a total of two times from (2.11) and (2.12), but this has no visible impact on the end result, (2.13). The estimate (2.13) generalises in a simple way to other intervals $I \in \mathcal{D}$, besides $I = [0, 1)$, but requires an additional piece of notation. Let $I \in \mathcal{D}$, and write $I_{0-} := I =: I_{0+}$. For $k \geq 1$, define $I_{k-} := (I_{(k-1)-})_-$ and $I_{k+} := (I_{(k-1)+})_+$. Now, for a fixed dyadic interval $I \subset [0, 1)$, and $N_1, N_2 \geq 0$, let $\mathrm{Tail}_I = \mathrm{Tail}_I(N_1, N_2)$ be the collection of subintervals of I, which includes I_{k-} for all $0 \leq k \leq N_1$ and $(I_-)_{k+}$ for all $0 \leq k \leq N_2$; see Figure 1. Then, the generalisation of (2.13) reads

$$\Delta_{\mu,\nu}(I)\mu(I) \leq \frac{C}{\tau} \sum_{J \in \mathrm{Tail}_I} \alpha_{\mu,\nu}(J)\mu(J) + \kappa \sum_{J \in \mathrm{Tail}_I} \Delta_{\mu,\nu}(J)\mu(J) + 2\mu(\mathrm{Tip}_I), \tag{2.14}$$

where $\mathrm{Tip}_I = I_{(N_1+1)-} \cup (I_-)_{(N_2+1)+}$. If $N_1 < \infty$ and $N_2 = \infty$, for instance, then $\mathrm{Tip}_I = I_{(N_1+1)-}$. The proof is nothing but an application of (2.13) to the measures μ_I and ν_I. For minor technical reasons, I also wish to allow the choice $N_1 = 0$ and $N_2 = -1$: by definition, this choice means that $\mathrm{Tail}_I = \{I\}$ and $\mathrm{Tip}_I := I_-$. It is easy to see that (2.14) remains valid in this case, with “2” replaced by “4” (for $I = [0, 1)$, this follows by applying (2.11) and (2.12) with the choices $N_1 = 0 = N_2$).

Now, the table is set to prove Proposition 2.4, which I recall here:

Proposition 2.4. *Let μ, ν be measures satisfying the assumptions of the section, and let $\mathcal{T} \subset \mathcal{D}$ be a tree. Moreover, assume that μ is $(\mathcal{T}, D)$-doubling for some constant $D \geq 1$. Then*

$$\sum_{I \in \mathcal{T}} \Delta^2_{\mu,\nu}(I)\mu(I) \lesssim_{D_\nu, D} \sum_{I \in \mathcal{T} \setminus \mathrm{Leaves}(\mathcal{T})} \alpha^2_{\mu,\nu}(I)\mu(I) + \mu(\mathrm{Top}(\mathcal{T})).$$

Proof. The sum over $I \in \mathrm{Leaves}(\mathcal{T})$ is evidently bounded by $4\mu(\mathrm{Top}(\mathcal{T}))$, so it suffices to consider

$$I \in \mathcal{T} \setminus \mathrm{Leaves}(\mathcal{T}) =: \mathcal{T}^-. \tag{2.15}$$

Let $I \in \mathcal{T}$, and define the number $N_1 = N_1(I) \geq 0$ as the smallest index such that $I_{(N_1+1)-} \in \mathrm{Leaves}(\mathcal{T})$. If no such index exists, set $N_1 = \infty$. If $I_- \in \mathrm{Leaves}(\mathcal{T})$, then $N_1 = 0$, and I define $N_2 = -1$: then $\mathrm{Tail}_I := \{I\}$, and $\mathrm{Tip}_I := I_-$. Otherwise, if $I_- \in \mathcal{T}^-$, let $N_2 \geq 0$ be the smallest index such that $(I_-)_{(N_2+1)+} \in \mathrm{Leaves}(\mathcal{T})$. If no such index exists, let $N_2 = \infty$. Now $\mathrm{Tail}_I \subset \mathcal{T}^-$ and $\mathrm{Tip}_I \subset \mathrm{Leaves}(\mathcal{T})$ are defined as after (2.14). Start by the following combination of (2.14) and Cauchy–Schwarz:

$$\Delta^2_{\mu,\nu}(I)\mu(I)^2 \lesssim \frac{1}{\tau^2}\Big(\sum_{J\in \mathrm{Tail}_I} \alpha^2_{\mu,\nu}(J)\mu(J)^{3/2}\Big)\Big(\sum_{J\in \mathrm{Tail}_I} \mu(J)^{1/2}\Big) + \kappa^2\Big(\sum_{J\in \mathrm{Tail}_I} \Delta^2_{\mu,\nu}(J)\mu(J)^{3/2}\Big)\Big(\sum_{J\in \mathrm{Tail}_I} \mu(J)^{1/2}\Big) + \mu(\mathrm{Tip}_I)^2. \tag{2.16}$$

The factors $\sum_{J\in \mathrm{Tail}_I} \mu(J)^{1/2}$ are under control, thanks to the $(\mathcal{T}, D)$-doubling hypothesis on μ, and the fact that $\mathrm{Tail}_I \subset \mathcal{T}$. Since Tail_I consists of two "branches" of nested intervals inside I, and the $(\mathcal{T}, D)$-doubling hypothesis implies that the μ-measures of intervals decay geometrically along these branches, one arrives at

$$\sum_{J\in \mathrm{Tail}_I} \mu(J)^{1/2} \lesssim_D \mu(I)^{1/2}.$$

Thus, by (2.16),

$$\Delta^2_{\mu,\nu}(I)\mu(I) \lesssim_D \frac{1}{\tau^2}\sum_{J\in \mathrm{Tail}_I} \alpha^2_{\mu,\nu}(J)\frac{\mu(J)^{3/2}}{\mu(I)^{1/2}} + \kappa^2 \sum_{J\in \mathrm{Tail}_I} \Delta^2_{\mu,\nu}(J)\frac{\mu(J)^{3/2}}{\mu(I)^{1/2}} + \frac{\mu(\mathrm{Tip}_I)^2}{\mu(I)}. \tag{2.17}$$

The constant $\kappa > 0$ will have to be chosen so small, eventually, that its product with the implicit constants above is notably less than 1. From now on, the precise restriction $J \in \mathrm{Tail}_I$ can be replaced by the conditions $J \in \mathcal{T}^-$ and $J \subset I$. With this in mind, observe first that

$$\sum_{I\in\mathcal{T}^-}\sum_{\substack{J\in\mathcal{T}^-\\ J\subset I}} \alpha^2_{\mu,\nu}(J)\frac{\mu(J)^{3/2}}{\mu(I)^{1/2}} = \sum_{J\in\mathcal{T}^-} \alpha^2_{\mu,\nu}(J)\mu(J)^{3/2} \sum_{\substack{I\in\mathcal{T}^-\\ I\supset J}} \frac{1}{\mu(I)^{1/2}} \lesssim_D \sum_{J\in\mathcal{T}^-} \alpha^2_{\mu,\nu}(J)\mu(J).$$

The final inequality uses, again, the geometric decay of μ-measures of intervals in $\mathcal{T}$. A similar estimate can be performed for the second term in (2.17). As for the third term,

$$\begin{aligned}\sum_{I\in\mathcal{T}^-} \frac{\mu(\mathrm{Tip}_I)^2}{\mu(I)} &\lesssim \sum_{I\in\mathcal{T}^-} \frac{\mu(I_{(N_1+1)-})^2 + \mu((I_-)_{(N_2+1)+})^2}{\mu(I)}\\ &\lesssim \sum_{J\in \mathrm{Leaves}(\mathcal{T})} \mu(J)^2 \sum_{\substack{I\in\mathcal{T}^-\\ I\supset J}} \frac{1}{\mu(I)} \lesssim_D \mu(\mathrm{Leaves}(\mathcal{T})),\end{aligned}$$

relying once more on the geometric decay of μ in $\mathcal{T}$. Combining all the estimates gives

$$\sum_{I\in\mathcal{T}^-} \Delta^2_{\mu,\nu}(I)\mu(I) \lesssim_D \frac{1}{\tau^2}\sum_{I\in\mathcal{T}^-} \alpha^2_{\mu,\nu}(I)\mu(I) + \kappa^2 \sum_{I\in\mathcal{T}^-} \Delta^2_{\mu,\nu}(I)\mu(I) + \mu(\mathrm{Leaves}(\mathcal{T})). \tag{2.18}$$

If the left-hand side is a priori finite, the proof of Proposition 2.4 is now completed by choosing κ small enough, depending on D. If not, consider any finite subtree $\mathcal{T}_j \subset \mathcal{T}$ with $\mathrm{Top}(\mathcal{T}_j) = \mathrm{Top}(\mathcal{T})$. Then, the proof above gives (2.18) with $\mathcal{T}_j$ in place of $\mathcal{T}$. Hence

$$\sum_{I\in\mathcal{T}_j^-}\Delta^2_{\mu,\nu}(I)\mu(I)\lesssim_D\sum_{I\in\mathcal{T}_j^-}\alpha^2_{\mu,\nu}(I)\mu(I)+\mu(\mathrm{Top}(\mathcal{T})),$$

where the constants do not depend on the choice of $\mathcal{T}_j$. Now the proposition follows by letting $\mathcal{T}_j\nearrow\mathcal{T}$. □

3. Absolute continuity of tree-adapted measures

Recall the concepts of tree, leaves and boundaries from Definition 2.2, and the notion of $(\mathcal{T},D)$-doubling measures from Definition 2.3. In the present section, I assume that $\mathcal{T}\subset\mathcal{D}$ is a tree, and μ,ν are two finite Borel measures, which satisfy the following two assumptions:

(A) $\min\{\mu(\mathrm{Top}(\mathcal{T})),\nu(\mathrm{Top}(\mathcal{T}))\}>0$.

(B) μ,ν are $(\mathcal{T},D)$-doubling for some constant $D\geq 1$.

In particular, the assumptions imply

$$\mu(I)>0\quad\text{and}\quad\nu(I)>0,\quad I\in\mathcal{T}.$$

For reasons to become apparent soon, I define the *$(\mathcal{T},\mu)$-adaptation of ν*,

$$\nu_{\mathcal{T}}:=\nu|_{\partial\mathcal{T}}+\sum_{I\in\mathrm{Leaves}(\mathcal{T})}\frac{\nu}{\mu}(I)\cdot\mu|_I,$$

where $(\nu/\mu)(I):=\nu(I)/\mu(I)$. Note that

$$\nu_{\mathcal{T}}(I)=\nu(I),\quad I\in\mathcal{T},\tag{3.1}$$

because $\partial\mathcal{T}$ is disjoint from the leaves, which are also pairwise disjoint. In particular, $\nu_{\mathcal{T}}(\mathrm{Top}(\mathcal{T}))=\nu(\mathrm{Top}(\mathcal{T}))$. The main result of the section is the following:

Proposition 3.2. *Assume* (A) *and* (B), *and that*

$$\sum_{I\in\mathcal{T}\setminus\mathrm{Leaves}(\mathcal{T})}\Delta^2_{\mu,\nu}(I)\mu(I)<\infty.$$

Then $\mu|_{\mathrm{Top}(\mathcal{T})}\ll\nu_{\mathcal{T}}$. In particular $\mu|_{\partial\mathcal{T}}\ll\nu$.

Remark 3.3. By the definition of $\nu_{\mathcal{T}}$, it is obvious that $\mu|_{\mathrm{Leaves}(\mathcal{T})}\ll\nu_{\mathcal{T}}$. So, the main point of Proposition 3.2 is to show that $\mu|_{\partial\mathcal{T}}\ll(\nu_{\mathcal{T}})|_{\partial\mathcal{T}}=\nu|_{\partial\mathcal{T}}$.

Since $\mu(\mathrm{Top}(\mathcal{T}))>0$ and $\nu(\mathrm{Top}(\mathcal{T}))>0$, one may assume without loss of generality that

$$\mu(\mathrm{Top}(\mathcal{T}))=1=\nu(\mathrm{Top}(\mathcal{T})).$$

The proof of Proposition 3.2 is based on a "product representation" for $\nu_{\mathcal{T}}$, relative to μ, in the spirit of [Fefferman et al. 1991, Theorem 3.22] of Fefferman, Kenig and Pipher. Recall that every interval $I\in\mathcal{D}$ has exactly two children: I_- and I_+. Define the μ-adapted Haar functions

$$h_I^\mu:=c_I^+\chi_{I_+}-c_I^-\chi_{I_-},\quad I\in\mathcal{T}\setminus\mathrm{Leaves}(\mathcal{T}),$$

where

$$c_I^+ := \frac{\mu(I)}{\mu(I_+)} \quad \text{and} \quad c_I^- := \frac{\mu(I)}{\mu(I_-)}.$$

This ensures that $\int h_I^\mu \, d\mu = 0$ for $I \in \mathcal{T} \setminus \text{Leaves}(\mathcal{T})$. Note that $\mu(I_+), \mu(I_-) > 0$, because $I_+, I_- \in \mathcal{T}$. Now, the plan is to define coefficients $a_J \in \mathbb{R}$, for $J \in \mathcal{T} \setminus \text{Leaves}(\mathcal{T})$, so that the following requirement is met:

$$\prod_{\substack{J \supsetneq I \\ J \in \mathcal{T}}} (1 + a_J h_J^\mu)(x) = \frac{\nu}{\mu}(I), \quad x \in I \in \mathcal{T}. \tag{3.4}$$

The left-hand side of (3.4) is certainly constant on I, so the equation has some hope; if $I = \text{Top}(\mathcal{T})$, then the product is empty, and the right-hand side of (3.4) equals 1 by the assumption $\mu(\text{Top}(\mathcal{T})) = \nu(\text{Top}(\mathcal{T})) = 1$. Now, assume that (3.4) holds for some interval $I \in \mathcal{T} \setminus \text{Leaves}(\mathcal{T})$. Then $I_-, I_+ \in \mathcal{T}$, so if (3.4) is supposed to hold for I_-, one has

$$\frac{\nu}{\mu}(I_-) = \prod_{\substack{J \supsetneq I_- \\ J \in \mathcal{T}}} (1 + a_J h_J^\mu) = (1 - c_I^- a_I) \prod_{\substack{J \supsetneq I \\ J \in \mathcal{T}}} (1 + a_J h_J^\mu) = (1 - c_I^- a_I) \frac{\nu}{\mu}(I), \tag{3.5}$$

and similarly

$$\frac{\nu}{\mu}(I_+) = (1 + c_I^+ a_I) \frac{\nu}{\mu}(I). \tag{3.6}$$

From (3.5) one solves

$$a_I = \frac{(\nu/\mu)(I) - (\nu/\mu)(I_-)}{(\nu/\mu)(I) c_I^-} = \frac{\mu(I_-)}{\mu(I)} - \frac{\nu(I_-)}{\nu(I)}, \tag{3.7}$$

and (3.6) gives

$$a_I = \frac{(\nu/\mu)(I_+) - (\nu/\mu)(I)}{(\nu/\mu)(I) c_I^+} = \frac{\nu(I_+)}{\nu(I)} - \frac{\mu(I_+)}{\mu(I)}. \tag{3.8}$$

Using that $\mu(I_-)/\mu(I) = 1 - \mu(I_+)/\mu(I)$ (and three other similar formulae), it is easy to see that the numbers on the right-hand sides of (3.7) and (3.8) agree. So, a_I can be defined consistently, and (3.4) holds for $I_+, I_- \in \mathcal{T}$. Moreover, the formulae for a_I look quite familiar:

Observation 1. $|a_I| = \Delta_{\mu,\nu}(I)$ for $I \in \mathcal{T} \setminus \text{Leaves}(\mathcal{T})$.

Now that the coefficients a_I have been successfully defined for $I \in \mathcal{T} \setminus \text{Leaves}(\mathcal{T})$, let g be the (at the moment) formal series

$$g(x) := \sum_{I \in \mathcal{T} \setminus \text{Leaves}(\mathcal{T})} a_I h_I^\mu(x).$$

Since the Haar functions h_I^μ are orthogonal in $L^2(\mu)$, and satisfy

$$\int (h_I^\mu)^2 \, d\mu \le \max\{c_I^+, c_I^-\}^2 \mu(I) \le D^2 \mu(I), \quad I \in \mathcal{T} \setminus \text{Leaves}(\mathcal{T}),$$

one arrives at

$$\|g\|_{L^2(\mu)}^2 = \sum_{I \in \mathcal{T} \setminus \text{Leaves}(\mathcal{T})} \Delta_{\mu,\nu}^2(I) \|h_I\|_{L^2(\mu)}^2 \le D^2 \sum_{I \in \mathcal{T} \setminus \text{Leaves}(\mathcal{T})} \Delta_{\mu,\nu}^2(I) \mu(I) < \infty,$$

by the assumption in Proposition 3.2. This means that the sequence

$$g_N := \sum_{\substack{I \in \mathcal{T} \setminus \mathrm{Leaves}(\mathcal{T}) \\ |I| > 2^{-N}}} a_I h_I^{\mu}$$

converges in $L^2(\mu)$. In particular, one can pick a subsequence $(g_{N_j})_{j \in \mathbb{N}}$ which converges pointwise μ almost everywhere (in fact, the entire sequence converges by basic martingale theory, but this is not needed). Now, recall that the goal was to prove that $\mu|_{\mathrm{Top}(\mathcal{T})} \ll \nu_{\mathcal{T}}$. To this end, one has to verify that

$$\liminf_{I \to x} \frac{\mu}{\nu_{\mathcal{T}}}(I) < \infty \tag{3.9}$$

at μ almost every $x \in \mathrm{Top}(\mathcal{T})$. This is clear for $x \in \mathrm{Leaves}(\mathcal{T})$, since the ratios $\mu(I)/\nu_{\mathcal{T}}(I)$, $I \ni x$, are eventually constant. So, it suffices to prove (3.9) at μ almost every point $x \in \partial\mathcal{T}$. Fix a point $x \in \partial\mathcal{T}$ with the properties that sequence $(g_{N_j}(x))_{j \in \mathbb{N}}$ converges, and also

$$\sum_{x \in J \in \mathcal{T}} a_J^2 = \sum_{x \in J \in \mathcal{T}} \Delta_{\mu,\nu}^2(I) < \infty. \tag{3.10}$$

These properties hold at μ almost every $x \in \partial\mathcal{T}$. Let $I \in \mathcal{D}$ be so small that $x \in I \in \mathcal{T}$, and note that

$$\log \frac{\nu_{\mathcal{T}}}{\mu}(I) = \log \frac{\nu}{\mu}(I) = \log \prod_{\substack{J \supsetneq I \\ J \in \mathcal{T}}} (1 + a_J h_J^{\mu}(x)) = \sum_{\substack{J \supsetneq I \\ J \in \mathcal{T}}} \log(1 + a_J h_J^{\mu}(x)).$$

Now, the plan is to use the estimate $\log(1+t) \geq t - C_\delta t^2$, which is valid as long as $t \geq \delta - 1$ for some $\delta > 0$. Observe that $a_J h_J^{\mu}(x) \in \{-c_J^- a_J, c_J^+ a_J\}$, where

$$-a_J c_J^- = \frac{(\nu/\mu)(J_-)}{(\nu/\mu)(J)} - 1 \geq \frac{1}{C} - 1 \quad \text{and} \quad a_J c_J^+ = \frac{(\nu/\mu)(J_+)}{(\nu/\mu)(J)} - 1 \geq \frac{1}{C} - 1. \tag{3.11}$$

Consequently, for $x \in I \in \mathcal{T}$ with $|I| = 2^{-N_j}$, one has

$$\log \frac{\nu_{\mathcal{T}}}{\mu}(I) \geq \sum_{\substack{J \supsetneq I \\ J \in \mathcal{T}}} a_J h_J^{\mu}(x) - C' \sum_{\substack{J \supsetneq I \\ J \in \mathcal{T}}} (a_J h_J^{\mu}(x))^2 \geq g_{N_j}(x) - C' D^2 \sum_{x \in J \in \mathcal{T}} a_J^2, \tag{3.12}$$

where $C' \lesssim_D 1$ only depends on the constant C in (3.11). Since the sequence $(g_{N_j}(x))_{j \in \mathbb{N}}$ converges and (3.10) holds, the right-hand side of (3.12) has a uniform lower bound $-M(x) > -\infty$. This implies

$$\limsup_{I \to x} \frac{\nu_{\mathcal{T}}}{\mu}(I) \geq \exp(-M(x)) > 0,$$

which gives (3.9) at x. The proof of Proposition 3.2 is complete.

4. Proof of Theorem 1.8(b)

In this section, Theorem 1.8(b) is proved via a simple tree construction, coupled with Propositions 2.4 and 3.2. Recall the statement of Theorem 1.8(b):

Theorem 4.1. *Assume that μ, ν are Borel probability measures on $[0, 1)$, μ does not charge the boundaries of dyadic intervals, and ν is dyadically doubling. Write $\mu = \mu_a + \mu_s$ for the Lebesgue decomposition of μ relative to ν, and let $\mathcal{S}_{\mathcal{D},\nu}(\mu)$ be the square function*

$$\mathcal{S}^2_{\mathcal{D},\nu}(\mu) = \sum_{I \in \mathcal{D}} \alpha^2_{\mu,\nu}(I)\chi_I.$$

Then, $\mathcal{S}_{\mathcal{D},\nu}(\mu)$ is infinite μ_s almost surely.

An equivalent statement is that the restriction of μ to the set

$$G := \{x \in [0, 1) : \mathcal{S}_{\mathcal{D},\nu}(\mu)(x) < \infty\}$$

is absolutely continuous with respect to ν; this is the formulation proven below. For the rest of the section, fix the measures μ, ν as in the statement above, and let D be the doubling constant of ν. I record a simple lemma, which says that the doubling of ν implies the doubling of μ on intervals, where the α-number is small enough.

Lemma 4.2. *There are constants $\epsilon > 0$ and $C \geq 1$, depending only on D, such that the following holds. For every interval $I \in \mathcal{D}$, if $\alpha_{\mu,\nu}(I) < \epsilon$, then*

$$\mu(I) \leq C \min\{\mu(I_-), \mu(I_+)\}. \tag{4.3}$$

Proof. Let $I_{--} \subset I_-$ and $I_{++} \subset I_+$ be intervals which lie at distance $\geq \frac{1}{8}|I|$ from the boundaries of I_- and I_+, respectively, and have length $\frac{1}{8}|I|$. Let ψ_- and $\psi_+ : \mathbb{R} \to [0, 1]$ be $(C'/|I|)$-Lipschitz functions which equal 1 on I_{--} and I_{++}, respectively, and are supported on I_- and I_+. Then

$$\frac{\mu(I_-)}{\mu(I)} \geq \frac{1}{\mu(I)} \int \psi_- \, d\mu \geq \frac{1}{\nu(I)} \int \psi_- \, d\nu - C'\alpha_{\mu,\nu}(I) \geq \frac{\nu(I_{--})}{\nu(I)} - C'\alpha_{\mu,\nu}(I),$$

and the analogous inequality holds for $\mu(I_+)/\mu(I)$. The ratio $\nu(I_{--})/\nu(I)$ is at least $1/D^3$, so if $\alpha_{\mu,\nu}(I) < 1/(2C'D^3) =: \epsilon$, then both $\mu(I_-) \geq [1/(2D^3)]\mu(I)$ and $\mu(I_+) \geq [1/(2D^3)]\mu(I)$. This gives (4.3) with $C = 2D^3$. □

In particular, if $\mathcal{T}$ is a tree, and $\alpha_{\mu,\nu}(I) < \epsilon$ for all $I \in \mathcal{T} \setminus \mathrm{Leaves}(\mathcal{T})$, then μ is $(\mathcal{T}, C)$-doubling. I will now describe how such trees $\mathcal{T}_j \subset \mathcal{D}$ are constructed, starting with $\mathcal{T}_0$. Let $[0, 1) = \mathrm{Top}(\mathcal{T}_0)$, and assume that some interval I is in $\mathcal{T}_0$. If

$$\sum_{I \subset J \subset [0,1)} \alpha^2_{\mu,\nu}(J) \geq \epsilon^2, \tag{4.4}$$

add I to $\mathrm{Leaves}(\mathcal{T}_0)$. The children I_- and I_+ become the tops of new trees. If (4.4) fails, add I_- and I_+ to $\mathcal{T}_0$. The construction of $\mathcal{T}_0$ is now complete. If a new top T_j was created in the process of constructing $\mathcal{T}_0$, and $\mu(T_j) > 0$, construct a new tree $\mathcal{T}_j$ with $\mathrm{Top}(\mathcal{T}_j) = T_j$ by repeating the algorithm above, only replacing $[0, 1)$ by T_j in the stopping criterion (4.4). Continue this process until all intervals in $\mathcal{D}$ belong to some tree, or all remaining tops T_j satisfy $\mu(T_j) = 0$. For all tops T_j with $\mu(T_j) = 0$, simply define $\mathcal{T}_j := \{I \in \mathcal{D} : I \subset T_j\}$, so there is no further stopping inside $\mathcal{T}_j$.

Remark 4.5. Let $\mathcal{T}$ be one of the trees constructed above, with $\mu(\mathrm{Top}(\mathcal{T})) > 0$. Then μ is $(\mathcal{T}, C)$-doubling by Lemma 4.2, since it is clear that $\alpha_{\mu,\nu}(I) < \epsilon$ for all $I \in \mathcal{T} \setminus \mathrm{Leaves}(\mathcal{T})$. In particular $\mu(I) > 0$ for all $I \in \mathcal{T}$.

The following observation is now rather immediate from the definitions:

Lemma 4.6. *Assume that $\mathcal{T}_0, \ldots, \mathcal{T}_{N-1}$ are distinct trees such that $x \in \mathrm{Leaves}(\mathcal{T}_j)$ for all $0 \leq j \leq N-1$. Then*

$$\mathcal{S}^2_{\mathcal{D},\nu}(\mu)(x) \geq \epsilon^2 N.$$

Proof. For $0 \leq j \leq N-1$, let $I_j \in \mathrm{Leaves}(\mathcal{T}_j)$ with $x \in I_j$. Then

$$\mathcal{S}^2_{\mathcal{D},\nu}(\mu)(x) \geq \sum_{j=0}^{N-1} \sum_{I_j \subset J \subset \mathrm{Top}(\mathcal{T}_j)} \alpha^2_{\mu,\nu}(J) \geq \epsilon^2 N,$$

as claimed. □

It follows that μ almost every point in $G = \{x \in [0,1) : \mathcal{S}_\nu(\mu)(x) < \infty\}$ belongs to $\mathrm{Leaves}(\mathcal{T}_j)$ for only finitely many trees $\mathcal{T}_j$. This is equivalent to saying that μ almost every point in G belongs to $\partial\mathcal{T}$ for some tree $\mathcal{T}$. The converse is also true: if x belongs to $\partial\mathcal{T}$ for some tree $\mathcal{T}$, then clearly $\mathcal{S}_\nu(\mu)(x) < \infty$. Consequently

$$\mu|_G = \sum_{\text{trees } \mathcal{T}} \mu|_{\partial\mathcal{T}}.$$

To prove Theorem 4.1, it now suffices to show that $\mu|_{\partial\mathcal{T}} \ll \nu$ for every tree $\mathcal{T}$. This is clear, if $\mu(\mathrm{Top}(\mathcal{T})) = 0$, so I exclude the trivial case to begin with. In the opposite case, note that

$$\sum_{I \in \mathcal{T} \setminus \mathrm{Leaves}(\mathcal{T})} \alpha^2_{\mu,\nu}(I)\mu(I) = \int \sum_{I \in \mathcal{T} \setminus \mathrm{Leaves}(\mathcal{T})} \alpha^2_{\mu,\nu}(I)\chi_I(x)\, d\mu x \leq \epsilon^2 \cdot \mu(\mathrm{Top}(\mathcal{T})). \tag{4.7}$$

It then follows from Proposition 2.4 that

$$\sum_{I \in \mathcal{T}} \Delta^2_{\mu,\nu}(I)\mu(I) \lesssim \mu(\mathrm{Top}(\mathcal{T})) < \infty,$$

and the claim $\mu|_{\partial\mathcal{T}} \ll \nu$ is finally a consequence of Proposition 3.2. The proof of Theorem 1.8(b) is complete.

5. The nondyadic square function

This section contains the proof of Theorem 1.9(b). The argument naturally contains many similarities to the one given above. The main novelty is that one needs to work with the smooth α-numbers, introduced in Definition 1.4 (or [Azzam et al. 2016, Section 5]).

Smooth α-numbers, and their properties. I recall the definition of the smooth α-numbers:

Definition 5.1 (smooth α-numbers). Write $\varphi(x) = \operatorname{dist}(x, \mathbb{R} \setminus (0,1))$. For an interval $I \subset \mathbb{R}$, define $\alpha_{s,\mu,\nu}(I) := \mathbb{W}_1(\mu_{\varphi,I}, \nu_{\varphi,I})$, where

$$\mu_{\varphi,I} := \frac{T_{I\sharp}(\mu|_I)}{\mu(\varphi_I)} \quad \text{and} \quad \nu_{\varphi,I} := \frac{T_{I\sharp}(\nu|_I)}{\nu(\varphi_I)}.$$

Here $\varphi_I = \varphi \circ T_I$, and $\mu(\varphi_I) = \int \varphi_I \, d\mu$. If $\mu(\varphi_I) = 0$ set $\mu_{\varphi,I} \equiv 0$, and if $\nu(\varphi_I) = 0$, set $\nu_{\varphi,I} \equiv 0$. Unwrapping the definition, if $\mu(\varphi_I), \nu(\varphi_I) > 0$, then

$$\alpha_{s,\mu,\nu}(I) = \sup_{\psi} \left| \frac{1}{\mu(\varphi_I)} \int \psi \circ T_I \, d\mu - \frac{1}{\nu(\varphi_I)} \int \psi \circ T_I \, d\nu \right| = \sup_{\psi} \left| \frac{\mu(\psi_I)}{\mu(\varphi_I)} - \frac{\nu(\psi_I)}{\nu(\varphi_I)} \right|,$$

where the sup is taken over test functions ψ.

Recall that the main reason to prefer the smooth α-numbers over the ones from Definition 1.3 is the following stability property: if $I \subset J$ are intervals of comparable length, then $\alpha_{s,\mu,\nu}(I) \lesssim \alpha_{s,\mu,\nu}(J)$, whenever either μ or ν is doubling. This fact is essentially [Azzam et al. 2017, Lemma 5.2], but I include a proof in Proposition 5.4 for completeness. Similar stability is not true for the numbers $\alpha_{\mu,\nu}(I)$ and $\alpha_{\mu,\nu}(J)$, even for very nice measures μ and ν, as the following example demonstrates:

Example 5.2. Fix $n \in \mathbb{N}$, and let $I_-^n := \left[\frac{1}{2} - 2^{-n}, \frac{1}{2}\right]$ and $I_+^n := \left(\frac{1}{2}, \frac{1}{2} + 2^{-n}\right]$. Let μ be the same measure as in Example 2.1:

$$\mu = \chi_{\mathbb{R} \setminus (I_-^n \cup I_+^n)} + \tfrac{1}{2}\chi_{I_-^n} + \tfrac{3}{2}\chi_{I_+^n}.$$

Let $\nu = \mathcal{L}$. It is clear that both μ and ν are doubling, with constants independent of n. It is also easy to check that $\alpha_{\mu,\nu}(I) \lesssim 2^{-2n}$ for any interval I with length $|I| \sim 1$ such that $I_-^n \cup I_+^n \subset I$ (this implies $\mu(I) = \nu(I)$). However, $\alpha_{\mu,\nu}\left(\left[0, \frac{1}{2}\right]\right) \sim 2^{-n}$, because $\nu_{[0,1/2)} = \chi_{[0,1]}$, while

$$\mu_{[0,1/2]} = \left(1 + \frac{2^{-n}}{1 - 2^{-n}}\right)\chi_{[0,1-2^{1-n})} + \frac{1}{2}\left(1 + \frac{2^{-n}}{1 - 2^{-n}}\right)\chi_{[1-2^{1-n},1]}.$$

So, for instance, it is clear that no inequality of the form $\alpha_{\mu,\nu}\left(\left[0, \frac{1}{2}\right]\right) \lesssim \alpha_{\mu,\nu}([-1,1])$ can hold.

Without any doubling assumptions, even the smooth α-numbers can behave badly:

Example 5.3. Let $\mu = \delta_{1/2}$, and $\nu = (1-\epsilon) \cdot \delta_{1/2+\epsilon} + \epsilon \cdot \delta_{1/4}$. Then $\alpha_{s,\mu,\nu}([-1,1]) \sim \epsilon$, but $\alpha_{s,\mu,\nu}\left(\left[0, \frac{1}{2}\right]\right) \sim 1$.

Proposition 5.4 (basic properties of the smooth α-numbers). *Let μ, ν be two Radon measures on $\mathbb{R}$, and let $I \subset \mathbb{R}$ be an interval. Then*

$$\alpha_{s,\mu,\nu}(I) \leq 2 \quad \textit{and} \quad \alpha_{s,\mu,\nu}(I) \leq \frac{2\alpha_{\mu,\nu}(I)}{\nu_I(\varphi)}.$$

Moreover, if ν is doubling with constant D, the following holds: if $I \subset J \subset \mathbb{R}$ are intervals with $|I| \geq \theta |J|$ for some $\theta > 0$, then

$$\alpha_{s,\mu,\nu}(I) \lesssim_{D,\theta} \alpha_{s,\mu,\nu}(J). \tag{5.5}$$

Proof. For the duration of the proof, fix an interval $I \subset \mathbb{R}$ with $\mu(\varphi_I), \nu(\varphi_I) > 0$. The cases where $\mu(\varphi_I) = 0$ or $\nu(\varphi_I) = 0$ always require a little case chase, which I omit. Recall that $\varphi = \chi_{[0,1]} \operatorname{dist}(\cdot, \{0, 1\})$. Note that any 1-Lipschitz function $\psi : \mathbb{R} \to \mathbb{R}$ supported on $[0, 1]$ must satisfy $|\psi| \leq \varphi$. Consequently $|\psi_I| \leq \varphi_I$ for any interval I, and so

$$\alpha_{s,\mu,\nu}(I) \leq \sup_{\psi} \left[\frac{\mu(|\psi_I|)}{\mu(\varphi_I)} + \frac{\nu(|\psi_I|)}{\nu(\varphi_I)} \right] \leq 2.$$

This proves the first inequality. For the second inequality, one may assume that $\alpha_{\mu,\nu}(I) > 0$, since otherwise $\mu|_{\operatorname{int} I} = c\nu|_{\operatorname{int} I}$ for some constant $c > 0$, and this also gives $\alpha_{s,\mu,\nu}(I) = 0$. After this observation, it is easy to reduce to the case $\mu(\varphi_I) > 0$ and $\nu(\varphi_I) > 0$. Fix a test function ψ. Using that $\mu_I(|\psi|) = \mu(|\psi_I|)/\mu(I) \leq \mu(\varphi_I)/\mu(I) = \mu_I(\varphi)$, one obtains

$$\begin{aligned}\left| \frac{\mu(\psi_I)}{\mu(\varphi_I)} - \frac{\nu(\psi_I)}{\nu(\varphi_I)} \right| &= \left| \frac{\mu_I(\psi)}{\mu_I(\varphi)} - \frac{\nu_I(\psi)}{\nu_I(\varphi)} \right| = \left| \frac{\mu_I(\psi)\nu_I(\varphi) - \nu_I(\psi)\mu_I(\varphi)}{\mu_I(\varphi)\nu_I(\varphi)} \right| \\ &\leq \frac{\mu_I(|\psi|)}{\mu_I(\varphi)\nu_I(\varphi)} |\mu_I(\varphi) - \nu_I(\varphi)| + \frac{\mu_I(\varphi)}{\mu_I(\varphi)\nu_I(\varphi)} |\mu_I(\psi) - \nu_I(\psi)| \leq \frac{2\alpha_{\mu,\nu}(I)}{\nu_I(\varphi)}.\end{aligned}$$

To prove the final claim, start with the following estimate for a test function ψ:

$$\left| \frac{\mu(\psi_I)}{\mu(\varphi_I)} - \frac{\nu(\psi_I)}{\nu(\varphi_I)} \right| \leq \frac{\nu(\varphi_J)}{\nu(\varphi_I)} \left| \frac{\mu(\psi_I)}{\mu(\varphi_J)} - \frac{\nu(\psi_I)}{\nu(\varphi_J)} \right| + \frac{\mu(|\psi_I|)}{\mu(\varphi_I)} \frac{\nu(\varphi_J)}{\nu(\varphi_I)} \left| \frac{\mu(\varphi_I)}{\mu(\varphi_J)} - \frac{\nu(\varphi_I)}{\nu(\varphi_J)} \right|.$$

Then, recall that $\mu(|\psi_I|) \leq \mu(\varphi_I)$. Further, it follows from the doubling of ν that $\nu(\varphi_J) \lesssim_{D,\theta} \nu(\varphi_I)$. Finally, notice that $\psi_I = (\psi_I \circ T_J^{-1}) \circ T_J$ and $\varphi_I = (\varphi_I \circ T_J^{-1}) \circ T_J$, where both

$$\psi_I \circ T_J^{-1} \quad \text{and} \quad \varphi_I \circ T_J^{-1}$$

are $(|J|/|I|)$-Lipschitz functions supported on $T_J(I) \subset [0, 1]$. Consequently,

$$\max\left\{ \left| \frac{\mu(\psi_I)}{\mu(\varphi_J)} - \frac{\nu(\psi_I)}{\nu(\varphi_J)} \right|, \left| \frac{\mu(\varphi_I)}{\mu(\varphi_J)} - \frac{\nu(\varphi_I)}{\nu(\varphi_J)} \right| \right\} \leq \frac{\alpha_{s,\mu,\nu}(J)}{\theta},$$

and the estimate (5.5) follows. □

***Proof of Theorem 1.9*(b).** In this section, ν is a globally doubling measure with constant $D \geq 1$, say. As in Section 4, it suffices to show that $\mu|_G \ll \nu$, where

$$G := \{x : \mathcal{S}_\nu(\mu)(x) < \infty\}.$$

Write

$$\alpha_{s,\mu,\nu}(J) =: \alpha(J), \quad J \subset \mathbb{R}.$$

Assume without loss of generality (or translate both measures μ and ν slightly) that $\mu(\partial I) = 0$ for all $I \in \mathcal{D}$. Also without loss of generality, one may assume that $\operatorname{spt} \mu \subset (0, 1)$: the reason is that the finiteness $\mathcal{S}_\nu(\mu)(x)$ is equivalent to the finiteness of $\mathcal{S}_\nu(\mu|_U)(x)$ for all $x \in U$, whenever $U \subset \mathbb{R}$ is open. So, it suffices to prove $\mu|_{U\cap G} \ll \nu$ for any bounded open set U. Whenever I write $\mathcal{D}$ in the sequel, I only mean the family $\{I \in \mathcal{D} : I \subset [0, 1)\}$.

I start with some standard discretisation arguments. For each $I \in \mathcal{D}$, associate a somewhat larger interval $B_I \supset I$ as follows. First, for $x \in \operatorname{spt}\mu$ and $k \in \mathbb{N}$, choose a radius $r_{x,k} > 0$ such that

$$\alpha(B(x, r_{x,k})) \leq 2 \inf\{\alpha(B(x,r)) : 1.1 \cdot 2^{-k-1} \leq r \leq 0.9 \cdot 2^{-k}\}. \tag{5.6}$$

Then

$$\alpha^2(B(x, r_{x,k})) \leq \left(\frac{1}{\ln[2 \cdot (0.9/1.1)]} \int_{1.1 \cdot 2^{-k-1}}^{0.9 \cdot 2^{-k}} 2\alpha(x,r) \, \frac{dr}{r} \right)^2 \lesssim \int_{2^{-k-1}}^{2^{-k}} \alpha^2(x,r) \, \frac{dr}{r}.$$

For $I \in \mathcal{D}$ with $|I| = 2^{-k}$ and $I \cap \operatorname{spt}\mu \neq \varnothing$, let B_I be some open interval of the form $B(x, r_{k-10})$, $x \in I$, such that

$$\alpha(B_I) \leq 2 \inf\{\alpha(B(y, r_{y,k-10})) : y \in I \cap \operatorname{spt}\mu\}.$$

The number "-10" simply ensures that $I \subset B_I$ with $\operatorname{dist}(I, \partial B_I) \sim |I|$, and

$$I \subset J \implies B_I \subset B_J \quad \text{for } I, J \in \mathcal{D}.$$

This implication also uses the slight separation between the scales, provided by the factors "1.1" and "0.9" in (5.6). For $I \in \mathcal{D}$ with $I \cap \operatorname{spt}\mu = \varnothing$, define $B_I := I$ (although this definition will never be really used). Now, a tree decomposition of $\mathcal{D}$ can be performed as in the previous section, replacing the stopping condition (4.4) by declaring $\operatorname{Leaves}(\mathcal{T})$ to consist of the maximal intervals $I \subset \operatorname{Top}(\mathcal{T})$ with

$$\sum_{I \subset J \subset \operatorname{Top}(\mathcal{T})} \alpha^2(B_I) \geq \epsilon^2,$$

where $\epsilon = \epsilon_D > 0$ is a suitable small number; in particular, $\epsilon > 0$ is chosen so small that $\alpha(B_I) \leq \epsilon$ implies $\mu(B_I) \lesssim \mu(I)$ (which is possible by a small modification of Lemma 4.2). If now $x \in \operatorname{Leaves}(\mathcal{T})$ for infinitely many different trees $\mathcal{T}$, then

$$\infty = \sum_{x \in I \in \mathcal{D}} \alpha^2(B_I) \leq 2 \sum_{k \in \mathbb{N}} \alpha^2(B(x, r_{x,k-10})) \lesssim \int_0^{2^{10}} \alpha^2(B(x,r)) \, \frac{dr}{r},$$

which implies that $x \notin G$. Repeating the argument from Section 4, this gives

$$\mu|_G \leq \sum_{\text{trees } \mathcal{T}} \mu|_{\partial \mathcal{T}}.$$

The converse inequality could also be deduced from the stability of the smooth α-numbers (Proposition 5.4), but it is not needed: the inequality already shows that it suffices to prove

$$\mu|_{\partial \mathcal{T}} \ll \nu \tag{5.7}$$

for any given tree $\mathcal{T}$. So, fix a tree $\mathcal{T}$. If $\epsilon > 0$ was chosen small enough (again depending on D), then μ is $(\mathcal{T}, C)$-doubling for some $C = C_D \geq 1$ in the usual sense:

$$\mu(\hat{I}) \leq C\mu(I), \quad I \in \mathcal{T} \setminus \operatorname{Top}(\mathcal{T}).$$

So, if one knew that

$$\sum_{I\in\mathcal{T}\setminus\mathrm{Leaves}(\mathcal{T})} \Delta^2_{\mu,\nu}(I)\mu(I) < \infty, \tag{5.8}$$

then the familiar Proposition 3.2 would imply (5.7), completing the entire proof.

The proof of (5.8) is based on the inequality

$$\sum_{I\in\mathcal{T}} \Delta^2_{\mu,\nu}(I)\mu(I) \lesssim \sum_{I\in\mathcal{T}\setminus\mathrm{Leaves}(\mathcal{T})} \alpha^2(B_I)\mu(B_I) + \mu(\mathrm{Top}(\mathcal{T})). \tag{5.9}$$

The right-hand side is finite by the same estimate as in (4.7) (start with $\mu(B_I) \lesssim \mu(I)$, using $\alpha(B_I) \le \epsilon$ for $I \in \mathcal{T} \setminus \mathrm{Leaves}(\mathcal{T})$). So, (5.9) implies (5.8). I start the proof of (5.9) by noting that if $I \in \mathcal{D}$, then

$$\begin{aligned}\Delta_{\mu,\nu}(I) &= \left|\frac{\nu(I_-)}{\nu(I)} - \frac{\mu(I_-)}{\mu(I)}\right| \\ &\le \frac{\nu(\varphi_{B_I})}{\nu(I)}\left|\frac{\nu(I_-)}{\nu(\varphi_{B_I})} - \frac{\mu(I_-)}{\mu(\varphi_{B_I})}\right| + \frac{\mu(I_-)}{\mu(I)}\frac{\nu(\varphi_{B_I})}{\nu(I)}\left|\frac{\mu(I)}{\mu(\varphi_{B_I})} - \frac{\nu(I)}{\nu(\varphi_{B_I})}\right|.\end{aligned} \tag{5.10}$$

Noting that $\nu(\varphi_{B_I})/\nu(I) \lesssim_D 1$, to prove (5.9), it suffices to control

$$\sum_{I\in\mathcal{T}\setminus\mathrm{Leaves}(\mathcal{T})} \left[\left|\frac{\nu(I_-)}{\nu(\varphi_{B_I})} - \frac{\mu(I_-)}{\mu(\varphi_{B_I})}\right|^2 + \left|\frac{\mu(I)}{\mu(\varphi_{B_I})} - \frac{\nu(I)}{\nu(\varphi_{B_I})}\right|^2\right]\mu(I) \tag{5.11}$$

by the right-hand side of (5.9). The main task is to find a suitable replacement for the "Tail-Tip" inequality (2.14), which I replicate here for comparison:

$$\Delta_{\mu,\nu}(I)\mu(I) \le \frac{C}{\tau}\sum_{J\in\mathrm{Tail}_I} \alpha_{\mu,\nu}(J)\mu(J) + \kappa\sum_{J\in\mathrm{Tail}_I} \Delta_{\mu,\nu}(J)\mu(J) + 2\mu(\mathrm{Tip}_I). \tag{5.12}$$

Glancing at (5.11), one sees that an analogue for the inequality above is actually needed for both the terms

$$\tilde{\Delta}_{B_I}(I_-) = \left|\frac{\nu(I_-)}{\nu(\varphi_{B_I})} - \frac{\mu(I_-)}{\mu(\varphi_{B_I})}\right| \quad\text{and}\quad \tilde{\Delta}_{B_I}(I) = \left|\frac{\mu(I)}{\mu(\varphi_{B_I})} - \frac{\nu(I)}{\nu(\varphi_{B_I})}\right|.$$

If $I_- \in \mathrm{Leaves}(\mathcal{T})$, then the trivial estimate $\tilde{\Delta}_{B_I}(I_-) \lesssim 1$ will suffice, so in the sequel I assume that

$$I, I_- \notin \mathrm{Leaves}(\mathcal{T}). \tag{5.13}$$

The goal is inequality (5.18) below. Fix B_I and $J \in \{I, I_-\}$. Assume for notational convenience that $|B_I| = 1$, and hence, also $|J| \sim 1$. In a familiar manner, start by writing

$$\chi_J = \sum_{k\in\mathbb{Z}} \psi_k, \tag{5.14}$$

where ψ_k is a nonnegative $C2^{|k|}$-Lipschitz function supported on either $J \subset B_I$ (for $k = 0$) or $J_{|k|-}$ (for negative k) or J_{k+} (for positive k). As in the proof of the original Tail-Tip inequality, it suffices to first estimate

$$\left|\frac{1}{\mu(\varphi_{B_I})}\int \Psi_0^+\,d\mu - \frac{1}{\nu(\varphi_{B_I})}\int \Psi_0^+\,d\nu\right|, \tag{5.15}$$

where $\Psi_0^+ = \sum_{k\geq 1} \psi_k + \frac{1}{2}\psi_0$, and more generally $\Psi_j^+ = \sum_{k\geq j} \psi_j$ for $j \geq 1$; eventually one can just replicate the argument for the function $\Psi_0^- = \sum_{k\leq -1} \psi_k + \frac{1}{2}\psi_0$, and summing the bounds gives control for $\tilde{\Delta}_{B_I}(J)$. Start with the following estimate, which only uses the triangle inequality, and the fact that $\frac{1}{2}\psi_0$ is a C-Lipschitz function supported on B_I:

$$\left|\frac{1}{\mu(\varphi_{B_I})}\int \Psi_0^+\, d\mu - \frac{1}{\nu(\varphi_{B_I})}\int \Psi_0^+\, d\nu\right| \leq C\alpha(B_I) + \frac{\mu(\varphi_{B_{J_+}})}{\mu(\varphi_{B_I})}\left|\frac{1}{\mu(\varphi_{B_{J_+}})}\int \Psi_1^+\, d\mu - \frac{1}{\nu(\varphi_{B_{J_+}})}\int \Psi_1^+\, d\nu\right| \\ + \left(\frac{1}{\nu(\varphi_{B_{J_+}})}\int \Psi_1^+\, d\nu\right)\left|\frac{\mu(\varphi_{B_{J_+}})}{\mu(\varphi_{B_I})} - \frac{\nu(\varphi_{B_{J_+}})}{\nu(\varphi_{B_I})}\right|. \quad (5.16)$$

Here

$$\frac{1}{\nu(\varphi_{B_{J_+}})}\int \Psi_1^+\, d\nu \lesssim 1,$$

since ν is doubling and Ψ_1^+ vanishes outside $J_+ \subset B_{J_+}$, and

$$\left|\frac{\mu(\varphi_{B_{J_+}})}{\mu(\varphi_{B_I})} - \frac{\nu(\varphi_{B_{J_+}})}{\nu(\varphi_{B_I})}\right| \leq \frac{|B_I|}{|B_{J_+}|}\cdot \alpha(B_I) \lesssim \alpha(B_I),$$

since $\varphi_{B_{J_+}} = (\varphi_{B_{J_+}} \circ T_{B_I}^{-1}) \circ T_{B_I}$, where $\varphi_{B_{J_+}} \circ T_{B_I}^{-1}$ is a $(|B_I|/|B_{J_+}|)$-Lipschitz function supported on [0, 1]. Consequently,

$$\left|\frac{1}{\mu(\varphi_{B_I})}\int \Psi_0^+\, d\mu - \frac{1}{\nu(\varphi_{B_I})}\int \Psi_0^+\, d\nu\right|\mu(\varphi_{B_I}) \leq C\alpha(B_I)\mu(\varphi_{B_I}) \\ + \left|\frac{1}{\mu(\varphi_{B_{J_+}})}\int \Psi_1^+\, d\mu - \frac{1}{\nu(\varphi_{B_{J_+}})}\int \Psi_1^+\, d\nu\right|\mu(\varphi_{B_{J_+}}).$$

Here Ψ_1^+ vanishes outside $J_+ \subset B_{J_+}$, so the estimate can be iterated. After $N \geq 0$ repetitions (the case $N = 0$ was seen above), one ends up with

$$\left|\frac{1}{\mu(\varphi_{B_I})}\int \Psi_0^+\, d\mu - \frac{1}{\nu(\varphi_{B_I})}\int \Psi_0^+\, d\nu\right|\mu(\varphi_{B_I}) \\ \leq C\sum_{k=0}^{N}\alpha(B_{J_{k+}})\mu(\varphi_{B_{J_{k+}}}) \\ + \mu(\varphi_{B_{J_{(N+1)+}}})\left|\frac{1}{\mu(\varphi_{B_{(N+1)+}})}\int \Psi_{N+1}^+\, d\mu - \frac{1}{\nu(B_{(N+1)+})}\int \Psi_{N+1}^+\, d\nu\right|, \quad (5.17)$$

where one needs to interpret $J_{0+} = I$ (which is different from J in the case $J = I_-$). What is a good choice for N? Let $N_1 \geq 0$ be the smallest number such that $J_{(N_1+1)+} \in \mathrm{Leaves}(\mathcal{T})$. If there is no such number, let $N_1 = \infty$. In the case $N_1 = \infty$, the term in the last line of (5.17) vanishes, since $\mu(B_{J_{N+}})$ decays rapidly as long as $N \in \mathcal{T}$ (using the doubling of ν, and the fact that $\alpha(B_I) \leq \epsilon$ for $I \in \mathcal{T}$). If $N_1 < \infty$, the term in the last line of (5.17) is clearly bounded by $\leq 2\mu(B_{J_{(N_1+1)+}})$, since $\Psi_{N_1+1}^+$ vanishes outside $J_{(N_1+1)+}$, which is well inside $B_{(N_1+1)+}$. Observing that also $\mu(I) \lesssim \mu(\varphi_{B_I})$, it follows that

$$\left|\frac{1}{\mu(\varphi_{B_I})}\int \Psi_0^+\, d\mu - \frac{1}{\nu(\varphi_{B_I})}\int \Psi_0^+\, d\nu\right|\mu(I) \lesssim \sum_{k=0}^{N_1}\alpha(B_{J_{k+}})\mu(B_{J_{k+}}) + \mu(B_{J_{(N_1+1)+}}).$$

Finally, by symmetry, the same argument can be carried out for the series $\Psi_0^- = \sum_{k<0} \psi_k + \frac{1}{2}\psi_0$. If $N_2 \geq 0$ is the smallest number such that $J_{(N_2+1)-} \in \mathrm{Leaves}(\mathcal{T})$, this leads to the following analogue of the Tail-Tip inequality:

$$\tilde{\Delta}_{B_I}(J)\mu(I) \lesssim \sum_{P \in \mathrm{Tail}_J} \alpha(B_P)\mu(B_P) + \mu(\mathrm{Tip}_J), \quad J \in \{I, I_-\},\ I \in \mathcal{T} \setminus \mathrm{Leaves}(\mathcal{T}). \tag{5.18}$$

Here Tail_J is the collection of dyadic intervals $\mathrm{Tail}_J = \{J_{N_2-}, \dots, J, \dots, J_{N_1+}\} \subset \mathcal{T} \setminus \mathrm{Leaves}(\mathcal{T})$, and $\mathrm{Tip}_J = B_{J_{(N_2+1)-}} \cup B_{J_{(N_1+1)+}}$. Finally, in the excluded special case, where $J = I_- \in \mathrm{Leaves}(\mathcal{T})$ (recall (5.13)), the same estimate holds if one defines $\mathrm{Tail}_J = \varnothing$ and $\mathrm{Tip}_J := J$ (noting that $I \in \mathcal{T}$, so $\mu(I) \lesssim \mu(J)$).

Armed with the Tail-Tip inequality (5.18), the proof of the main estimate (5.9) is a replica of the argument in the dyadic case, namely the proof of Proposition 2.4. I only sketch the details. For $I \in \mathcal{T} \setminus \mathrm{Leaves}(\mathcal{T})$ and $J \in \{I, I_-\}$, start with

$$\begin{aligned}\tilde{\Delta}^2_{B_I}(J)\mu(I) &\lesssim \sum_{P \in \mathrm{Tail}_J} \alpha^2(B_P)\frac{\mu(B_P)^{3/2}}{\mu(I)^{1/2}} + \frac{\mu(\mathrm{Tip}_J)^2}{\mu(I)} \\ &\leq \sum_{\substack{P \in \mathcal{T} \setminus \mathrm{Leaves}(\mathcal{T}) \\ P \subset I}} \alpha^2(B_P)\frac{\mu(B_P)^{3/2}}{\mu(I)^{1/2}} + \frac{\mu(\mathrm{Tip}_J)^2}{\mu(I)}.\end{aligned}$$

The second inequality is trivial, and the first is proved with the same Cauchy–Schwarz argument as (2.17), using the fact that $\sum_{P \in \mathrm{Tail}_J} \mu(B_P)^{1/2} \lesssim \mu(I)^{1/2}$, which follows from $\mathrm{Tail}_J \subset \mathcal{T} \setminus \mathrm{Leaves}(\mathcal{T})$, and in particular the geometric decay of the measures $\mu(B_P)$ for $P \in \mathcal{T} \setminus \mathrm{Leaves}(\mathcal{T})$. Now, the inequality above can be summed for $I \in \mathcal{T} \setminus \mathrm{Leaves}(\mathcal{T})$ precisely as in the proof of (2.18). In particular, one should first use the estimate

$$\mu(\mathrm{Tip}_J) \leq \mu(B_{J_{(N_2+1)-}}) + \mu(B_{J_{(N_1+1)+}}) \lesssim \mu(J_{(N_2+1)-}) + \mu(J_{(N_1+1)+}),$$

which follows from $\alpha(B_{J_{N_1+}}), \alpha(B_{J_{N_2-}}) < \epsilon$, if ϵ is small enough, depending on the doubling constant of ν. The conclusion is

$$\sum_{I \in \mathcal{T} \setminus \mathrm{Leaves}(\mathcal{T})} \tilde{\Delta}^2_{B_I}(J)\mu(I) \lesssim \sum_{P \in \mathcal{T} \setminus \mathrm{Leaves}(\mathcal{T})} \alpha^2(B_P)\mu(B_P) + \mu(\mathrm{Leaves}(\mathcal{T}))$$

for $J \in \{I, I_-\}$. As observed in and around (5.11), this implies (5.9).

Remark 5.19. In the proof of (5.9), the uniform bound $\alpha(B_I) < \epsilon$, $I \in \mathcal{T} \setminus \mathrm{Leaves}(\mathcal{T})$, was only used to guarantee that μ is sufficiently doubling along, and inside, the balls B_I. If such properties are assumed a priori in some given tree $\mathcal{T}$, then (5.9) continues to hold for $\mathcal{T}$. In particular, if μ is doubling on the whole real line, and the Carleson condition

$$\int_{B(x,2r)} \int_0^{2r} \alpha^2_{\mu,\nu}(B(y,t))\, \frac{dt\, d\mu y}{t} \leq C\mu(B(x,r))$$

holds, then the dyadic Carleson condition of Theorem 1.12 holds for any dyadic system $\mathcal{D}$ (a family of half-open intervals covering $\mathbb{R}$, where every interval has length of the form 2^{-k} for some $k \in \mathbb{Z}$, and every interval is the union of two further intervals in the family; the proof of Theorem 1.12 seen in Section 2 works for any such system). It follows from this that $\mu \in A_\infty^{\mathcal{D}}(\nu)$ for every dyadic system $\mathcal{D}$, and consequently $\mu \in A_\infty(\nu)$. (To see this, pick a finite collection $\mathcal{D}_1, \dots, \mathcal{D}_N$ of dyadic systems such that the max of the corresponding dyadic maximal functions $M_\nu^{\mathcal{D}_i}$,

$$M_\nu^{\mathcal{D}_i} f(x) = \sup_{x \in I \in \mathcal{D}_i} \frac{1}{\nu(I)} \int_I |f| \, d\nu,$$

bounds the usual Hardy–Littlewood maximal function M_ν, up to a constant depending only on the doubling of ν. The construction of such systems is well known, and in $\mathbb{R}$ as few as two systems do the trick; for a reference, see for instance Section 5 in [Muscalu et al. 2002]. Then, for every $1 \le i \le N$, there exists $p_i < \infty$ such that $\mu \in A_{p_i}^{\mathcal{D}_i}(\nu)$; see [Grafakos 2014, Theorem 9.33(f)]. In particular $\mu \in A_p^{\mathcal{D}_i}(\nu)$ for $p := \max p_i$, and hence $\|M_\nu^{\mathcal{D}_i}\|_{L^p(\mu) \to L^p(\mu)} < \infty$ for $1 \le i \le N$. It follows that $\|M_\nu\|_{L^p(\mu) \to L^p(\mu)} < \infty$, which is one possible definition for $\mu \in A_\infty(\nu)$. For much more information, see [Grafakos 2014, Section 9.11].) This proves the “continuous” part of Theorem 1.12.

6. Parts (a) of the main theorems

Parts (a) of Theorems 1.8 and 1.9 are proved in this section: $\mathcal{S}_{\mathcal{D},\nu}(\mu)$ and $\mathcal{S}_\nu(\mu)$ are finite μ_a almost everywhere, where μ_a is the absolutely continuous part of μ relative to ν. The strategy is to prove the statement first for the dyadic square function $\mathcal{S}_{\mathcal{D},\nu}(\mu)$, but allow $\mathcal{D}$ to be a slightly generalised system: a family $\mathcal{D} = \bigcup \mathcal{D}_k$, $k \ge 0$, of half-open intervals of length at most 1 such that

(D1) each $\mathcal{D}_k$ is a partition of $\mathbb{R}$,

(D2) each interval in $\mathcal{D}_k$ has length 2^{-k}, and

(D3) each interval $I \in \mathcal{D}_k$ has two children in $\mathcal{D}_{k+1}$, denoted by $\mathrm{ch}(I)$.

The added generality makes no difference in the proof, which closely follows previous arguments of Tolsa [2009; 2015]. The benefit is that the nondyadic square function $\mathcal{S}_\nu(\mu)$ can, eventually, be bounded by a finite sum of dyadic square functions $\mathcal{S}_{\mathcal{D}_1,\nu}(\mu), \dots, \mathcal{S}_{\mathcal{D}_N,\nu}(\mu)$, so the nondyadic problem easily reduces to the dyadic one.

With the strategy in mind, fix a dyadic system $\mathcal{D}$ satisfying (D1)–(D3), and let $\mathcal{S}_{\mathcal{D},\nu}(\mu)$ be the associated square function.

Lemma 6.1. *Assume that μ, ν are Radon measures on $\mathbb{R}$, with μ finite, and ν dyadically doubling (relative to $\mathcal{D}$). Then $\mathcal{S}_\nu(\mu)$ is finite μ_a almost surely.*

The proof of Lemma 6.1 is a combination of two arguments of Tolsa: the proofs of [Tolsa 2009, Theorem 1.1] and [Tolsa 2015, Lemma 2.2]. I start with an analogue of the first:

Lemma 6.2. *Assume that $\mu \in L^2(\nu)$. Then*

$$\sum_{\substack{I \in \mathcal{D} \\ \nu(I) > 0}} \alpha_{\mu,\nu}^2(I) \frac{\mu(I)^2}{\nu(I)} \lesssim \|\mu\|_{L^2(\nu)}^2.$$

Proof. It suffices to sum over the intervals $I \subset \mathcal{D}$ with $\mu(I) > 0$ and $\nu(I) > 0$; fix one of these I, and a 1-Lipschitz function $\psi : \mathbb{R} \to \mathbb{R}$, supported on $[0, 1]$. Then, write

$$\left| \int \psi \, d\mu_I - \int \psi \, d\nu_I \right| = \left| \frac{1}{\mu(I)} \int_I (\psi \circ T_I) g \, d\nu - \frac{1}{\nu(I)} \int_I (\psi \circ T_I) \, d\nu \right|, \tag{6.3}$$

where g is the Radon–Nikodym derivative $d\mu/d\nu \in L^2(\nu)$. Express $g\chi_I$ in terms of standard (ν-adapted) martingale differences:

$$g\chi_I = \langle g \rangle_I^\nu \chi_I + \sum_{J \in \mathcal{D}(I)} \Delta_J^\nu g, \tag{6.4}$$

where $\mathcal{D}(I) := \{J \in \mathcal{D} : J \subset I\}$, the sum converges in $L^2(\nu)$, and

$$\langle g \rangle_I^\nu = \frac{1}{\nu(I)} \int g \, d\nu = \frac{\mu(I)}{\nu(I)} \quad \text{and} \quad \Delta_J^\nu g = -\langle g \rangle_J^\nu \chi_J + \sum_{J' \in \mathrm{ch}(J)} \langle g \rangle_{J'}^\nu \chi_{J'}.$$

Note that $\Delta_J^\nu g$ is supported on J and has ν-mean zero. By (6.4),

$$\frac{1}{\mu(I)} \int_J (\psi \circ T_I) g \, d\nu = \frac{1}{\nu(I)} \int_I (\psi \circ T_I) \, d\nu + \sum_{J \in \mathcal{D}(I)} \frac{1}{\mu(I)} \int_J (\psi \circ T_I) \Delta_J^\nu g \, d\nu. \tag{6.5}$$

Since the first term on the right-hand side of (6.5) cancels out the last term in (6.3), one can continue as follows:

$$\begin{aligned} (6.3) &\leq \sum_{J \in \mathcal{D}(I)} \frac{1}{\mu(I)} \left| \int_J (\psi \circ T_I) \Delta_J^\nu g \, d\nu \right| \\ &= \sum_{J \in \mathcal{D}(I)} \frac{1}{\mu(I)} \left| \int_J [(\psi \circ T_I) - (\psi \circ T_I(x_J))] \Delta_J^\nu g \, d\nu \right|. \end{aligned}$$

Above, x_J is the midpoint of J, and the zero-mean property of $\Delta_J^\nu g$ was used. Finally, recalling that ψ is 1-Lipschitz, one obtains

$$(6.3) \leq \sum_{J \in \mathcal{D}(I)} \frac{\ell(T_I(J))}{\mu(I)} \|\Delta_J^\nu g\|_{L^1(\nu)} \leq \sum_{J \in \mathcal{D}(I)} \frac{\ell(J)\nu(J)^{1/2}}{\mu(I)\ell(I)} \|\Delta_J^\nu g\|_{L^2(\nu)}.$$

Taking a sup over admissible functions $\psi : \mathbb{R} \to \mathbb{R}$ gives

$$\alpha_{\mu,\nu}(I) \leq \sum_{J \in \mathcal{D}(I)} \frac{\ell(J)\nu(J)^{1/2}}{\mu(I)\ell(I)} \|\Delta_J^\nu g\|_{L^2(\nu)}. \tag{6.6}$$

Now, using (6.6) and Cauchy–Schwarz, we may sum over $I \in \mathcal{D}$ as follows (we suppress the requirement $\nu(I) > 0$ from the notation):

$$\begin{aligned} \sum_{I \in \mathcal{D}} \alpha_{\mu,\nu}^2(I) \frac{\mu(I)^2}{\nu(I)} &\leq \sum_{J \in \mathcal{D}} \left(\sum_{J \in \mathcal{D}(I)} \frac{\ell(J)\nu(J)^{1/2}}{\ell(I)} \|\Delta_J^\nu g\|_{L^2(\nu)} \right)^2 \frac{1}{\nu(I)} \\ &\leq \sum_{I \in \mathcal{D}} \left(\sum_{J \in \mathcal{D}(I)} \frac{\ell(J)}{\ell(I)} \|\Delta_J^\nu g\|_{L^2(\nu)}^2 \right) \sum_{J \in \mathcal{D}(I)} \frac{\ell(J)\nu(J)}{\ell(I)\nu(I)}. \end{aligned}$$

Clearly,

$$\sum_{J \in \mathcal{D}(I)} \frac{\ell(J)\nu(J)}{\ell(I)\nu(I)} \lesssim 1,$$

so

$$\sum_{J \in \mathcal{D}} \alpha_{\mu,\nu}(I)^2 \frac{\mu(I)^2}{\nu(I)} \lesssim \sum_{J \in \mathcal{D}} \|\Delta_J^\nu g\|_{L^2(\nu)}^2 \sum_{I \supset J} \frac{\ell(J)}{\ell(I)} \lesssim \sum_{J \in \mathcal{D}} \|\Delta_J^\nu g\|_{L^2(\nu)}^2 \le \|g\|_{L^2(\nu)}^2,$$

as claimed. □

Corollary 6.7. *If $\mu \in L^2(\nu)$, then $\mathcal{S}_{\mathcal{D},\nu}(\mu)$ is finite μ almost everywhere.*

Proof. By Lemma 6.2, and the Lebesgue differentiation theorem, the following conditions hold μ almost everywhere:

$$\sum_{x \in I \in \mathcal{D}} \alpha_{\mu,\nu}(I)^2 \frac{\mu(I)}{\nu(I)} < \infty \quad \text{and there exists} \quad \lim_{I \to x} \frac{\mu(I)}{\nu(I)} = \mu(x) > 0.$$

Clearly $\mathcal{S}_{\mathcal{D},\nu}(\mu)(x) < \infty$ for such $x \in [0, 1)$. □

Now, we can prove Lemma 6.1 by an argument similar to [Tolsa 2015, Lemma 2.2]:

Proof of Lemma 6.1. Perform a Calderón–Zygmund decomposition of μ with respect to ν, at some level $\lambda \ge 1$. More precisely, let $\mathcal{B}$ be the family of maximal intervals $I \in \mathcal{D}$ with $\mu(I) > \lambda\nu(I)$, and set $\mu = g + b$, where

$$g = \mu|_G + \sum_{I \in \mathcal{B}} \frac{\mu(I)}{\nu(I)} \nu|_I, \quad G := [0, 1) \setminus \bigcup_{I \in \mathcal{B}} I,$$

and

$$b = \sum_{I \in \mathcal{B}} \left[\mu|_I - \frac{\mu(I)}{\nu(I)} \nu|_I \right] =: \sum_{I \in \mathcal{B}} b_I.$$

Then $\|g\|_{L^\infty(\nu)} \lesssim \lambda$ (the implicit constants depend on the doubling of ν), and

$$\nu([0, 1) \setminus G) = \sum_{I \in \mathcal{B}} \nu(I) < \frac{1}{\lambda} \sum_{I \in \mathcal{B}} \mu(I) \le \frac{1}{\lambda}.$$

Since $\mu_a \in L^1(\nu)$ (recall that μ is a finite measure), it follows that $\mu_a([0, 1) \setminus G) \to 0$ as $\lambda \to \infty$. Hence, it suffices to show that

$$\mathcal{S}_{\mathcal{D},\nu}(\mu)(x) < \infty \quad \text{for } \mu \text{ almost every } x \in G \cap \operatorname{spt}_{\mathcal{D}} \mu,$$

where $\operatorname{spt}_{\mathcal{D}} \mu = \{x \in \mathbb{R} : \mu(I) > 0 \text{ for all } x \in I \in \mathcal{D}\}$. Let $\mathcal{G} \subset \mathcal{D}$ be the intervals, which are not contained in any interval in $\mathcal{B}$. Fix $x \in G \cap \operatorname{spt}_{\mathcal{D}} \mu$, and note that if $x \in I \in \mathcal{D}$, then $I \in \mathcal{G}$. Observe that $\mu(I) = g(I)$ for $I \in \mathcal{G}$, and consequently

$$\begin{aligned} \left| \int \psi \, d\mu_I - \int \psi \, d\nu_I \right| &\le \left| \int \psi \, d\mu_I - \int \psi \, dg_I \right| + \alpha_{g,\nu}(I) \\ &= \frac{1}{\mu(I)} \left| \int_I (\psi \circ T_I) \, db \right| + \alpha_{g,\nu}(I), \quad I \ni x, \end{aligned}$$

for any 1-Lipschitz function $\psi \colon \mathbb{R} \to \mathbb{R}$ supported on $[0,1]$. Using the zero-mean property of the measures b_J, one can estimate further as follows:

$$\left|\int (\psi \circ T_I)\, db\right| \le \sum_{J \in \mathcal{B}(I)} \left|\int (\psi \circ T_I)\, db_J\right| = \sum_{J \in \mathcal{B}(I)} \left|\int [(\psi \circ T_I) - (\psi \circ T_I(x_J)]\, db_J\right|,$$

where $\mathcal{B}(I) := \{J \in \mathcal{B} : J \subset I\}$, and x_J is the midpoint of J. Using the fact that ψ is 1-Lipschitz, one has

$$\frac{1}{\mu(I)}\left|\int_I (\psi \circ T_I)\, db\right| \le \frac{1}{\mu(I)}\left|\int [(\psi \circ T_I) - (\psi \circ T_I(x_J)]\, db_J\right| \le \frac{\ell(T_I(J))}{\mu(I)}\|b_J\| \lesssim \frac{\ell(J)\mu(J)}{\ell(I)\mu(I)},$$

and finally

$$\mathcal{S}^2_{\mathcal{D},\nu}(\mu)(x) \lesssim \mathcal{S}_{\mathcal{D},\nu}(g)^2(x) + \sum_{x \in I \in \mathcal{G}} \left(\sum_{J \in \mathcal{B}(I)} \frac{\ell(J)\mu(J)}{\ell(I)\mu(I)}\right)^2 =: \mathcal{S}_{\mathcal{D},\nu}(g)^2(x) + S^2(x).$$

Since $\mathcal{S}_{\mathcal{D},\nu}(g)$ is finite g almost everywhere by Corollary 6.7, and in particular $\mathcal{S}_{\mathcal{D},\nu}(g)(x) < \infty$ for μ almost every $x \in G$, it remains to prove that $S(x) < \infty$ for μ almost every $x \in \mathbb{R}$. First, note that

$$\sum_{J \in \mathcal{B}(I)} \frac{\ell(J)\mu(J)}{\ell(I)\mu(I)} \le \frac{1}{\mu(I)} \sum_{J \in \mathcal{B}(I)} \mu(J) \le 1,$$

as the intervals in $\mathcal{B}(I)$ are disjoint. Consequently,

$$\begin{aligned}\int S^2\, d\mu &\le \int \sum_{x \in I \in \mathcal{G}} \sum_{J \in \mathcal{B}(I)} \frac{\ell(J)\mu(J)}{\ell(I)\mu(I)}\, d\mu(x) = \sum_{I \in \mathcal{G}} \sum_{J \in \mathcal{B}(I)} \frac{\ell(J)\mu(J)}{\ell(I)} \\ &= \sum_{J \in \mathcal{B}} \mu(J) \sum_{J \subset I \in \mathcal{G}} \frac{\ell(J)}{\ell(I)} \lesssim \sum_{J \in \mathcal{B}} \mu(J) \le \|\mu\| < \infty.\end{aligned}$$

It follows that $S^2(x) < \infty$ for μ almost every $x \in \mathbb{R}$. This completes the proof of Lemma 6.1, and Theorem 1.8(a). □

Bounding the nondyadic square function. It remains to prove Theorem 1.9(a). Assume that μ, ν are Radon measures on $\mathbb{R}$, with ν doubling, and recall that $\mathcal{S}_\nu(\mu)$ is the square function

$$\mathcal{S}^2_\nu(\mu)(x) = \int_0^1 \alpha^2_{s,\mu,\nu}(B(x,r))\, \frac{dr}{r}, \quad x \in \mathbb{R}.$$

The claim is that $\mathcal{S}_\nu(\mu)$ is finite μ_a almost everywhere; since this is a local problem, one may assume that μ is a finite measure. Now, as in Remark 5.19 (or see [Muscalu et al. 2002, Section 5]), pick a finite number of dyadic systems $\mathcal{D}_1, \dots, \mathcal{D}_N$ with the following property: for any interval $I \subset \mathbb{R}$, there exists $j \in \{1, \dots, N\}$, depending on I, and an interval $J \in \mathcal{D}_j$ such that $I \subset J_i$ and $|J_i| \sim |I|$. As a little technical point, we actually need to restrict $\mathcal{D}_j$ to intervals of length at most 1, so also the defining property above only holds for intervals $I \subset \mathbb{R}$ of length $|I| \le r_0$, say.

Then, apply Lemma 6.1 to each of the corresponding square functions $\mathcal{S}_{\mathcal{D}_j,\nu}(\mu)$ to infer

$$\mathcal{S}_{\mathcal{D},\nu}(\mu)(x) := \sum_{j=1}^{N} \mathcal{S}_{\mathcal{D}_j,\nu}(\mu)(x) < \infty$$

for μ_a almost every $x \in \mathbb{R}$ (note that ν is dyadically doubling relative to every $\mathcal{D}_j$). So, it suffices to argue that $\mathcal{S}_{\mathcal{D},\nu}(\mu)$ dominates $\mathcal{S}_\nu(\mu)$. Using the stability of the smooth α-numbers, and the fact that they are dominated by the regular α-numbers whenever ν is doubling (see Proposition 5.4), one has

$$\alpha^2_{s,\mu,\nu}(B(x,r)) \lesssim \alpha^2_{\mu,\nu}(I^j_{x,r}), \quad x \in \mathbb{R},\ 0 < r < r_0,$$

where $j \in \{1, \dots, N\}$, and $I^j_{x,r} \in \mathcal{D}_j$ is a dyadic interval of length at most 1, satisfying $x \in B(x,r) \subset I_{x,r}$ and $|I_{x,r}| \sim r$. The existence follows from the construction of the systems $\mathcal{D}_j$. It is now clear that $\mathcal{S}_\nu(\mu) \lesssim \mathcal{S}_{\mathcal{D},\nu}(\mu)$, and the proof of Theorem 1.9(a) is complete.

Remark 6.8. Lemma 5.4 in [Azzam et al. 2017] implies

$$\int_{1/4}^{1/2} \alpha_{\mu,\nu}(B(0,t))\,dt \lesssim \alpha_{s,\mu,\nu}(B(0,1)),$$

whenever ν is doubling, and $\nu\big(B\big(0,\tfrac{1}{4}\big)\big) > 0$, $\mu\big(B\big(0,\tfrac{1}{4}\big)\big) > 0$. So, at the level of L^1-averages over scales, the smooth and regular α-numbers are comparable. One would need a similar comparison at the level of L^2-averages to answer Question 1.

Acknowledgements

I am grateful to Jonas Azzam, David Bate, and Antti Käenmaki for useful discussions during the preparation of the manuscript. I would also like to thank the referees for good comments, and for asking me to prove parts (a) of Theorems 1.8 and 1.9.

References

[Azzam et al. 2016] J. Azzam, G. David, and T. Toro, “Wasserstein distance and the rectifiability of doubling measures: part I”, *Math. Ann.* **364**:1-2 (2016), 151–224. MR Zbl

[Azzam et al. 2017] J. Azzam, G. David, and T. Toro, “Wasserstein distance and the rectifiability of doubling measures: part II”, *Math. Z.* **286**:3-4 (2017), 861–891. MR Zbl

[Buckley 1993] S. M. Buckley, “Summation conditions on weights”, *Michigan Math. J.* **40**:1 (1993), 153–170. MR Zbl

[Fefferman et al. 1991] R. A. Fefferman, C. E. Kenig, and J. Pipher, “The theory of weights and the Dirichlet problem for elliptic equations”, *Ann. of Math.* (2) **134**:1 (1991), 65–124. MR Zbl

[Grafakos 2014] L. Grafakos, *Modern Fourier analysis*, 3rd ed., Graduate Texts in Mathematics **250**, Springer, 2014. MR Zbl

[Muscalu et al. 2002] C. Muscalu, T. Tao, and C. Thiele, “Multi-linear operators given by singular multipliers”, *J. Amer. Math. Soc.* **15**:2 (2002), 469–496. MR Zbl

[Tolsa 2009] X. Tolsa, “Uniform rectifiability, Calderón–Zygmund operators with odd kernel, and quasiorthogonality”, *Proc. Lond. Math. Soc.* (3) **98**:2 (2009), 393–426. MR Zbl

[Tolsa 2015] X. Tolsa, “Characterization of n-rectifiability in terms of Jones’ square function: part I”, *Calc. Var. Partial Differential Equations* **54**:4 (2015), 3643–3665. MR Zbl

Received 15 Mar 2017. Revised 12 Jun 2018. Accepted 14 Jul 2018.

TUOMAS ORPONEN: tuomas.orponen@helsinki.fi
Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland

ANALYSIS AND PDE
Vol. 12, No. 4, 2019
dx.doi.org/10.2140/apde.2019.12.997

GLOBAL WELL-POSEDNESS FOR THE TWO-DIMENSIONAL MUSKAT PROBLEM WITH SLOPE LESS THAN 1

STEPHEN CAMERON

We prove the existence of global, smooth solutions to the two-dimensional Muskat problem in the stable regime whenever the product of the maximal and minimal slope is less than 1. The curvature of these solutions decays to 0 as t goes to infinity, and they are unique when the initial data is $C^{1,\epsilon}$. We do this by getting a priori estimates using a nonlinear maximum principle first introduced in a paper by Kiselev, Nazarov, and Volberg (2007), where the authors proved global well-posedness for the surface quasigeostraphic equation.

1. Introduction

The Muskat problem was originally introduced in [Muskat 1934] in order to model the interface between water and oil in tar sands. In general, it describes the interface between two incompressible, immiscible fluids of different constant densities in a porous media. The fluids evolve according to Darcy's law, giving an evolution of the interface (see [Córdoba and Gancedo 2007] for derivation of equations), and the problem in two dimensions is analogous to the two-phase Hele-Shaw cell (see [Saffman and Taylor 1958]). In the case that the two fluids are of equal viscosity and the interface is given by the graph $y = f(t, x)$ with the denser fluid on bottom (i.e., the stable regime), the function f satisfies

$$f_t(t,x) = \int_{\mathbb{R}} \frac{(f_x(t,y) - f_x(t,x))(y-x)}{(f(t,y) - f(t,x))^2 + (y-x)^2}\, dy, \tag{1-1}$$

after the appropriate renormalization. By making a change of variables, see the proof of Lemma 5.1 of [Córdoba and Gancedo 2009], we get the equivalent system

$$f_t(t,x) = \int_{\mathbb{R}} \frac{f(t,y) - f(t,x) - (y-x) f_x(t,x)}{(f(t,y) - f(t,x))^2 + (y-x)^2}\, dy, \tag{1-2}$$

which will be more useful for our purposes. Since the function f is Lipschitz, the above integral can be viewed as a nonlinear perturbation of the half Laplacian. In fact, it is easy to see that linearizing around a flat solution gives

$$f_t(t,x) = -c(-\Delta)^{1/2} f(t,x), \tag{1-3}$$

demonstrating the natural parabolicity of the problem.

MSC2010: 35K55, 35Q35, 35R09.
Keywords: Muskat problem, porous media, fluid interface, global well-posedness.

The Muskat problem is known to be locally well-posed in H^k for $k \geq 3$ with solutions satisfying L^∞ and L^2 maximum principles, but neither imply any gain of derivatives; see [Córdoba and Gancedo 2009; Constantin et al. 2013].

Under the assumption $\|f_0'\|_{L^\infty} < 1$, there have been a number of positive results. In [Constantin et al. 2013] the authors proved an L^∞ maximal principle for the slope f_x along with the existence of global weak Lipschitz solutions using a regularized system. Recently, [Gancedo 2017] improved the L^2 energy estimate of [Constantin et al. 2013] (which holds for any solution) to one analogous with the energy estimate from the linear equation under this assumption on the slope. When the initial data f_0 is in $H^2(\mathbb{R})$ with $\|f_0\|_1 = \||\xi|\hat{f}_0(\xi)\|_{L^1_\xi}$ less than some explicit constant $\approx \frac{1}{3}$ (which implies slope less than 1), [Constantin et al. 2016] proved that a unique global strong solution exists. In this case [Patel and Strain 2017] proved optimal decay estimates on the norms $\|f(t,\cdot)\|_s = \||\xi|^s \hat{f}(t,\xi)\|_{L^1_\xi}$, matching the estimates for the linear equation.

Recently, [Deng et al. 2017] was also able to prove the existence of global weak solutions for arbitrarily large monotonic initial data. They did this using the regularized system from [Constantin et al. 2013] to prove that both f and f_x still obey the maximum principle under this monotonicity assumption.

Because solutions to (1-2) have the natural scaling $(1/r)f(rt,rx)$, we see that L^∞ or sign bounds on the slope f_x are scale-invariant properties. We fit these two types of assumptions into the same framework by showing that the critical quantity is in fact the product of the maximal and minimal slopes,

$$\beta(f_0') := \Big(\sup_x f_0'(x)\Big)\Big(\sup_y -f_0'(y)\Big). \tag{1-4}$$

As we shall see in Section 3, the derivative f_x obeys the equation

$$(f_x)_t(t,x) = f_{xx}(t,x)\int_{\mathbb{R}} \frac{-h}{\delta_h f(t,x)^2 + h^2}\,dh + \int_{\mathbb{R}} \delta_h f_x(t,x) K(t,x,h)\,dh, \tag{1-5}$$

where $\delta_h f(t,x) := f(t,x+h) - f(t,x)$ and the kernel K is uniformly elliptic of order 1 whenever $\beta(f_0') < 1$. Thus we naturally get regularizing effects from the equation whenever the initial data satisfies this bound. It's clear that $\|f_0'\|_{L^\infty} < 1$ implies $\beta(f_0') < 1$, and for bounded monotonic data we get $\beta(f_0') = 0$ since either $\sup f_0' = 0$ or $\inf f_0' = 0$. Thus this $\beta(f_0') < 1$ provides a natural interpolation between these two types of assumptions.

In contrast to the positive results, [Castro et al. 2012] showed that there is an open subset of initial data in H^4 such that the Rayleigh–Taylor condition breaks down in finite time. That is,

$$\lim_{t \to t_0-} \|f_x(t,\cdot)\|_{L^\infty} = \infty$$

for some time t_0, after which the interface between the fluids can no longer be described by a graph.

The authors of [Constantin et al. 2017] made great progress towards proving global regularity. They proved that if the initial data $f_0 \in H^k$, then the solution f will exist and remain in H^k so long as the slope $f_x(t,\cdot)$ remains bounded and uniformly continuous. Thus the natural next step is to prove the generation of a modulus of continuity for f_x.

Theorem 1.1. *Let $f_0 \in W^{1,\infty}(\mathbb{R})$ with*

$$\beta(f_0') := \big(\sup_x f_0'(x)\big)\big(\sup_y -f_0'(y)\big) < 1. \tag{1-6}$$

Then there exists a classical solution

$$f \in C([0,\infty)\times\mathbb{R}) \cap C^{1,\alpha}_{\mathrm{loc}}((0\,\infty)\times\mathbb{R}) \cap L^\infty_{\mathrm{loc}}((0,\infty); C^{1,1}) \tag{1-7}$$

to (1-2) with f_x satisfying both the maximum principle and

$$f_x(t,x) - f_x(t,y) \le \rho\left(\frac{|x-y|}{t}\right), \quad t > 0,\ x \ne y \in \mathbb{R}, \tag{1-8}$$

for some Lipschitz modulus of continuity ρ depending solely on $\beta(f_0')$, $\|f_0'\|_{L^\infty}$. In the case that $f_0 \in C^{1,\epsilon}(\mathbb{R})$ for some $\epsilon > 0$, the solution f is unique with $f \in L^\infty([0,\infty); C^{1,\epsilon})$.

The uniqueness statement follows essentially from the uniqueness theorem of [Constantin et al. 2017]. We note in the Appendix the few small changes needed to their proof in order to apply it here.

The most vital part of Theorem 1.1 is the spontaneous generation of the modulus $\rho(\cdot/t)$, as everything else will follow from that. The spontaneous generation/propagation of a general modulus of continuity has old roots as classical Hölder estimates, but it's only recently that the idea to tailor-make moduli for specific equations emerged. The technique first appeared in [Kiselev et al. 2007], where the authors used it to prove global well-posedness for the surface quasigeostraphic equation. It has had great success at proving regularity for a number of active scalar equations, that is, equations of the form

$$\theta_t + (u\cdot\nabla)\theta + \mathcal{L}\theta = 0, \tag{1-9}$$

where u is a flow depending on θ and $\mathcal{L}$ is some diffusive operator. See [Kiselev 2010; Dabkowski et al. 2014] for a good overview of results using this method.

To date, these tailor-made moduli have only been applied to cases where all the nonlinearity is in the flow velocity u, and the diffusive term $\mathcal{L}$ is rather nice (typically $(-\Delta)^\alpha$, or at least a Fourier multiplier). We will be applying this method to f_x, which solves the active scalar equation (1-5). Note that in this equation, the kernel K defined in (3-4) is a highly nonlinear function of f, f_x. Thus this is the first time the method has been applied in a fully nonlinear equation.

We prove Theorem 1.1 by deriving a priori estimates for smooth solutions to (1-2) with initial data $f_0 \in C_c^\infty(\mathbb{R})$ depending primarily on $\beta(f_0')$, $\|f_0'\|_{L^\infty}$. We prove enough estimates that by approximating in $W^{1,\infty}_{\mathrm{loc}}$ with smooth compactly supported initial data, we get solutions f^ϵ which will converge along subsequences in C^1_{loc} to a solution f solving (1-2) for arbitrary initial data $f_0 \in W^{1,\infty}(\mathbb{R})$ with $\beta(f_0') < 1$.

The rest of the paper is organized as follows. We begin by repeating the breakthrough argument of [Kiselev et al. 2007] in Section 2. In Section 3, we differentiate (1-2) to derive the equation for f_x, showing that it satisfies the maximum principle when $\beta(f_0') < 1$. In Section 4, we state how a modulus of continuity ω interacts with the equation in our main technical lemma. In Sections 5 and 6 we then derive the bounds on the drift and diffusion terms necessary to prove that lemma. In Section 7, we apply our main technical lemma to a specific modulus of continuity, and finally in Section 8 we complete the proof of (1-8) by choosing the correct modulus ρ. In Section 9, we then use (1-8) to prove a few estimates on regularity in time, guaranteeing enough compactness to prove that there are classical solutions for rough initial data. Finally in the Appendix, we give a quick outline for how to modify the uniqueness proof of [Constantin et al. 2017] to work for initial data $f_0 \in C^{1,\epsilon}(\mathbb{R})$ with $\beta(f_0') < 1$.

2. Breakthrough scenario

Assume that $f_0 \in C_c^\infty(\mathbb{R})$, with $\beta(f_0') < 1$, so that there exists a solution $f \in C^1((0, T_+); H^k)$ for k arbitrarily large and some $T_+ > 0$ by [Córdoba and Gancedo 2009]. Note that under the assumption that $\beta(f_0') < 1$, we will show that the maximum principle holds (see Proposition 3.1) and hence $\|f_x\|_{L^\infty([0,T_+)\times\mathbb{R})} \leq \|f_0'\|_{L^\infty}$ is uniformly bounded. Fix a Lipschitz modulus ρ which we will define later. For sufficiently small times, $f_x(t,\cdot)$ will have modulus $\rho(\cdot/t)$ since it is smooth and bounded. It then follows by the main theorem of [Constantin et al. 2017] that as long as $f_x(t,\cdot)$ continues to have modulus $\rho(\cdot/t)$, the solution f will exist with $T_+ > t$.

So, we proceed as in the proof for the quasigeostraphic equation in [Kiselev et al. 2007]. Suppose that $f_x(t,\cdot)$ satisfies (1-8) for all $t < T$. Then by continuity,

$$f_x(T,x) - f_x(T,y) \leq \rho\left(\frac{|x-y|}{T}\right) \quad \text{for all } x \neq y \in \mathbb{R}. \tag{2-1}$$

We first prove that if we have the strict inequality $f_x(T,x) - f_x(T,y) < \rho(|x-y|/T)$, then $f_x(t,\cdot)$ will have modulus $\rho(\cdot/t)$ for $t \leq T + \epsilon$.

Lemma 2.1. *Let $f \in C([0,T_+); C_0^3(\mathbb{R}))$ and $T \in (0, T_+)$. Suppose that $f(T,\cdot)$ satisfies*

$$f_x(T,x) - f_x(T,y) < \rho\left(\frac{|x-y|}{T}\right) \quad \textit{for all } x \neq y \in \mathbb{R}, \tag{2-2}$$

for some Lipschitz modulus of continuity ρ with $\rho''(0) = -\infty$. Then

$$f_x(T+\epsilon,x) - f_x(T+\epsilon,y) < \rho\left(\frac{|x-y|}{T+\epsilon}\right) \quad \textit{for all } x \neq y \in \mathbb{R}, \tag{2-3}$$

for all $\epsilon > 0$ sufficiently small.

Proof. To begin, note that for any compact subset $K \subset \mathbb{R}^2 \setminus \{(x,x) : x \in \mathbb{R}\}$,

$$f_x(T,x) - f_x(T,y) < \rho\left(\frac{|x-y|}{T}\right) \quad \text{for all } (x,y) \in K$$
$$\implies \quad f_x(T+\epsilon,x) - f_x(T+\epsilon,y) < \rho\left(\frac{|x-y|}{T+\epsilon}\right) \quad \text{for all } (x,y) \in K, \tag{2-4}$$

for $\epsilon > 0$ sufficiently small by uniform continuity. So, we only need to focus on pairs (x,y) that are either close to the diagonal, or that are large.

To handle (x,y) near the diagonal, we start by noting that $f(T,\cdot) \in C^3(\mathbb{R})$ and $\rho''(0) = -\infty$. Thus for every x we get

$$|f_{xx}(T,x)| < \frac{\rho'(0)}{T}. \tag{2-5}$$

Since $f \in C([0, T_+); C_0^3(\mathbb{R}))$, we have $f_{xx}(T, x) \to 0$ as $x \to \infty$. Thus we can take the point where $\max_x |f_{xx}(T, x)|$ is achieved to get

$$\|f_{xx}(T, \cdot)\|_{L^\infty} < \frac{\rho'(0)}{T}. \tag{2-6}$$

By the continuity of f_{xx}, we thus have $\|f_{xx}(T+\epsilon, \cdot)\|_{L^\infty} < \rho'(0)/(T+\epsilon)$ for $\epsilon > 0$ sufficiently small. Hence,

$$f_x(T+\epsilon, x) - f_x(T+\epsilon, y) < \rho\left(\frac{|x-y|}{T+\epsilon}\right), \quad |x-y| < \delta, \tag{2-7}$$

for ϵ, δ sufficiently small.

Now let $R_1, R_2 > 0$ be such that

$$\rho(R_1/(T+\epsilon)) > \operatorname{osc}_{\mathbb{R}} f_x(T+\epsilon, \cdot) \tag{2-8}$$

and that $|x| > R_2$ implies

$$|f_x(T+\epsilon, x)| < \frac{1}{2}\rho\left(\frac{\delta}{T+\epsilon}\right) \tag{2-9}$$

for $\epsilon > 0$ sufficiently small. Taking $R = R_1 + R_2$, it's easy to check that $|x| > R$ implies

$$|f_x(T+\epsilon, x) - f_x(T+\epsilon, y)| < \rho\left(\frac{|x-y|}{T+\epsilon}\right) \quad \text{for all } y \neq x. \tag{2-10}$$

Finally, taking

$$K = \{(x, y) \in \mathbb{R}^2 : |x-y| \geq \delta,\ x, y \in \bar{B}_R\},$$

we're done. □

Thus by Lemma 2.1, if f_x was to lose its modulus after time T, we must have that there exist $x \neq y \in \mathbb{R}$ with

$$f_x(T, x) - f_x(T, y) = \rho\left(\frac{|x-y|}{T}\right). \tag{2-11}$$

We will show for a smooth solution f of (1-2) and the correct choice of ρ that in this case

$$\frac{d}{dt}(f_x(t, x) - f_x(t, y))\Big|_{t=T} < \frac{d}{dt}\left(\rho\left(\frac{|x-y|}{t}\right)\right)\Bigg|_{t=T}, \tag{2-12}$$

contradicting the fact that f_x had modulus $\rho(\cdot/t)$ for time $t < T$.

Thus we just need to prove (2-12) to complete the proof of the generation of modulus of continuity (1-8) of Theorem 1.1.

3. Equation for f_x

To begin proving (2-12), we need to examine the equation that f_x solves. Since everything we will be doing is for some fixed time $T > 0$, we will suppress the time variable from now on. Differentiating (1-2), we see that f_x solves

$$(f_x)_t(x) = f_{xx}(x)\int_{\mathbb{R}} \frac{x-y}{(f(y)-f(x))^2+(y-x)^2}\,dy \\ + \int_{\mathbb{R}} (f(y)-f(x)-(y-x)f_x(x)) \frac{2\big((f(y)-f(x))f_x(x)+(y-x)\big)}{\big((f(y)-f(x))^2+(y-x)^2\big)^2}\,dy. \tag{3-1}$$

To simplify notation, we reparametrize (3-1) by taking $y = x+h$, and letting

$$\delta_h f(x) := f(x+h) - f(x),$$

we get

$$(f_x)_t(x) = f_{xx}(x)\int_{\mathbb{R}} \frac{-h}{(\delta_h f(x))^2+h^2}\,dh + \int_{\mathbb{R}} (\delta_h f(x) - hf_x(x)) \frac{2(\delta_h f(x) f_x(x)+h)}{(\delta_h f(x)^2+h^2)^2}\,dh. \tag{3-2}$$

Note that

$$\delta_h f(x) - hf_x(x) = \begin{cases} \int_0^h \delta_s f_x(x)\,ds & \text{for } h>0, \\ -\int_h^0 \delta_s f_x(x)\,ds & \text{for } h<0. \end{cases}$$

With that in mind, define

$$k(x,s) = \frac{2(\delta_s f(x) f_x(x)+s)}{(\delta_s f(x)^2+s^2)^2}, \tag{3-3}$$

and

$$K(x,h) = \begin{cases} \int_h^\infty k(x,s)\,ds, & h>0, \\ \int_{-\infty}^h -k(x,s)\,ds, & h<0. \end{cases} \tag{3-4}$$

Then integrating (3-2) by parts, we have that f_x solves the equation

$$(f_x)_t(x) = f_{xx}(x)\int_{\mathbb{R}} \frac{-h}{\delta_h f(x)^2+h^2}\,dh + \int_{\mathbb{R}} \delta_h f_x(x) K(x,h)\,dh. \tag{3-5}$$

As

$$\frac{-\beta(f_x)}{s} \le \frac{f_x(x)\delta_s f(x)}{s} \le \frac{\|f_x\|_{L^\infty}^2}{s}, \tag{3-6}$$

we see that

$$\frac{2(1-\beta(f_x))}{(1+\|f_x\|_{L^\infty}^2)^2}\frac{1}{|s|^3} \le \operatorname{sgn}(s)k(x,s) \le \frac{2(1+\|f_x\|_{L^\infty}^2)}{|s|^3},$$

and hence

$$\frac{1-\beta(f_x)}{(1+\|f_x\|_{L^\infty}^2)^2}\frac{1}{h^2} \le K(x,h) \le \frac{1+\|f_x\|_{L^\infty}^2}{h^2}. \tag{3-7}$$

Thus in the case that $\beta(f_x) \le 1$, we then have that the kernel K is a nonnegative, from which we get immediately:

Proposition 3.1 (maximum principle). *Let f_x be a sufficiently smooth solution to (3-5) with $\beta(f_0') \le 1$. Then for any $0 \le s \le t$, we have*

$$\inf_y f_x(s, y) \le \inf_y f_x(t, y) \le \sup_y f_x(t, y) \le \sup_y f_x(s, y). \tag{3-8}$$

In particular, since $\beta(f_0') < 1$, the maximum principle tells us that

$$\beta(f_x) \le \beta(f_0') < 1, \quad \|f_x\|_{L^\infty} \le \|f_0'\|_{L^\infty} < \infty. \tag{3-9}$$

Thus we get

$$0 < \frac{\lambda}{h^2} \le K(x, h) \le \frac{\Lambda}{h^2}, \tag{3-10}$$

where

$$\lambda = \frac{1-\beta(f_0')}{(1+\|f_0'\|_{L^\infty}^2)^2}, \qquad \Lambda = 1 + \|f_0'\|_{L^\infty}^2. \tag{3-11}$$

Thus K is comparable to the kernel for $(-\Delta)^{1/2}$, so f_x solves the uniformly elliptic equation (3-5). Note that the sole reason we require $\beta(f_0') < 1$ is to ensure this ellipticity of K.

4. Moduli estimates

Our goal is to show that if $f_x(T, \cdot)$ has modulus $\rho(\cdot / T)$ and equality is achieved at two points (2-11), then (2-12) must hold, contradicting the assumptions of the breakthrough argument (see Section 2). To that end, we first need to understand how a modulus of continuity interacts with the equation for f_x (3-5). Hence:

Lemma 4.1. *Let $f : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ be a bounded smooth solution to (1-2) with $\beta(f_0') < 1$, and $\omega : [0, \infty) \to [0, \infty)$ be some fixed modulus of continuity. Assume that at some fixed time T*

$$\begin{aligned} \delta_h f_x(T, x) &\le \omega(|h|), \\ f_x(T, \xi/2) - f_x(T, -\xi/2) &= \omega(\xi) \end{aligned} \tag{4-1}$$

for all $h \in \mathbb{R}$ and for some $\xi > 0$. Then

$$\begin{aligned} &\frac{d}{dt}(f_x(t, \xi/2) - f_x(t, -\xi/2))\Big|_{t=T} \\ &\quad \le A\omega'(\xi)\left(\int_0^\xi \frac{\omega(h)}{h}\,dh + \xi\int_\xi^\infty \frac{\omega(h)}{h^2}\,dh + \ln(M+1)\omega(\xi)\right) \\ &\qquad + A\omega(\xi)\int_{M\xi}^\infty \frac{\omega(h)}{h^2}\,dh + 2(\Lambda-\lambda)\int_\xi^{M\xi} \frac{(\omega(h-\xi)-\omega(\xi))_+}{h^2}\,dh \\ &\qquad + 2\lambda\int_0^\xi \frac{\delta_h\omega(\xi)+\delta_{-h}\omega(\xi)}{h^2}\,dh + 2\lambda\int_\xi^\infty \frac{\omega(h+\xi)-\omega(h)-\omega(\xi)}{h^2}\,dh \end{aligned} \tag{4-2}$$

for any $M \ge 1$, where A depends only on $\|f_0'\|_{L^\infty}$ and λ, Λ are as in (3-11).

This is the main technical lemma that we need. Since solutions to (1-2) are closed under translation and sign change, it suffices to consider the above situation for our proof of (2-12).

Note that (4-2) holds for any value of the parameter $M \geq 1$. Later in Lemma 6.1, we will essentially use two different values of M depending on the size of ξ. In the small ξ regime we can simply take $M = 1$, but in the large ξ regime we will need to take M to be a sufficiently large constant depending only on initial data (but not on the exact size of ξ) in order to control the size of the error term

$$\omega(\xi) \int_{M\xi}^{\infty} \frac{\omega(h)}{h^2}\, dh.$$

The proof for Lemma 4.1 is essentially a nondivergence form argument; our function f_x is touched from above at $\xi/2$ by our modulus ω, and it is touched from below at $-\xi/2$ by $-\omega$. Specifically,

$$\begin{aligned} \delta_h f_x(\xi/2) &\leq \delta_h \omega(\xi) && \text{for all } h > -\xi, \\ \delta_h f_x(-\xi/2) &\geq -\delta_{-h} \omega(\xi) && \text{for all } h < \xi. \end{aligned} \tag{4-3}$$

From (4-3), we want to derive as much information as we can and bound $\frac{d}{dt}(f_x(\xi/2) - f_x(-\xi/2))$. To that end, by dividing (4-3) through by h and taking the limit as $h \to 0$, we then get

$$f_{xx}(\xi/2) = f_{xx}(-\xi/2) = \omega'(\xi). \tag{4-4}$$

Hence by our equation for f_x (3-5), we have

$$\begin{aligned} &\frac{d}{dt}(f_x(\xi/2) - f_x(-\xi/2)) \\ &\quad = \omega'(\xi) \int_{\mathbb{R}} \left(\frac{-h}{\delta_h f(\xi/2)^2 + h^2} - \frac{-h}{\delta_h f(-\xi/2)^2 + h^2} \right) dh \\ &\qquad + \int_{\mathbb{R}} \delta_h f_x(\xi/2) K(\xi/2, h) - \delta_h f_x(-\xi/2) K(-\xi/2, h)\, dh \\ &\quad = \omega'(\xi) \int_{\mathbb{R}} \left(\frac{-h}{\delta_h f(\xi/2)^2 + h^2} - \frac{-h}{\delta_h f(-\xi/2)^2 + h^2} \right) dh + \omega'(\xi) \int_{-M\xi}^{M\xi} (hK(\xi/2, h) - hK(-\xi/2, h))\, dh \\ &\qquad + \int_{-M\xi}^{M\xi} (\delta_h f_x(\xi/2) - h\omega'(\xi)) K(\xi/2, h) - (\delta_h f_x(-\xi/2) - h\omega'(\xi)) K(-\xi/2, h)\, dh \\ &\qquad + \int_{|h| > M\xi} \delta_h f_x(\xi/2) K(\xi/2, h) - \delta_h f_x(-\xi/2) K(-\xi/2, h)\, dh \end{aligned} \tag{4-5}$$

for any $M \geq 1$. The first two terms of the far right-hand side of (4-5) act as a drift, giving rise to the first two error terms of (4-2). The latter two terms of (4-5) act as a diffusion, giving rise to both the helpful (negative) terms in (4-2), as well as additional error terms (the middle terms of (4-2)) arising from the difference in the kernels, $|K(\xi/2, h) - K(-\xi/2, h)|$.

5. Bounds on drift terms

We begin proving Lemma 4.1 by bounding the drift terms of (4-5), starting with:

Lemma 5.1. *Under the assumptions of Lemma 4.1*,

$$\omega'(\xi)\left|\int_{\mathbb{R}} \frac{-h}{\delta_h f(\xi/2)^2+h^2}-\frac{-h}{\delta_h f(-\xi/2)^2+h^2}\,dh\right| \lesssim \omega'(\xi)\left(\int_0^{\xi}\frac{\omega(h)}{h}\,dh+\xi\int_{\xi}^{\infty}\frac{\omega(h)}{h^2}\,dh\right). \tag{5-1}$$

Proof. We want to bound (5-1) by symmetrizing the kernels for $|h|<\xi$ and then using the continuity in the first variable for $|h|>\xi$. To that end,

$$\begin{aligned}
&\omega'(\xi)\int_{\mathbb{R}}\left(\frac{-h}{\delta_h f(\xi/2)^2+h^2}-\frac{-h}{\delta_h f(-\xi/2)^2+h^2}\right)dh\\
&\leq \omega'(\xi)\int_0^{\xi} h\left|\frac{\delta_h f(\xi/2)^2-\delta_{-h} f(\xi/2)^2}{(\delta_h f(\xi/2)^2+h^2)(\delta_{-h} f(\xi/2)^2+h^2)}+\frac{\delta_h f(-\xi/2)^2-\delta_{-h} f(-\xi/2)^2}{(\delta_h f(-\xi/2)^2+h^2)(\delta_{-h} f(-\xi/2)^2+h^2)}\right|dh\\
&\qquad+\omega'(\xi)\int_{|h|>\xi}|h|\left|\frac{\delta_h f(\xi/2)^2-\delta_h f(-\xi/2)^2}{(\delta_h f(\xi/2)^2+h^2)(\delta_h f(-\xi/2)^2+h^2)}\right|dh.
\end{aligned} \tag{5-2}$$

We bound the first integral using

$$\begin{gathered}
|\delta_h f(x)|\lesssim |h|,\\
|\delta_h f(x)+\delta_{-h} f(x)| = \left|\int_0^h f_x(x+s)-f_x(x+s-h)\,ds\right| \leq \omega(h)h.
\end{gathered} \tag{5-3}$$

Thus we get, for $0\leq h<\xi$,

$$\left|\frac{\delta_h f(x)^2-\delta_{-h} f(x)^2}{(\delta_h f(x)^2+h^2)(\delta_{-h} f(x)^2+h^2)}\right| \lesssim \frac{\omega(h)}{h^2}, \tag{5-4}$$

and hence

$$\int_0^{\xi} h\left|\frac{\delta_h f(\xi/2)^2-\delta_{-h} f(\xi/2)^2}{(\delta_h f(\xi/2)^2+h^2)(\delta_{-h} f(\xi/2)^2+h^2)}\,dh\right| \lesssim \int_0^{\xi}\frac{\omega(h)}{h}\,dh. \tag{5-5}$$

For $|h|\geq\xi$, we bound $|\delta_h f(\xi/2)+\delta_h f(-\xi/2)|\lesssim|h|$ and

$$\begin{aligned}
\left|\delta_h f(\xi/2)-\delta_h f(-\xi/2)\right| &= \left|\int_0^h f_x(\xi/2+s)-f_x(-\xi/2+s)\,ds\right|\\
&= \left|\int_0^{\xi} f_x(h-\xi/2+s)-f_x(-\xi/2+s)\,ds\right| \leq \xi\omega(|h|),
\end{aligned} \tag{5-6}$$

in order to get

$$\int\limits_{|h|>\xi} |h| \left| \frac{\delta_h f(\xi/2)^2 - \delta_h f(-\xi/2)^2}{(\delta_h f(\xi/2)^2 + h^2)(\delta_h f(-\xi/2)^2 + h^2)} \right| dh \lesssim \xi \int\limits_{\xi}^{\infty} \frac{\omega(h)}{h^2}\, dh. \tag{5-7}$$

Putting (5-5) and (5-7) together, we thus have

$$\omega'(\xi) \int\limits_{\mathbb{R}} \left(\frac{-h}{\delta_h f(\xi/2)^2 + h^2} - \frac{-h}{\delta_h f(-\xi/2)^2 + h^2} \right) dh \lesssim \omega'(\xi) \left(\int\limits_0^{\xi} \frac{\omega(h)}{h}\, dh + \xi \int\limits_{\xi}^{\infty} \frac{\omega(h)}{h^2}\, dh \right), \tag{5-8}$$

completing the proof. $\square$

That leaves us with the second drift term of (4-5):

Lemma 5.2. *Under the assumptions of Lemma 4.1, for any* $M \ge 1$

$$\omega'(\xi) \left| \int\limits_{-M\xi}^{M\xi} hK(\xi/2, h) - hK(-\xi/2, h)\, dh \right| \lesssim \omega'(\xi) \left(\int\limits_0^{\xi} \frac{\omega(h)}{h}\, dh + \xi \int\limits_{\xi}^{\infty} \frac{\omega(h)}{h^2}\, dh + \ln(M+1)\omega(\xi) \right). \tag{5-9}$$

Proof. To begin, we note

$$\begin{aligned} &\omega'(\xi) \left| \int\limits_{-M\xi}^{M\xi} hK(\xi/2, h) - hK(-\xi/2, h)\, dh \right| \\ &\qquad \le \omega'(\xi) \int\limits_0^{M\xi} h \big| K(\xi/2, h) - K(\xi/2, -h) - K(-\xi/2, h) + K(-\xi/2, -h) \big|\, dh. \end{aligned} \tag{5-10}$$

Recall the definition of K, (3-4),

$$K(x, h) = \begin{cases} \int_h^{\infty} k(x, s)\, ds, & h > 0, \\ \int_{-\infty}^{h} -k(x, s)\, ds, & h < 0, \end{cases} \qquad k(x, s) = \frac{2(\delta_s f(x) f_x(x) + s)}{(\delta_s f(x)^2 + s^2)^2}. \tag{5-11}$$

So, to control (5-10) we first need to bound $|k(x, s) + k(x, -s)|$ for $0 \le s < \xi$, and $|k(\xi/2, s) - k(-\xi/2, s)|$ for $|s| > \xi$. For the first, using the bounds (5-3) we see that

$$\begin{aligned} &|k(x, s) + k(x, -s)| \\ &\quad = \left| \frac{2(\delta_s f(x) f_x(x) + s)}{(\delta_s f(x)^2 + s^2)^2} + \frac{2(\delta_{-s} f(x) f_x(x) - s)}{(\delta_{-s} f(x)^2 + s^2)^2} \right| \\ &\quad \le \frac{2|\delta_s f(x) + \delta_{-s} f(x)|\,|f_x(x)|}{(\delta_{-s} f(x)^2 + s^2)^2} + 2|\delta_s f(x) f_x(x) + s| \left| \frac{(\delta_s f(x)^2 + s^2)^2 - (\delta_{-s} f(x)^2 + s^2)^2}{(\delta_s f(x)^2 + s^2)^2 (\delta_{-s} f(x)^2 + s^2)^2} \right| \\ &\quad \lesssim \frac{\omega(s)}{s^3} + s \left| \frac{\delta_s f(x)^4 - \delta_{-s} f(x)^4 + 2s^2(\delta_s f(x)^2 - \delta_{-s} f(x)^2)}{s^8} \right| \\ &\quad \lesssim \frac{\omega(s)}{s^3}. \end{aligned} \tag{5-12}$$

For the second, using (5-3), (5-6), and (4-1) we get

$$\begin{aligned}
|k(\xi/2,s)-k(-\xi/2,s)| &= \left|\frac{2(\delta_s f(\xi/2) f_x(\xi/2)+s)}{(\delta_s f(\xi/2)^2+s^2)^2} - \frac{2(\delta_s f(-\xi/2) f_x(-\xi/2)+s)}{(\delta_s f(-\xi/2)^2+s^2)^2}\right| \\
&\le 2\frac{|\delta_s f(\xi/2) f_x(\xi/2) - \delta_s f(-\xi/2) f_x(-\xi/2)|}{(\delta_s f(-\xi/2)^2+s^2)^2} \\
&\qquad + 2|\delta_s f(\xi/2) f_x(\xi/2)+s| \left|\frac{(\delta_s f(\xi/2)^2+s^2)^2-(\delta_s f(-\xi/2)^2+s^2)^2}{(\delta_s f(\xi/2)^2+s^2)^2(\delta_s f(-\xi/2)^2+s^2)^2}\right| \\
&\lesssim \frac{|\delta_s f(\xi/2)-\delta_s f(-\xi/2)|\,|f_x(\xi/2)|}{s^4} + \frac{|\delta_s f(-\xi/2)|\,|f_x(\xi/2)-f_x(-\xi/2)|}{s^4} \\
&\qquad + |s| \left|\frac{\delta_s f(\xi/2)^4 - \delta_s f(-\xi/2)^4 + s^2(\delta_s f(\xi/2)^2 - \delta_s f(-\xi/2)^2)}{s^8}\right| \\
&\lesssim \frac{\xi\omega(s)}{s^4} + \frac{\omega(\xi)}{s^3}.
\end{aligned} \tag{5-13}$$

So using (5-12) and (5-13), we can first bound

$$\begin{aligned}
\int_0^\xi h\big|K(\xi/2,h)-K(\xi/2,-h)-K(-\xi/2,h)+K(-\xi/2,-h)\big|\,dh \\
\lesssim \int_0^\xi h \int_h^\xi \frac{\omega(s)}{s^3}\,ds\,dh + \int_0^\xi h \int_\xi^\infty \frac{\xi\omega(s)}{s^4} + \frac{\omega(\xi)}{s^3}\,ds\,dh \\
\lesssim \int_0^\xi \frac{\omega(s)}{s^3} \int_0^s h\,dh\,ds + \int_\xi^\infty \frac{\xi^3\omega(s)}{s^4} + \frac{\xi^2\omega(\xi)}{s^3}\,ds \\
\lesssim \int_0^\xi \frac{\omega(s)}{s}\,ds + \xi \int_\xi^\infty \frac{\omega(s)}{s^2}\,ds + \omega(\xi).
\end{aligned} \tag{5-14}$$

For the rest of (5-10), we use (5-13) again to also bound

$$\begin{aligned}
\int_{M\xi>|h|>\xi} |h|\big|K(\xi/2,h)-K(-\xi/2,h)\big|\,dh &\lesssim \int_\xi^{M\xi} h \int_h^\infty \frac{\omega(\xi)}{s^3} + \frac{\xi\omega(s)}{s^4}\,ds\,dh \\
&\lesssim \omega(\xi)\int_\xi^{M\xi} \frac{1}{h}\,dh + \xi \int_\xi^{M\xi} \frac{\omega(h)}{h^2}\,dh \\
&\lesssim \ln(M)\omega(\xi) + \xi \int_\xi^\infty \frac{\omega(h)}{h^2}\,dh,
\end{aligned} \tag{5-15}$$

completing the proof. □

6. Bounds on diffusive terms

Now we move on to proving an upper bound for the diffusive terms of (4-5). We can rewrite them as

$$
\begin{aligned}
&\int_{-M\xi}^{M\xi} (\delta_h f_x(\xi/2) - h\omega'(\xi))K(\xi/2,h) - (\delta_h f_x(-\xi/2) - h\omega'(\xi))K(-\xi/2,h)\,dh \\
&\qquad\qquad\qquad + \int_{|h|>M\xi} \delta_h f_x(\xi/2)K(\xi/2,h) - \delta_h f_x(-\xi/2)K(-\xi/2,h)\,dh \\
&\quad = \int_{-M\xi}^{M\xi} (\delta_h f_x(\xi/2) - h\omega'(\xi))K(\xi/2,h) - (\delta_h f_x(-\xi/2) - h\omega'(\xi))K(-\xi/2,h)\,dh \\
&\qquad + \int_{|h|>M\xi} [\delta_h f_x(\xi/2) - \delta_h f_x(-\xi/2)]K(\xi/2,h)\,dh \\
&\qquad\quad + \int_{|h|>M\xi} \delta_h f_x(-\xi/2)[K(\xi/2,h) - K(-\xi/2,h)]\,dh. \qquad (6\text{-}1)
\end{aligned}
$$

We begin by bounding the last term, which is an error term.

Lemma 6.1. *Under the assumptions of Lemma 4.1*,

$$
\left| \int_{|h|>M\xi} \delta_h f_x(-\xi/2)[K(\xi/2,h) - K(-\xi/2,h)] \right| dh \lesssim \omega(\xi)\int_{M\xi}^{\infty} \frac{\omega(h)}{h^2}\,dh + \omega'(\xi)\xi\int_{\xi}^{\infty} \frac{\omega(h)}{h^2}\,dh. \qquad (6\text{-}2)
$$

Proof. Using the fact that f_x has modulus ω and the bounds (5-13), it follows that

$$
\begin{aligned}
&\int_{|h|>M\xi} \delta_h f_x(-\xi/2)[K(\xi/2,h) - K(-\xi/2,h)]\,dh \\
&\qquad \lesssim \int_{M\xi}^{\infty} \omega(h) \int_{h}^{\infty} \frac{\omega(\xi)}{s^3} + \frac{\xi\omega(s)}{s^4}\,ds\,dh \\
&\qquad \lesssim \omega(\xi)\int_{M\xi}^{\infty} \frac{\omega(h)}{h^2}\,dh + \int_{M\xi}^{\infty} \omega(h) \int_{h}^{\infty} \frac{\xi\omega(\xi) + \xi\omega'(\xi)(s-\xi)}{s^4}\,ds\,dh \\
&\qquad \lesssim \omega(\xi)\int_{M\xi}^{\infty} \frac{\omega(h)}{h^2}\,dh + \omega(\xi)\int_{M\xi}^{\infty} \frac{\xi\omega(h)}{h^3}\,dh + \omega'(\xi)\xi\int_{M\xi}^{\infty} \frac{\omega(h)}{h^2}\,dh \\
&\qquad \lesssim \omega(\xi)\int_{M\xi}^{\infty} \frac{\omega(h)}{h^2}\,dh + \omega'(\xi)\xi\int_{\xi}^{\infty} \frac{\omega(h)}{h^2}\,dh, \qquad (6\text{-}3)
\end{aligned}
$$

completing the proof. □

For the other two terms in (6-1), we bound them in two stages.

Lemma 6.2. *Under the assumptions of Lemma 4.1*,

$$\int_{-M\xi}^{M\xi} (\delta_h f_x(\xi/2) - h\omega'(\xi))K(\xi/2,h) - (\delta_h f_x(-\xi/2) - h\omega'(\xi))K(-\xi/2,h)\,dh \\ + \int_{|h|>M\xi} [\delta_h f_x(\xi/2) - \delta_h f_x(-\xi/2)]K(\xi/2,h)\,dh$$

$$\le \lambda \int_{\mathbb{R}} \frac{\delta_h f_x(\xi/2) - \delta_h f_x(-\xi/2)}{h^2}\,dh + 2(\Lambda-\lambda)\int_{\xi}^{M\xi} \frac{(\omega(h-\xi)-\omega(\xi))_+}{h^2}\,dh \\ + \omega'(\xi)\int_{\xi<|h|<M\xi} \big|h[K(\xi/2,h) - K(-\xi/2,h)]\big|\,dh. \tag{6-4}$$

Proof. We can bound the second term of (6-4) rather easily. Since

$$\delta_h f_x(\xi/2) - \delta_h f_x(-\xi/2) = (f_x(h+\xi/2) - f_x(h-\xi/2)) - \omega(\xi) \le 0, \tag{6-5}$$

by the uniform ellipticity of K,

$$\int_{|h|>M\xi} [\delta_h f_x(\xi/2) - \delta_h f_x(-\xi/2)]K(\xi/2,h)\,dh \le \lambda \int_{|h|>M\xi} \frac{\delta_h f_x(\xi/2) - \delta_h f_x(-\xi/2)}{h^2}\,dh. \tag{6-6}$$

To bound the first term, we first define

$$G(\xi,h) = (\delta_h f_x(\xi/2) - h\omega'(\xi))K(\xi/2,h) - (\delta_h f_x(-\xi/2) - h\omega'(\xi))K(-\xi/2,h). \tag{6-7}$$

Note that since ω is concave and touches f_x from above, see (4-3), it follows that

$$\begin{aligned} \delta_h f_x(\xi/2) - \omega'(\xi)h &\le \delta_h\omega(\xi) - \omega'(\xi)h \le 0, && h \ge -\xi, \\ \delta_h f_x(-\xi/2) - \omega'(\xi)h &\ge -\delta_{-h}\omega(\xi) - h\omega'(\xi) \ge 0, && h \le \xi. \end{aligned} \tag{6-8}$$

Thus for $|h| \le \xi$, by the uniform ellipticity of K we have the bound

$$G(\xi,h) \le \lambda \frac{\delta_h f_x(\xi/2) - \delta_h f_x(-\xi/2)}{h^2}. \tag{6-9}$$

That just leaves us with the case $\xi \le |h| \le M\xi$ to analyze. Note that we can write G in two distinct ways:

$$\begin{aligned} G(\xi,h) &= (\delta_h f_x(\xi/2) - \delta_h f_x(-\xi/2))K(\xi/2,h) + (\delta_h f_x(-\xi/2) - h\omega'(\xi))(K(\xi/2,h) - K(-\xi/2,h)) \\ &= (\delta_h f_x(\xi/2) - \delta_h f_x(-\xi/2))K(-\xi/2,h) + (\delta_h f_x(\xi/2) - h\omega'(\xi))(K(\xi/2,h) - K(-\xi/2,h)). \end{aligned} \tag{6-10}$$

By (6-8), $\delta_h f_x(\xi/2) - h\omega'(\xi) \le 0$ for all $h > \xi$. Thus if $K(\xi/2,h) - K(-\xi/2,h) \ge 0$, then

$$G(\xi,h) \le \lambda \frac{\delta_h f_x(\xi/2) - \delta_h f_x(-\xi/2)}{h^2} \quad \text{if } K(\xi/2,h) - K(-\xi/2,h) \ge 0. \tag{6-11}$$

On the other hand, since

$$\delta_h f_x(-\xi/2) = \delta_{h-\xi} f(\xi/2) + \omega(\xi) \geq -\omega(h-\xi) + \omega(\xi) \tag{6-12}$$

for $h \geq \xi$, we see that

$$\begin{aligned} G(\xi, h) &\leq \lambda \frac{\delta_h f_x(\xi/2) - \delta_h f_x(-\xi/2)}{h^2} + (\delta_h f_x(-\xi/2) - h\omega'(\xi))(K(\xi/2, h) - K(-\xi/2, h)) \\ &\leq \lambda \frac{\delta_h f_x(\xi/2) - \delta_h f_x(-\xi/2)}{h^2} + (\Lambda - \lambda) \frac{(\omega(h-\xi) - \omega(\xi))_+}{h^2} \\ &\qquad + h\omega'(\xi)|K(\xi/2, h) - K(-\xi/2, h)| \quad \text{if } K(\xi/2, h) - K(-\xi/2, h) \leq 0. \end{aligned} \tag{6-13}$$

Putting these two together, we get

$$\begin{aligned} G(\xi, h) \leq \lambda \frac{\delta_h f_x(\xi/2) - \delta_h f_x(-\xi/2)}{h^2} + (\Lambda - \lambda) \frac{(\omega(h-\xi) - \omega(\xi))_+}{h^2} \\ + h\omega'(\xi)|K(\xi/2, h) - K(-\xi/2, h)| \end{aligned} \tag{6-14}$$

for $h \geq \xi$. A similar argument can be made in the case that $h \leq -\xi$.

Putting this all together,

$$\begin{aligned} &\int_{-M\xi}^{M\xi} G(\xi, h)\, dh + \int_{|h|>M\xi} [\delta_h f_x(\xi/2) - \delta_h f_x(-\xi/2)] K(\xi/2, h)\, dh \\ &\quad \leq \lambda \int_{\mathbb{R}} \frac{\delta_h f_x(\xi/2) - \delta_h f_x(-\xi/2)}{h^2}\, dh + 2(\Lambda - \lambda) \int_{\xi}^{M\xi} \frac{(\omega(h-\xi) - \omega(\xi))_+}{h^2}\, dh \\ &\qquad + \omega'(\xi) \int_{\xi<|h|<M\xi} \big| h[K(\xi/2, h) - K(-\xi/2, h)] \big|\, dh, \end{aligned} \tag{6-15}$$

completing the proof. □

It's clear that we can bound $\int_{\xi<|h|<M\xi} \big|h[K(\xi/2, h) - K(-\xi/2, h)]\big|\, dh$ as in (5-15). Thus the only thing remaining to prove (4-2) is:

Lemma 6.3. *Under the assumptions of Lemma 4.1*,

$$\begin{aligned} &\lambda \int_{\mathbb{R}} \frac{\delta_h f_x(\xi/2) - \delta_h f_x(-\xi/2)}{h^2}\, dh \\ &\qquad \leq 2\lambda \int_0^{\xi} \frac{\delta_h \omega(\xi) + \delta_{-h}\omega(\xi)}{h^2}\, dh + 2\lambda \int_{\xi}^{\infty} \frac{\omega(\xi+h) - \omega(h) - \omega(\xi)}{h^2}\, dh. \end{aligned} \tag{6-16}$$

Proof. To see this, note that formally we should have

$$\int_{\mathbb{R}} \frac{\delta_h f_x(\xi/2) - \delta_h f_x(-\xi/2)}{h^2}\, dh = \int_{\mathbb{R}} f_x(y) \left(\frac{1}{(y-\xi/2)^2} - \frac{1}{(y+\xi/2)^2} \right) - \frac{\omega(\xi)}{y^2}\, dy. \tag{6-17}$$

Thus in order to get an upper bound on (6-17), we should be taking an upper bound on $f_x(y)$ when $y > 0$ and a lower bound when $y < 0$. Note by (4-3) that

$$\begin{aligned} f_x(y) &\le f_x(\xi/2) + \omega(y+\xi/2) - \omega(\xi) = f_x(-\xi/2) + \omega(y+\xi/2), \qquad && y > -\xi/2, \\ f_x(y) &\ge f_x(-\xi/2) - \omega(-y+\xi/2) + \omega(\xi) = f_x(\xi/2) - \omega(-y+\xi/2), && y < \xi/2. \end{aligned} \tag{6-18}$$

In particular, using the upper bounds on $\delta_h f_x(\pm\xi/2)$ for $h > 0$ and the lower bounds for $\delta_h f_x(\pm\xi/2)$ for $h < 0$ gives the result. To rigorously justify this though, we will bound

$$\int_\epsilon^\infty \frac{\delta_h f_x(\xi/2) - \delta_h f_x(-\xi/2)}{h^2}\,dh$$

from above. Taking $\epsilon \to 0$, we'll get

$$\int_0^\infty \frac{\delta_h f_x(\xi/2) - \delta_h f_x(-\xi/2)}{h^2}\,dh \le \int_0^\xi \frac{\delta_h \omega(\xi) + \delta_{-h}\omega(\xi)}{h^2}\,dh + \int_\xi^\infty \frac{\omega(\xi+h) - \omega(h) - \omega(\xi)}{h^2}\,dh. \tag{6-19}$$

The bound for $\int_{-\infty}^0$ follows from identical arguments.

So, fix some $\epsilon \ll \xi$. By splitting the integral into a several pieces and reparametrizing, we get

$$\begin{aligned} &\int_\epsilon^\infty \frac{\delta_h f_x(\xi/2) - \delta_h f_x(-\xi/2)}{h^2}\,dh \\ &\qquad = \int_{\epsilon+\xi/2}^\infty \frac{f_x(y)}{(y-\xi/2)^2}\,dy - \int_{\epsilon-\xi/2}^\infty \frac{f_x(y)}{(y+\xi/2)^2}\,dy - \int_\epsilon^\infty \frac{\omega(\xi)}{y^2}\,dy \\ &\qquad = \int_{\epsilon+\xi/2}^\infty f_x(y)\left(\frac{1}{(y-\xi/2)^2} - \frac{1}{(y+\xi/2)^2}\right)dy - \int_\epsilon^\infty \frac{\omega(\xi)}{y^2}\,dy - \int_{\epsilon-\xi/2}^{\epsilon+\xi/2} \frac{f_x(y)}{(y+\xi/2)^2}\,dy. \end{aligned} \tag{6-20}$$

In the first integral of the third line, since $y > \xi/2$ we have $(y-\xi/2)^{-2} > (y+\xi/2)^{-2}$. So applying the upper bound in (6-18) gives an upper bound on the integral,

$$\begin{aligned} &\int_{\epsilon+\xi/2}^\infty f_x(y)\left(\frac{1}{(y-\xi/2)^2} - \frac{1}{(y+\xi/2)^2}\right)dy \\ &\qquad \le \int_{\epsilon+\xi/2}^\infty (f_x(\xi/2) + \omega(y+\xi/2) - \omega(\xi))\left(\frac{1}{(y-\xi/2)^2} - \frac{1}{(y+\xi/2)^2}\right)dy \\ &\qquad = \int_{\epsilon+\xi/2}^\infty \frac{f_x(\xi/2) + \omega(y+\xi/2) - \omega(\xi)}{(y-\xi/2)^2}\,dy - \int_{\epsilon+\xi/2}^\infty \frac{f_x(\xi/2) + \omega(y+\xi/2) - \omega(\xi)}{(y+\xi/2)^2}\,dy. \end{aligned} \tag{6-21}$$

By reparametrizing back, we get

$$\int_{\epsilon+3\xi/2}^{\infty} \frac{f_x(\xi/2)+\omega(y+\xi/2)-\omega(\xi)}{(y-\xi/2)^2}\,dy - \int_{\epsilon+\xi/2}^{\infty} \frac{f_x(\xi/2)+\omega(y+\xi/2)-\omega(\xi)}{(y+\xi/2)^2}\,dy - \int_{\epsilon+\xi}^{\infty} \frac{\omega(\xi)}{y^2}\,dy$$
$$= \int_{\epsilon+\xi}^{\infty} \frac{\omega(\xi+h)-\omega(h)-\omega(\xi)}{h^2}\,dh. \tag{6-22}$$

Hence combining (6-20), (6-21), and (6-22) gives us

$$\begin{aligned}
&\int_{\epsilon}^{\infty} \frac{\delta_h f_x(\xi/2)-\delta_h f_x(-\xi/2)}{h^2}\,dh \\
&\le \int_{\epsilon+\xi}^{\infty} \frac{\omega(\xi+h)-\omega(h)-\omega(\xi)}{h^2}\,dh + \int_{\epsilon}^{\epsilon+\xi} \frac{f_x(\xi/2)+\omega(\xi+h)-\omega(\xi)}{h^2}\,dh - \int_{\epsilon}^{\epsilon+\xi} \frac{\omega(\xi)}{h^2}\,dh - \int_{\epsilon}^{\epsilon+\xi} \frac{f_x(h-\xi/2)}{h^2}\,dh \\
&= \int_{\epsilon+\xi}^{\infty} \frac{\omega(\xi+h)-\omega(h)-\omega(\xi)}{h^2}\,dh + \int_{\epsilon}^{\epsilon+\xi} \frac{\delta_h\omega(\xi)+f_x(\xi/2)-f_x(h-\xi/2)-\omega(\xi)}{h^2}\,dh.
\end{aligned} \tag{6-23}$$

Now for $h<\xi$, we have $f_x(\xi/2)-f_x(h-\xi/2)\le\omega(\xi-h)$, and thus

$$\int_{\epsilon}^{\xi} \frac{\delta_h\omega(\xi)+f_x(\xi/2)-f_x(h-\xi/2)-\omega(\xi)}{h^2}\,dh \le \int_{\epsilon}^{\xi} \frac{\delta_h\omega(\xi)+\delta_{-h}\omega(\xi)}{h^2}\,dh. \tag{6-24}$$

Taking the limit as $\epsilon\to 0$, we then get

$$\int_{0}^{\infty} \frac{\delta_h f_x(\xi/2)-\delta_h f_x(-\xi/2)}{h^2}\,dh \le \int_{0}^{\xi} \frac{\delta_h\omega(\xi)+\delta_{-h}\omega(\xi)}{h^2}\,dh + \int_{\xi}^{\infty} \frac{\omega(\xi+h)-\omega(h)-\omega(\xi)}{h^2}\,dh. \qquad \square$$

7. Modulus inequality

Combining all the estimates from the previous two sections, we get a proof of Lemma 4.1. Thus under the assumptions (4-1), we have

$$\begin{aligned}
\frac{d}{dt}\big(f_x(\xi/2)-f_x(-\xi/2)\big) \le A\omega'(\xi)\bigg(&\int_0^{\xi} \frac{\omega(h)}{h}\,dh+\xi\int_{\xi}^{\infty}\frac{\omega(h)}{h^2}\,dh+\ln(M+1)\omega(\xi)\bigg) \\
&+A\omega(\xi)\int_{M\xi}^{\infty}\frac{\omega(h)}{h^2}\,dh+2(\Lambda-\lambda)\int_{\xi}^{M\xi}\frac{(\omega(h-\xi)-\omega(\xi))_+}{h^2}\,dh \\
&+2\lambda\int_0^{\xi}\frac{\delta_h\omega(\xi)+\delta_{-h}\omega(\xi)}{h^2}\,dh+2\lambda\int_{\xi}^{\infty}\frac{\omega(h+\xi)-\omega(h)-\omega(\xi)}{h^2}\,dh
\end{aligned} \tag{7-1}$$

for any $M\ge 1$, where A is a constant depending only on $\|f_0'\|_{L^\infty}$.

In [Kiselev et al. 2007], the authors showed that the modulus

$$\begin{aligned} \omega(\xi) &= \xi - \xi^{3/2}, && 0 \le \xi \le \delta, \\ \omega'(\xi) &= \frac{\gamma}{\xi(4+\log(\xi/\delta))}, && \xi \ge \delta \end{aligned} \tag{7-2}$$

satisfies

$$A\omega'(\xi)\left(\int_0^\xi \frac{\omega(h)}{h}\,dh + \xi\int_\xi^\infty \frac{\omega(h)}{h^2}\,dh\right) + \lambda\int_0^\xi \frac{\delta_h\omega(\xi)+\delta_{-h}\omega(\xi)}{h^2}\,dh + \lambda\int_\xi^\infty \frac{\omega(h+\xi)-\omega(h)-\omega(\xi)}{h^2}\,dh < 0 \tag{7-3}$$

for all $\xi \in \mathbb{R}$ so long as δ, γ are sufficiently small.

With that in mind, we will show:

Lemma 7.1. *Under the assumptions of Lemma 4.1 for the modulus ω defined in (7-2),*

$$\frac{d}{dt}(f_x(\xi/2) - f_x(-\xi/2)) < -\omega'(\xi)\omega(\xi), \tag{7-4}$$

as long as δ, γ are taken sufficiently small depending on $\beta(f_0')$, $\|f_0'\|_{L^\infty}$.

Proof. By Lemma 4.1 and (7-3) which was proven in [Kiselev et al. 2007], it suffices to show

$$A\omega'(\xi)\ln(M+1)\omega(\xi) + A\omega(\xi)\int_{M\xi}^\infty \frac{\omega(h)}{h^2}\,dh + 2(\Lambda-\lambda)\int_\xi^{M\xi} \frac{(\omega(h-\xi)-\omega(\xi))_+}{h^2}\,dh + \lambda\int_0^\xi \frac{\delta_h\omega(\xi)+\delta_{-h}\omega(\xi)}{h^2}\,dh + \lambda\int_\xi^\infty \frac{\omega(h+\xi)-\omega(h)-\omega(\xi)}{h^2}\,dh \le -\omega'(\xi)\omega(\xi) \tag{7-5}$$

for the correct choices of M, and δ, γ sufficiently small.

We proceed very similarly to [Kiselev et al. 2007]. To begin, for $\xi \le \delta$ we take $M = 1$. Then we just need to show that

$$A\omega'(\xi)\omega(\xi) + A\omega(\xi)\int_\xi^\infty \frac{\omega(h)}{h^2}\,dh + \lambda\int_0^\xi \frac{\delta_h\omega(\xi)+\delta_{-h}\omega(\xi)}{h^2}\,dh + \lambda\int_\xi^\infty \frac{\omega(h+\xi)-\omega(h)-\omega(\xi)}{h^2}\,dh \le -\omega'(\xi)\omega(\xi). \tag{7-6}$$

In this regime, note that we have the bounds

$$\int_{\xi}^{\delta} \frac{\omega(h)}{h^2}\,dh \le \log(\delta/\xi),$$

$$\int_{\delta}^{\infty} \frac{\omega(h)}{h^2}\,dh = \frac{\omega(\delta)}{\delta} + \gamma \int_{\delta}^{\infty} \frac{1}{h^2(4+\log(h/\delta))}\,dh \le 1 + \frac{\gamma}{4\delta} \le 2 \quad \text{if we take } \gamma < 4\delta, \tag{7-7}$$

$$\omega'(\xi) \le 1, \quad \omega(\xi) \le \xi,$$

$$\int_{0}^{\xi} \frac{\omega(\xi+h)+\omega(\xi-h)-2\omega(\xi)}{h^2}\,dh \le \xi\omega''(\xi) = -\tfrac{3}{2}\xi\xi^{-1/2}.$$

Putting this all together, we get

$$\begin{aligned}(A+1)\omega'(\xi)\omega(\xi) + A\omega(\xi)\int_{\xi}^{\infty} \frac{\omega(h)}{h^2}\,dh \\ + \lambda \int_{0}^{\xi} \frac{\omega(\xi+h)+\omega(\xi-h)-2\omega(\xi)}{h^2}\,dh + \lambda \int_{\xi}^{\infty} \frac{\omega(\xi+h)-\omega(h)-\omega(\xi)}{h^2}\,dh \\ \le \xi\big((A+1)(3+\log(\delta/\xi)) - \tfrac{3}{2}\lambda\xi^{-1/2}\big) < 0,\end{aligned} \tag{7-8}$$

assuming that δ is sufficiently small.

Now assume that $\xi \ge \delta$. Then what we need to show is

$$\begin{aligned}A\omega'(\xi)\ln(M+1)\omega(\xi) + A\omega(\xi)\int_{M\xi}^{\infty} \frac{\omega(h)}{h^2}\,dh + 2(\Lambda-\lambda)\int_{\xi}^{M\xi} \frac{(\omega(h-\xi)-\omega(\xi))_+}{h^2}\,dh \\ + \lambda\int_{0}^{\xi} \frac{\delta_h\omega(\xi)+\delta_{-h}\omega(\xi)}{h^2}\,dh + \lambda\int_{\xi}^{\infty} \frac{\omega(h+\xi)-\omega(h)-\omega(\xi)}{h^2}\,dh \le -\omega'(\xi)\omega(\xi).\end{aligned} \tag{7-9}$$

We first bound our new error terms. Using the definition of ω and integrating by parts, we see that

$$\begin{aligned}2(\Lambda-\lambda)\int_{2\xi}^{M\xi} \frac{\omega(h-\xi)-\omega(\xi)}{h^2}\,dh \le 2(\Lambda-\lambda)\int_{\xi}^{\infty} \frac{\omega(h)-\omega(\xi)}{h^2}\,dh \le 2(\Lambda-\lambda)\int_{\xi}^{\infty} \frac{\gamma}{h^2(4+\log(h/\delta))}\,dh \\ \le \frac{2(\Lambda-\lambda)\gamma}{\xi} \le \frac{\lambda}{4}\frac{\omega(\delta)}{\xi} \le \frac{\lambda}{4}\frac{\omega(\xi)}{\xi},\end{aligned} \tag{7-10}$$

assuming $\gamma \le (\lambda/(8(\Lambda-\lambda)))\omega(\delta)$.

In order to bound our other new error term, we will be taking M sufficiently large and then γ sufficiently small depending on M, δ. Noting that $\omega(\xi) \le 2\|f_0'\|_{L^\infty}$, we can bound our other new error term by integrating by parts,

$$\begin{aligned}A\omega(\xi)\int_{M\xi}^{\infty}\frac{\omega(h)}{h^2}\,dh &\leq \frac{2A\|f_0'\|_{L^\infty}}{M}\frac{\omega(M\xi)}{\xi}+2A\|f_0'\|_{L^\infty}\int_{M\xi}^{\infty}\frac{\gamma}{h^2(4+\log(h/\delta))}\,dh\\ &\leq \frac{2A\|f_0'\|_{L^\infty}}{M}\frac{\omega(M\xi)}{\xi}+\frac{2A\|f_0'\|_{L^\infty}}{M}\frac{\gamma}{\xi}\\ &\leq \frac{\lambda}{16}\frac{\omega(M\xi)}{\xi}+\frac{\lambda}{8}\frac{\omega(\xi)}{\xi},\end{aligned}\tag{7-11}$$

assuming that

$$M\geq\frac{32A\|f_0'\|_{L^\infty}}{\lambda},$$

and then γ is sufficiently small so that

$$\frac{2\|f_0'\|_{L^\infty}A}{M}\gamma\leq\frac{\lambda}{8}\omega(\delta)\leq\frac{\lambda}{8}\omega(\xi).$$

Note that this is where we set a value for M, and that γ is taken sufficiently small depending on M. Now that the value for M is fixed, we can also control the value $\omega(M\xi)$ by taking γ sufficiently small that

$$\begin{aligned}\omega(M\xi)=\omega(\xi)+\int_{\xi}^{M\xi}\frac{\gamma}{h(4+\log(h/\delta))}\,dh\leq\omega(\xi)+\gamma\ln(M)\leq\omega(\xi)+\omega(\delta)\\ \leq 2\omega(\xi).\end{aligned}\tag{7-12}$$

Hence,

$$A\omega(\xi)\int_{M\xi}^{\infty}\frac{\omega(h)}{h^2}\,dh\leq\frac{\lambda}{16}\frac{\omega(M\xi)}{\xi}+\frac{\lambda}{8}\frac{\omega(\xi)}{\xi}\leq\frac{\lambda}{4}\frac{\omega(\xi)}{\xi}.\tag{7-13}$$

Using the same integration-by-parts tricks, we can also show

$$\lambda\int_{\xi}^{\infty}\frac{\omega(h+\xi)-\omega(h)-\omega(\xi)}{h^2}\,dh\leq-\frac{3}{4}\lambda\frac{\omega(\xi)}{\xi}\tag{7-14}$$

for γ sufficiently small.

So combining these, we get

$$\begin{aligned}A\omega(\xi)\int_{M\xi}^{\infty}\frac{\omega(h)}{h^2}\,dh+2(\Lambda-\lambda)\int_{2\xi}^{M\xi}\frac{\omega(h-\xi)-\omega(\xi)}{h^2}\,dh\\ +\lambda\int_{\xi}^{\infty}\frac{\omega(h+\xi)-\omega(h)-\omega(\xi)}{h^2}\,dh\leq-\frac{\lambda}{4}\frac{\omega(\xi)}{\xi}.\end{aligned}\tag{7-15}$$

Since $\omega'(\xi)\omega(\xi) \le \gamma\omega(\xi)/\xi$, we finally get

$$(A\ln(M+1)+1)\omega'(\xi)\omega(\xi) - \frac{\lambda}{4}\frac{\omega(\xi)}{\xi} \le \frac{\omega(\xi)}{\xi}\left((A\ln(M+1)+1)\gamma - \frac{\lambda}{4}\right) < 0, \tag{7-16}$$

if γ is taken sufficiently small. □

8. Our choice for the modulus ρ

We've now shown that for the modulus defined in (7-2) if the assumptions (4-1) hold then

$$\frac{d}{dt}(f_x(t,\xi/2) - f_x(t,-\xi/2))\big|_{t=T} < -\omega'(\xi)\omega(\xi). \tag{8-1}$$

We claim that in fact (8-1) will hold for any rescaling $\omega_r(h) = \omega(rh)$ as well. To see this, fix some $r > 0$, and suppose that $f(t,x)$ satisfies the conditions of Lemma 4.1 for ω_r at time T and distance ξ. Take $\tilde{f}(t,x) = rf(t/r, x/r)$, which is also a solution of (1-2). Then $\tilde{f}_x$ is a solution of (3-5) with $\beta(\tilde{f}_0') = \beta(f_0')$, $\|\tilde{f}_0'\|_{L^\infty} = \|f_0'\|_{L^\infty}$, and satisfying the conditions of Lemma 4.1 for ω at time rT and distance $r\xi$. Hence by Lemma 7.1

$$\begin{aligned}\frac{d}{dt}(f_x(t,\xi/2) - f_x(t,-\xi/2))\big|_{t=T} &= r\frac{d}{dt}(\tilde{f}_x(t,r\xi/2) - \tilde{f}_x(t,-r\xi/2))\big|_{t=rT} \\ &< -r\omega'(r\xi)\omega(r\xi) = -\omega_r'(\xi)\omega_r(\xi).\end{aligned} \tag{8-2}$$

So, (8-1) will hold for any rescaling ω_r. Also note that for $f_x(T,\xi/2) - f_x(T,-\xi/2) = \omega(\xi)$ to hold, we must necessarily have $\omega(\xi) \le 2\|f_x(T,\cdot)\|_{L^\infty} < 2\|f_0'\|_{L^\infty}$. Thus taking

$$C = \sup_{0<h<\omega^{-1}(2\|f_0'\|_{L^\infty})} \frac{h}{\omega(h)} = \frac{\omega^{-1}(2\|f_0'\|_{L^\infty})}{2\|f_0'\|_{L^\infty}}, \tag{8-3}$$

we see that

$$\omega(h) \ge \frac{h}{C} \tag{8-4}$$

for all relevant h. Define

$$\rho(h) := \omega(Ch) \tag{8-5}$$

so that

$$\rho(h) \ge h \tag{8-6}$$

for all $h \in [0, \rho^{-1}(2\|f_0'\|_{L^\infty})]$.

Now, suppose that at time T, f satisfies the assumptions (4-1) for $\rho(\cdot/T)$. Then since $\rho(\cdot/T)$ is a rescaling of ω, we have

$$\begin{aligned}\frac{d}{dt}(f_x(T,\xi/2) - f_x(T,-\xi/2)) &< -\frac{d}{dh}\rho(h/T)\big|_{h=\xi}\rho(\xi/T) \\ &= -\frac{1}{T}\rho'(\xi/T)\rho(\xi/T) \le -\frac{\xi}{T^2}\rho'(\xi/T) = \frac{d}{dt}\rho(\xi/t)\big|_{t=T}.\end{aligned} \tag{8-7}$$

Thus we've constructed a modulus ρ which satisfies (2-12), completing the proof of the generation of a Lipschitz modulus of continuity (1-8) in our main theorem.

9. Regularity in time

With the construction of the modulus ρ, we get universal Lipschitz bounds in space for $f_x(t,\cdot)$. By the structure of (1-2), we also get regularity in space for f_t.

Proposition 9.1. *Let $f:(0,T)\times\mathbb{R}\to\mathbb{R}$ be a classical solution to (1-2) with $\|f(t,\cdot)\|_{W^{1,\infty}}$ bounded and $\|f_{xx}(t,\cdot)\|_{L^\infty}\lesssim 1/t$. Then $f_t(t,\cdot)$ is log-Lipschitz in space with*

$$\begin{gathered}|f_t(t,\cdot)|\lesssim \max\{-\log(t),1\},\\ |f_t(t,x)-f_t(t,y)|\lesssim -\log(|x-y|)|x-y|\left(1+\frac{1}{t}\right),\quad 0<|x-y|<\tfrac{1}{2}.\end{gathered}\tag{9-1}$$

Proof. For $t<1$, we have

$$\begin{aligned}|f_t(t,x)|&=\left|\int_{\mathbb{R}}\frac{\delta_h f(t,x)-hf_x(t,x)}{\delta_h f(t,x)^2+h^2}\,dh\right|\\ &\le\left|\int_0^\infty\frac{\delta_h f(t,x)+\delta_{-h}f(t,x)}{\delta_{-h}f(t,x)^2+h^2}\,dh\right|+\left|\int_0^\infty\frac{(\delta_h f(t,x)-hf_x(t,x))(\delta_h f(t,x)^2-\delta_{-h}f(t,x)^2)}{(\delta_h f(t,x)^2+h^2)(\delta_{-h}f(t,x)^2+h^2)}\,dh\right|\\ &\lesssim\int_0^t\frac{1}{t}\,dh+\int_t^1\frac{1}{h}\,dh+\int_1^\infty\frac{1}{h^2}+\frac{1}{h^3}\,dh\\ &\lesssim-\log(t)+1.\end{aligned}\tag{9-2}$$

For $t>1$, we can similarly show $|f_t(t,x)|\lesssim 1$, proving the first bound.

For regularity in space, we see that

$$\begin{aligned}&f_t(t,x)-f_t(t,y)\\ &=\int_{\mathbb{R}}\frac{\delta_h f(t,x)-hf_x(t,x)}{\delta_h f(t,x)^2+h^2}-\frac{\delta_h f(t,y)-hf_x(t,y)}{\delta_h f(t,y)^2+h^2}\,dh\\ &=\int_{\mathbb{R}}\frac{\delta_h f(t,x)-hf_x(t,x)-(\delta_h f(t,y)-hf_x(t,y))}{\delta_h f(t,y)^2+h^2}+\frac{(\delta_h f(t,x)-hf_x(t,x))(\delta_h f(t,x)^2-\delta_h f(t,y)^2)}{(\delta_h f(t,x)^2+h^2)(\delta_h f(t,y)^2+h^2)}\,dh\\ &\le\left|\int_{|h|<|x-y|}\right|+\left|\int_{|x-y|<|h|<1}\right|+\left|\int_{|h|>1}\right|.\end{aligned}\tag{9-3}$$

For $|h|<|x-y|$, we can bound much as before to get

$$\left|\int_{|h|<|x-y|}\right|\lesssim\int_0^{|x-y|}\frac{1}{t}\,dh=\frac{|x-y|}{t}.\tag{9-4}$$

For midsize $|x-y|<|h|<1$, we have

$$\begin{aligned}\left|\delta_h f(t,x)-hf_x(t,x)-(\delta_h f(t,y)-hf_x(t,y))\right| &= \left|\int_0^h \delta_s f_x(t,x)-\delta_s f_x(t,y)\,ds\right| \lesssim \frac{|x-y|h}{t},\\ \left|\delta_h f(t,x)-\delta_h f(t,y)\right| &= \left|\int_0^h f_x(t,x+s)-f_x(t,y+s)\,ds\right| \lesssim \frac{|x-y|h}{t}.\end{aligned}\tag{9-5}$$

Thus

$$\left|\int_{|x-y|<|h|<1}\right| \lesssim \frac{|x-y|}{t}\int_{|x-y|}^{1}\frac{1}{h}\,dh = \frac{-\ln(|x-y|)|x-y|}{t}.\tag{9-6}$$

Finally, we use L^∞ bounds on f to get

$$\begin{aligned}\left|\int_{|h|>1}\right| &\le \left|\int_{|h|>1}\frac{\delta_h f(t,x)-\delta_h f(t,y)}{\delta_h f(t,y)^2+h^2}+\frac{(\delta_h f(t,x)-hf_x(t,x))(\delta_h f(t,x)^2-\delta_h f(t,y)^2)}{(\delta_h f(t,x)^2+h^2)(\delta_h f(t,y)^2+h^2)}\,dh\right|\\ &\qquad + |f_x(t,x)-f_x(t,y)|\left|\int_{|h|>1}\frac{-h}{\delta_h f(t,y)^2+h^2}\,dh\right|\\ &\lesssim |x-y|\int_1^\infty \frac{1}{h^2}+\frac{1}{h^3}\,dh+\frac{|x-y|}{t}\int_1^\infty\frac{1}{h^3}\,dh \lesssim \left(1+\frac{1}{t}\right)|x-y|.\end{aligned}\tag{9-7}$$

Putting this all together, we thus have

$$|f_t(t,x)-f_t(t,y)| \lesssim -\ln(|x-y|)|x-y|\left(1+\frac{1}{t}\right),\tag{9-8}$$

completing the proof. □

Recall that in Section 2, we assumed that our initial data f_0 was in $C_c^\infty(\mathbb{R})$ so that, by the local existence results of [Córdoba and Gancedo 2009], there was a unique solution $f \in C^1((0,T_+);H^k)$ for k arbitrarily large and some $T_+>0$. We were then able to prove the existence of the modulus ρ as in Theorem 1.1 depending only on $\beta(f_0')$, $\|f_0'\|_{L^\infty}$, and hence with the solution f existing for all time by the main theorem of [Constantin et al. 2017]. For an arbitrary $f_0 \in W^{1,\infty}(\mathbb{R})$ with $\beta(f_0')<1$, the same result holds true by compactness. Let $\eta \in C_c^\infty(\mathbb{R})$ be a smooth mollifier and $\phi \in C_c^\infty(\mathbb{R})$ be a smooth cutoff function. For $f_0 \in W^{1,\infty}(\mathbb{R})$ with $\beta(f_0')<1$, take $f_0^{(\epsilon)}(x) := (f_0 * \eta_\epsilon)(x)\phi(\epsilon x)$. Then $f_0^{(\epsilon)} \to f_0$ in $W^{1,\infty}_{\mathrm{loc}}$, with $\beta(f_0^{(\epsilon)\prime}) \to \beta(f_0')$ and $\|f_0^{(\epsilon)}\|_{W^{1,\infty}(\mathbb{R})} \to \|f_0\|_{W^{1,\infty}(\mathbb{R})}$ as $\epsilon \to 0$. Thus for ϵ sufficiently small, $\beta(f_0^{(\epsilon)\prime})<1$ and the results of the previous section hold for the solution to the mollified problem $f^{(\epsilon)}$. The L^∞ bound on $f_t^{(\epsilon)}$ proven above along with the maximum principle for $f_x^{(\epsilon)}$ is enough to ensure that there a subsequence $f^{(\epsilon_k)}$ converging in $C_{\mathrm{loc}}([0,\infty)\times\mathbb{R})$ to a Lipschitz (weak) solution f to the original problem. In order to get a classical C^1 solution, we need regularity estimates for $f_x^{(\epsilon)}$, $f_t^{(\epsilon)}$ in both time and space. The modulus ρ and Proposition 9.1 give the regularity in space that we need for f_x, f_t. All that leaves is to prove regularity in time.

Proposition 9.2. *Let f be a sufficiently smooth solution to (1-2) with $\beta(f_0') < 1$. Then $f_x, f_t \in C^\alpha_{\text{loc}}((0,\infty)\times\mathbb{R})$ with*

$$\|f_x\|_{C^\alpha(Q_{t/4}(t,x))}, \|f_t\|_{C^\alpha(Q_{t/4}(t,x))} \le C(\beta(f_0'), \|f\|_{L^\infty_t((t/2,3t/2);W^{2,\infty}_x(\mathbb{R}))})\max\{t^{-\alpha},1\}, \tag{9-9}$$

where $Q_r(s,y) = (s-r,s]\times B_r(y)$, and $\alpha > 0$ depends only on $\beta(f_0')$, $\|f_0'\|_{L^\infty}$.

Proof. We have that f_x solves

$$(f_x)_t(t,x) = f_{xx}(t,x)\int_{\mathbb{R}} \frac{-h}{\delta_h f(t,x)^2+h^2}\,dh + \int_{\mathbb{R}} \delta_h f_x(t,x) K(t,x,h)\,dh, \tag{9-10}$$

where $\lambda/h^2 \le K(t,x,h) \le \Lambda/h^2$ is uniformly elliptic with ellipticity constants λ, Λ depending on $\beta(f_0')$, $\|f_0'\|_{L^\infty}$. Rewriting this, we have that f_x satisfies

$$\begin{aligned}(f_x)_t - \int_{\mathbb{R}} \delta_h f_x(t,x)\left(\frac{K(t,x,h)+K(t,x,-h)}{2}\right)dh \\ = f_{xx}(t,x)\int_{\mathbb{R}} \frac{-h}{\delta_h f(t,x)^2+h^2}\,dh + \int_{\mathbb{R}} \delta_h f_x(t,x)\left(\frac{K(t,x,h)-K(t,x,-h)}{2}\right)dh.\end{aligned} \tag{9-11}$$

Let $F(t,x)$ denote the right-hand side of (9-11). Then $F(t,x)$ is locally bounded with $|F(t,x)|$ controlled by $\|f(t,\cdot)\|_{W^{2,\infty}}$. Then since $(K(t,x,h)+K(t,x,-h))/2$ is a symmetric uniformly elliptic kernel, it follows that we have local C^α bounds for $\alpha \le \alpha_0$ for some α_0 depending on ellipticity constants; see [Silvestre 2011].

So, all we have to do is give bounds on $F(t,x)$ depending only on $\|f(t,\cdot)\|_{W^{2,\infty}}$. Similar to the proof of Lemma 5.1,

$$\int_{\mathbb{R}} \frac{-h}{\delta_h f(t,x)^2+h^2}\,dh = \int_0^\infty h\frac{\delta_h f(t,x)^2-\delta_{-h}f(t,x)^2}{(\delta_h f(t,x)^2+h^2)(\delta_{-h}f(t,x)^2+h^2)}\,dh \lesssim \int_0^1 1\,dh + \int_1^\infty \frac{1}{h^3}\,dh \lesssim 1. \tag{9-12}$$

Also similar to the proof of Lemma 5.2, specifically (5-12), we have

$$|K(t,x,h)-K(t,x,-h)| \lesssim \min\left\{\frac{1}{h},\frac{1}{h^3}\right\}, \tag{9-13}$$

so

$$\left|\int_{\mathbb{R}} \delta_h f_x(t,x)\left(\frac{K(t,x,h)-K(t,x,-h)}{2}\right)dh\right| \lesssim \int_0^1 1\,dh + \int_1^\infty \frac{1}{h^3}\,dh \lesssim 1. \tag{9-14}$$

Thus since we've bounded the right-hand side of (9-11) depending only on $\|f(t,\cdot)\|_{W^{2,\infty}}$, we have our local C^α bounds for f_x for all α sufficiently small. A C^α bound that is uniform in x for f_x then gives a log C^α estimate for f_t, similar to the proof for regularity in space in Proposition 9.1. Thus we have C^α estimates for both f_x, f_t. □

Appendix: Uniqueness

We now prove that if our initial data f_0 is in $C^{1,\epsilon}(\mathbb{R})$ with $\beta(f_0') < 1$, then the solution f given by Theorem 1.1 is unique with $f \in L^\infty([0,\infty); C^{1,\epsilon})$. As mentioned before, this essentially follows from the uniqueness theorem given in [Constantin et al. 2017], which under our assumptions simplifies to:

Theorem A.1 [Constantin et al. 2017]. *Let $f \in L^\infty([0,T]; W^{1,\infty})$ be a classical, C^1 solution to (1-2) with initial data $f(0,x) = f_0(x)$. Assume that $\lim_{x\to\infty} f(t,x) = 0$, and that there is some modulus of continuity $\tilde{\rho}$ such that*

$$f_x(t,x) - f_x(t,y) \le \tilde{\rho}(|x-y|) \quad \textit{for all } 0 \le t \le T,\ x \ne y \in \mathbb{R}. \tag{A-1}$$

Then the solution f is unique.

The authors of [Constantin et al. 2017] note that the uniform continuity assumption should be the only real assumption; the decay is assumed for convenience in their proof. So, we start by proving that if $f_0 \in C^{1,\epsilon}(\mathbb{R})$, then the solution f is in $L^\infty([0,\infty); C^{1,\epsilon})$. To begin, suppose that $f_0 \in C^{1,1}(\mathbb{R})$. Then necessarily f_0' has modulus $\rho(\,\cdot\,/\delta)$ for some $\delta > 0$ sufficiently small. The same proof for the instantaneous generation of the modulus ρ will give that $f_x(t,\cdot)$ has modulus $\rho(\,\cdot\,/t+\delta)$. Hence $f_x(t,\cdot)$ has modulus $\rho(\,\cdot\,/\delta)$ for all $t \ge 0$.

If $f_0 \in C^{1,\epsilon}(\mathbb{R})$, we can make the same essential argument by changing the definitions of ρ, ω. You can repeat the arguments of Sections 7 and 8 for the modulus

$$\begin{aligned} \omega^{(\epsilon)}(\xi) &= \xi^\epsilon, && 0 \le \xi \le \delta, \\ \omega^{(\epsilon)\prime}(\xi) &= \frac{\gamma}{\xi(4+\log(\xi/\delta))}, && \xi \ge \delta. \end{aligned} \tag{A-2}$$

All the error terms for $\xi \le \delta$ are of order $\xi^{2\epsilon-1}$, while the diffusion term is of order $\xi^{\epsilon-1}$, so there are no problems as long as δ is sufficiently small. The argument for $\xi \ge \delta$ is identical to the original. Taking $\rho^{(\epsilon)}$ to be some suitable rescaling of $\omega^{(\epsilon)}$, we then have that if f_0' has modulus $\rho^{(\epsilon)}(\,\cdot\,/\delta)$, then $f_x(t,\cdot)$ will have modulus $\rho^{(\epsilon)}(\,\cdot\,/t+\delta)$.

Thus if $f_0 \in C^{1,\epsilon}(\mathbb{R})$, then the solution f given by Theorem 1.1 will satisfy the main uniform continuity assumption of Theorem A.1. Our solution f will not decay as $x \to \infty$, but that assumption isn't truly necessary.

Let f_1, f_2 be two uniformly continuous, classical solutions to (1-2) with the same initial data, and let $M(t) = \|f_1(t,\cdot) - f_2(t,\cdot)\|_{L^\infty}$. With the decay assumption, the authors of [Constantin et al. 2017] are able to assume that for almost every t, there is a point $x(t) \in \mathbb{R}$ such that

$$M(t) = |f_1(t,x(t)) - f_2(t,x(t))|, \qquad \frac{d}{dt}M(t) = \Big(\frac{d}{dt}|f_1 - f_2|\Big)(t,x(t)). \tag{A-3}$$

They then bound $\frac{d}{dt}|f_1(t,x(t)) - f_2(t,x(t))|$ using equation (1-2), $\tilde{\rho}$, and $W^{1,\infty}$ bounds.

Without the decay assumption, we instead use that

$$\frac{d}{dt}M(t) \le \sup\Big\{\frac{d}{dt}|f_1(t,x) - f_2(t,x)| : |f_1(t,x) - f_2(t,x)| \ge M(t) - \delta\Big\}, \tag{A-4}$$

where $\delta > 0$ is arbitrary. When we go to bound $\frac{d}{dt}|f_1(t,x) - f_2(t,x)|$, we then get new error terms which can be bounded by

$$C\big(\tilde{\rho}, \max_i \|f_i(t,\cdot)\|_{W^{1,\infty}}, M(t)\big)(\delta + |f_{1,x}(t,x) - f_{2,x}(t,x)|). \tag{A-5}$$

Since $f_{i,x}(t,x)$ is bounded and has modulus $\tilde{\rho}$, it then follows that

$$|f_{1,x}(t,x) - f_{2,x}(t,x)| = o_\delta(1). \tag{A-6}$$

Thus by taking δ sufficiently small depending on $\tilde{\rho}$, $\max_i \|f_i(t,\cdot)\|_{W^{1,\infty}}$, $M(t)$, we can guarantee that the new error terms $\lesssim M(t)$. Then the original proof of [Constantin et al. 2017] goes through.

Acknowledgement

I would like to thank my advisor Luis Silvestre for suggesting the problem, pointing me towards good resources, and just giving good advice in general.

References

[Castro et al. 2012] A. Castro, D. Córdoba, C. Fefferman, F. Gancedo, and M. López-Fernández, "Rayleigh–Taylor breakdown for the Muskat problem with applications to water waves", *Ann. of Math.* (2) **175**:2 (2012), 909–948. MR Zbl

[Constantin et al. 2013] P. Constantin, D. Córdoba, F. Gancedo, and R. M. Strain, "On the global existence for the Muskat problem", *J. Eur. Math. Soc.* (*JEMS*) **15**:1 (2013), 201–227. MR Zbl

[Constantin et al. 2016] P. Constantin, D. Córdoba, F. Gancedo, L. Rodríguez-Piazza, and R. M. Strain, "On the Muskat problem: global in time results in 2D and 3D", *Amer. J. Math.* **138**:6 (2016), 1455–1494. MR Zbl

[Constantin et al. 2017] P. Constantin, F. Gancedo, R. Shvydkoy, and V. Vicol, "Global regularity for 2D Muskat equations with finite slope", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **34**:4 (2017), 1041–1074. MR Zbl

[Córdoba and Gancedo 2007] D. Córdoba and F. Gancedo, "Contour dynamics of incompressible 3-D fluids in a porous medium with different densities", *Comm. Math. Phys.* **273**:2 (2007), 445–471. MR Zbl

[Córdoba and Gancedo 2009] D. Córdoba and F. Gancedo, "A maximum principle for the Muskat problem for fluids with different densities", *Comm. Math. Phys.* **286**:2 (2009), 681–696. MR Zbl

[Dabkowski et al. 2014] M. Dabkowski, A. Kiselev, L. Silvestre, and V. Vicol, "Global well-posedness of slightly supercritical active scalar equations", *Anal. PDE* **7**:1 (2014), 43–72. MR Zbl

[Deng et al. 2017] F. Deng, Z. Lei, and F. Lin, "On the two-dimensional Muskat problem with monotone large initial data", *Comm. Pure Appl. Math.* **70**:6 (2017), 1115–1145. MR Zbl

[Gancedo 2017] F. Gancedo, "A survey for the Muskat problem and a new estimate", *SeMA J.* **74**:1 (2017), 21–35. MR Zbl

[Kiselev 2010] A. Kiselev, "Regularity and blow up for active scalars", *Math. Model. Nat. Phenom.* **5**:4 (2010), 225–255. MR Zbl

[Kiselev et al. 2007] A. Kiselev, F. Nazarov, and A. Volberg, "Global well-posedness for the critical 2D dissipative quasi-geostrophic equation", *Invent. Math.* **167**:3 (2007), 445–453. MR Zbl

[Muskat 1934] M. Muskat, "Two fluid systems in porous media: the encroachment of water into an oil sand", *J. Appl. Phys.* **5**:9 (1934), 250–264. Zbl

[Patel and Strain 2017] N. Patel and R. M. Strain, "Large time decay estimates for the Muskat equation", *Comm. Partial Differential Equations* **42**:6 (2017), 977–999. MR Zbl

[Saffman and Taylor 1958] P. G. Saffman and G. Taylor, "The penetration of a fluid into a porous medium or Hele-Shaw cell containing a more viscous liquid", *Proc. Roy. Soc. London. Ser. A* **245** (1958), 312–329. MR Zbl

[Silvestre 2011] L. Silvestre, "On the differentiability of the solution to the Hamilton–Jacobi equation with critical fractional diffusion", *Adv. Math.* **226**:2 (2011), 2020–2039. MR Zbl

Received 9 May 2017. Revised 14 Jan 2018. Accepted 30 Jul 2018.

STEPHEN CAMERON: scameron@math.uchicago.edu
Department of Mathematics, University of Chicago, Chicago, IL, United States

ANALYSIS AND PDE
Vol. 12, No. 4, 2019
dx.doi.org/10.2140/apde.2019.12.1023

GLOBAL WELL-POSEDNESS AND SCATTERING FOR THE RADIAL, DEFOCUSING, CUBIC WAVE EQUATION WITH INITIAL DATA IN A CRITICAL BESOV SPACE

BENJAMIN DODSON

We prove that the cubic wave equation is globally well-posed and scattering for radial initial data lying in $B^2_{1,1} \times B^1_{1,1}$. This space of functions is a scale-invariant subspace of $\dot{H}^{1/2} \times \dot{H}^{-1/2}$.

1. Introduction

The three-dimensional cubic nonlinear wave equation,

$$u_{tt} - \Delta u = -u^3 = F(u), \qquad u(0,x) = u_0, \quad u_t(0,x) = u_1, \quad x \in \mathbb{R}^3, \tag{1-1}$$

has been a topic of recent interest in the study of dispersive partial differential equations. This is due to the fact that the Hamiltonian for (1-1),

$$E(u(t)) = \frac{1}{2}\int |\nabla u(t,x)|^2\,dx + \frac{1}{2}\int u_t(t,x)^2\,dx + \frac{1}{4}\int u(t,x)^4\,dx = E(u(0)), \tag{1-2}$$

does not control the critical Sobolev norm.

A solution to (1-1) obeys the scaling symmetry that if $u(t,x)$ solves (1-1), then for any $\lambda > 0$

$$\lambda u(\lambda t, \lambda x) \tag{1-3}$$

also solves (1-1) with initial data $(\lambda u_0(\lambda x), \lambda^2 u_1(\lambda x))$. It is a general rule that, for any dimension $d \geq 1$,

$$\|u_0\|_{\dot{H}^{(d-2)/2}(\mathbb{R}^d)} = \|\lambda u_0(\lambda x)\|_{\dot{H}^{(d-2)/2}(\mathbb{R}^d)}, \quad \|u_1\|_{\dot{H}^{(d-4)/2}(\mathbb{R}^d)} = \|\lambda^2 u_1(\lambda x)\|_{\dot{H}^{(d-4)/2}(\mathbb{R}^d)}. \tag{1-4}$$

Thus in three dimensions (1-1) is called $\dot{H}^{1/2}$-critical.

Local well-posedness theory for (1-1) in L^2-based Sobolev spaces is completely determined by the critical $s_c = \frac{1}{2}$.

Negatively, using the arguments found in [Christ et al. 2003; Lindblad and Sogge 1995], one can show that the initial value problem (1-1) fails to be even locally well-posed for data lying in spaces less regular than $\dot{H}^{1/2} \times \dot{H}^{-1/2}$, that is, any space $\dot{H}^s \times \dot{H}^{s-1}$, $s < \frac{1}{2}$.

Positively:

Lemma 1.1. *Equation (1-1) is locally well-posed in $\dot{H}^s \times \dot{H}^{s-1}$ for any $s \geq \frac{1}{2}$.*

Proof. See [Lindblad and Sogge 1995]. □

MSC2010: 35L05, 35B40.
Keywords: defocusing, nonlinear wave equation, scattering, global well-posedness.

Local well-posedness is defined in the usual way.

Definition (locally well-posed). The initial value problem (1-1) is said to be locally well-posed if there exists an open interval $I \subset \mathbb{R}$ containing 0 such that:

(1) A unique solution $u \in L_t^\infty \dot{H}^s(I \times \mathbb{R}^3) \cap L_{t,\mathrm{loc}}^4 L_x^4(I \times \mathbb{R}^3)$, $u_t \in L_t^\infty \dot{H}^{s-1}(I \times \mathbb{R}^3)$ exists.

(2) u is continuous in time, $u \in C(I; \dot{H}^s(\mathbb{R}^3))$, $u_t \in C(I; \dot{H}^{s-1}(\mathbb{R}^3))$.

(3) u depends continuously on the initial data. That is, for any compact $J \subset I$, if $\|u_0 - u_0^*\|_{\dot{H}^s} < \epsilon$ and $\|u_1 - u_1^*\|_{\dot{H}^{s-1}} < \epsilon$ for some $\epsilon < \epsilon_0(J) > 0$ sufficiently small, then

$$\|u^* - u\|_{L_{t,x}^4(J \times \mathbb{R}^3)} + \|u^* - u\|_{L_t^\infty \dot{H}^s(J \times \mathbb{R}^3)} + \|u_t^* - u_t\|_{L_t^\infty \dot{H}^{s-1}(J \times \mathbb{R}^3)} \le \delta(\epsilon), \tag{1-5}$$

where u is the unique solution with initial data (u_0, u_1) and u^* is the solution with initial data (u_0^*, u_1^*) and $\delta(\epsilon)$ is a continuous function of ϵ with $\delta(0) = 0$.

The defocusing, energy-critical nonlinear wave equation, obtained from (1-1) either by changing $-u^3$ to $-u^5$ or by changing from three dimensions to four has now been completely worked out. For initial data in the energy class (which occurs for $u_0 \in \dot{H}^1$ and $u_1 \in L^2$) a priori bounds on scattering norms and concentration compactness properties of solutions have been established in [Struwe 1988; Grillakis 1990; Ginibre et al. 1992; Shatah and Struwe 1993; Bahouri and Shatah 1998; Bahouri and Gérard 1999; Nakanishi 1999; Tao 2006b].

Remark. The focusing case (obtained by changing the sign of the nonlinearity) is considerably more complicated. Focusing problems are not addressed at all in this paper, and so the interested reader is referred to [Kenig 2015].

For the radial version of (1-1) in three dimensions,

$$u_{tt} - u_{rr} - \frac{2}{r}u_r + u^3 = u_{tt} - \frac{1}{r}\partial_{rr}(ru) + u^3 = 0, \qquad u(0,r) = u_0(r), \quad u_t(0,r) = u_1(r), \tag{1-6}$$

the lack of control of the $\dot{H}^{1/2} \times \dot{H}^{-1/2}$ norm is the only obstacle to proving global well-posedness and scattering for (1-6). Indeed:

Theorem 1.2. *Suppose $u_0 \in \dot{H}^{1/2}(\mathbb{R}^3)$ and $u_1 \in \dot{H}^{-1/2}(\mathbb{R}^3)$ are radial functions, and u solves (1-1) on a maximal interval $0 \in I \subset \mathbb{R}$ with*

$$\sup_{t \in I} \|u(t)\|_{\dot{H}^{1/2}(\mathbb{R}^3)} + \|u_t(t)\|_{\dot{H}^{-1/2}(\mathbb{R}^3)} < \infty. \tag{1-7}$$

Then $I = \mathbb{R}$ and the solution u scatters both forward and backward in time.

Proof. See [Dodson and Lawrie 2015]. □

Scattering is also defined in the usual way.

Definition (scattering). A solution to (1-1) is said to scatter forward in time if there exist some $u_0^+ \in \dot{H}^{1/2}$, $u_1^+ \in \dot{H}^{-1/2}$ such that

$$\lim_{t \to \infty} \|u(t) - S(t)(u_0^+, u_1^+)\|_{\dot{H}^{1/2}(\mathbb{R}^3)} + \|u_t(t) - \partial_t S(t)(u_0^+, u_1^+)\|_{\dot{H}^{-1/2}(\mathbb{R}^3)} = 0, \tag{1-8}$$

where $u(t) = S(t)(u_0, u_1)$ is the solution to the linear wave equation

$$u_{tt} - \Delta u = 0, \qquad u(0, x) = u_0, \quad u_t(0, x) = u_1. \tag{1-9}$$

A solution to (1-1) is said to scatter backward in time if there exist $u_0^- \in \dot{H}^{1/2}$, $u_1^- \in \dot{H}^{-1/2}$ such that

$$\lim_{t \to -\infty} \|u(t) - S(t)(u_0^-, u_1^-)\|_{\dot{H}^{1/2}(\mathbb{R}^3)} + \|u_t(t) - \partial_t S(t)(u_0^-, u_1^-)\|_{\dot{H}^{-1/2}(\mathbb{R}^3)} = 0. \tag{1-10}$$

A solution which scatters both forward and backward in time is called scattering.

In this paper, (1-1) is proved to be globally well-posed and scattering for initial data lying in a critical space. The proof of global well-posedness is fairly general, and could be applied to a broad range of nonlinearities. The proof of scattering utilizes hyperbolic coordinates and relies on the fact that the cubic exponent $3 = (d+3)/(d-1)$ is the conformal exponent in three dimensions.

Hyperbolic coordinates were utilized by Tataru [2001] to prove weighted Strichartz estimates that extended previous results of [Georgiev et al. 1997]. Miao et al. [2018] recently proved a result similar to Theorem 1.4 for the five-dimensional problem, also for the conformal exponent. The conformal exponent is $\dot{H}^{1/2}$-critical, and it is straightforward to prove that the energy of a solution to (1-1) in hyperbolic coordinates scales like the $\dot{H}^{1/2} \times \dot{H}^{-1/2}$ norm.

Recently, Shen [2017], also working in hyperbolic coordinates, was able to prove a scattering result for data lying in a weighted energy space. Later, Dodson [2016] combined the result of Shen [2017] with the I-method, proving:

Theorem 1.3. *Suppose there exists a positive constant $\epsilon > 0$ such that*

$$\|u_0\|_{\dot{H}^{1/2+\epsilon}(\mathbb{R}^3)} + \||x|^{2\epsilon} u_0\|_{\dot{H}^{1/2+\epsilon}(\mathbb{R}^3)} \le A < \infty, \tag{1-11}$$

$$\|u_1\|_{\dot{H}^{-1/2+\epsilon}(\mathbb{R}^3)} + \||x|^{2\epsilon} u_1\|_{\dot{H}^{-1/2+\epsilon}(\mathbb{R}^3)} \le A < \infty. \tag{1-12}$$

Then (1-1) has a global solution and there exists some $C(A, \epsilon) < \infty$ such that

$$\int_{\mathbb{R}} \int (u(t,x))^4 \, dx \, dt \le C(A, \epsilon), \tag{1-13}$$

which proves that u scatters both forward and backward in time.

Remark. A straightforward application of the Strichartz estimates of [Ginibre and Velo 1995; Strichartz 1977] shows that

$$\|u\|_{L^4_{t,x}(\mathbb{R} \times \mathbb{R}^3)} < \infty \tag{1-14}$$

is equivalent to scattering.

Note that conditions (1-11) and (1-12) fall just short of lying in the critical Sobolev space $\dot{H}^{1/2} \times \dot{H}^{-1/2}$, and are not quite invariant under the scaling (1-3). In this paper we will study the radial, nonlinear wave equation in three dimensions,

$$u_{tt} - u_{rr} - \frac{2}{r} u_r + u^3 = 0, \qquad u_0 \in B^2_{1,1}, \quad u_1 \in B^1_{1,1}. \tag{1-15}$$

The Besov spaces $B^s_{q,r}$ will be defined in the next section. By the Sobolev embedding theorem, this space is a subspace of $\dot{H}^{1/2} \times \dot{H}^{-1/2}$, and the norm is invariant under (1-3).

We believe that this is the first result in which large-data scattering was proved for initial data in a scale-invariant space for which the norm was not controlled by a conserved quantity.

Theorem 1.4. *The initial value problem* (1-1) *is globally well-posed and scattering for* $u_0 \in B^2_{1,1}(\mathbb{R}^3)$, *radial, and* $u_1 \in B^1_{1,1}(\mathbb{R}^3)$, *radial. Moreover,*

$$\|u\|_{L^4_{t,x}(\mathbb{R}\times\mathbb{R}^3)} \le C(\|u_0\|_{B^2_{1,1}}, \|u_1\|_{B^1_{1,1}}). \tag{1-16}$$

The proof of this theorem utilizes the fact that the free solution with such initial data is only singular at the origin in space and time, $t = 0$ and $x = 0$. Thus, using a Gronwall-type inequality, the local solution to (1-1) can be extended to a global solution that is the sum of a solution to the free wave equation combined with a finite energy term. A Morawetz estimate in hyperbolic coordinates then proves scattering.

The proof of Theorem 1.4 will occupy the remainder of this paper. In Section 2 we will begin by defining the Besov spaces and recalling basic Strichartz estimates. Then in Section 3 the local theory of (1-1) will be discussed. Global well-posedness will then be proved in Section 4. In Section 5 we will switch to hyperbolic coordinates to prove scattering. Finally in Section 6 we will use a profile decomposition to show that the bounds obtained for any $u_0 \in B^2_{1,1}$, $u_1 \in B^1_{1,1}$ depend only on size.

2. Besov spaces and linear estimates

We now present some harmonic analysis estimates that will be used in this paper. None of these results are new.

Theorem 2.1 (Hardy–Littlewood–Sobolev inequality). *For any* $0 < s < 1$, *if*

$$\frac{1}{q} = \frac{1}{p} + s - 1,$$

then

$$\left\| \frac{1}{|t|^s} * F(t) \right\|_{L^q(\mathbb{R})} \lesssim_s \|F\|_{L^q(\mathbb{R})}. \tag{2-1}$$

Definition (Littlewood–Paley decomposition). Let $\phi \in C_0^\infty(\mathbb{R}^3)$ be a radial, decreasing function supported on $|x| \le 2$ and $\phi(x) = 1$ for $|x| \le 1$. Then for any $j \in \mathbb{Z}$ let P_j be the Fourier multiplier

$$P_j f = \mathcal{F}^{-1}\big((\phi(2^{-j}\xi) - \phi(2^{-j+1}\xi))\hat{f}(\xi)\big), \tag{2-2}$$

where

$$\hat{f}(\xi) = (2\pi)^{-d/2} \int e^{-ix\cdot\xi} f(x)\,dx, \tag{2-3}$$

$$\mathcal{F}^{-1} g = (2\pi)^{-d/2} \int e^{ix\cdot\xi} g(\xi)\,d\xi. \tag{2-4}$$

Then for any Schwartz function f,

$$f = \sum_{j\in\mathbb{Z}} P_j f. \tag{2-5}$$

Let $K_j(x)$ be the kernel of the Littlewood–Paley multiplier P_j. By direct computation using stationary phase estimates, for any N,

$$|K_j(x)| \lesssim_{d,N} \frac{2^{jd}}{(1+2^j|x|)^N}. \tag{2-6}$$

This implies K_j has an L^1 norm that is uniformly bounded in j, so for any $1 \le p \le \infty$,

$$\|P_j f\|_{L^p(\mathbb{R}^d)} \lesssim_d \|f\|_{L^p(\mathbb{R}^d)}. \tag{2-7}$$

A direct computation also gives Bernstein's inequality

$$\|P_j f\|_{L^p(\mathbb{R}^d)} \lesssim_d 2^{-j} \|\nabla f\|_{L^p(\mathbb{R}^d)}, \tag{2-8}$$

along with the Sobolev embedding estimate, for $1 \le p \le q \le \infty$,

$$\|P_j f\|_{L^q(\mathbb{R}^d)} \lesssim_d 2^{jd\left(\frac{1}{p}-\frac{1}{q}\right)} \|f\|_{L^p(\mathbb{R}^d)}. \tag{2-9}$$

The Littlewood–Paley decomposition is foundational to the definition of Besov spaces.

Definition (Besov spaces). Suppose $1 \le p \le \infty$, $1 \le r \le \infty$, and $s \ge 0$. Then

$$\|f\|_{B^s_{r,p}(\mathbb{R}^d)} = \left(\sum_{j\in\mathbb{Z}} 2^{jsr} \|P_j f\|^r_{L^p(\mathbb{R}^d)}\right)^{1/r}. \tag{2-10}$$

The Besov space $B^s_{r,p}$ is then the completion of the Schwartz space under this norm. $B^s_{r,p}$ is a Banach space under this topology.

Remark. Observe that for any $s \in \mathbb{R}$ we have $B^s_{2,2}(\mathbb{R}^d) = \dot{H}^s(\mathbb{R}^d)$.

The Besov spaces are well-behaved with respect to multiplying by smooth cutoff functions.

Lemma 2.2. *Suppose $\chi(x) \in C_0^\infty(\mathbb{R}^3)$. Then*

$$\|\chi(x)u\|_{B^{1/2}_{1,2}(\mathbb{R}^3)} \lesssim \|u\|_{B^{1/2}_{1,2}(\mathbb{R}^3)}, \tag{2-11}$$

$$\|\chi(x)u\|_{B^{-1/2}_{1,2}(\mathbb{R}^3)} \lesssim \|u\|_{B^{-1/2}_{1,2}(\mathbb{R}^3)}. \tag{2-12}$$

Also if $\chi(x) = 1$ on $|x| \le 1$ then if $u_0 \in B^2_{1,1}$ and $u_1 \in B^1_{1,1}$, we have

$$\lim_{R\to\infty} \left\|\left(1-\chi\left(\frac{x}{R}\right)\right)u_0\right\|_{B^{1/2}_{1,2}(\mathbb{R}^3)} + \left\|\left(1-\chi\left(\frac{x}{R}\right)\right)u_1\right\|_{B^{-1/2}_{1,2}(\mathbb{R}^3)} = 0. \tag{2-13}$$

Proof. Split $P_j(\chi f)$ as

$$P_j(\chi f) = \chi(P_j f) + [P_j, \chi] f. \tag{2-14}$$

By Hölder's inequality,

$$\sum_j 2^{j/2} \|\chi(P_j f)\|_{L^2} \lesssim \sum_j 2^{j/2} \|P_j f\|_{L^2}, \tag{2-15}$$

so it only remains to compute

$$\sum_j 2^{j/2} \|[P_j, \chi] f\|_{L^2}. \tag{2-16}$$

By (2-6) and the fundamental theorem of calculus,

$$\int K_j(x-y)[\chi(y)f(y)-\chi(x)f(y)] \lesssim \sum_k \int |K_j(x-y)||x-y||P_k f(y)|\,dy, \tag{2-17}$$

and therefore by Bernstein's inequality and (2-6),

$$\sum_{j\geq 0} 2^{j/2}\|P_j(\chi P_{\geq 0} f)\|_{L^2} \lesssim \sum_{j\geq 0} 2^{-j/2} \sum_{k\geq 0} 2^{-k/2}\|P_k f\|_{\dot{H}^{1/2}} \lesssim \sum_k 2^{k/2}\|P_k f\|_{\dot{H}^{1/2}} \lesssim \|f\|_{B^{1/2}_{1,2}}. \tag{2-18}$$

Also by Bernstein's inequality, the Sobolev embedding theorem, and Hölder's inequality,

$$\sum_{j\geq 0} 2^{j/2}\|P_j(\chi P_{\leq 0} f)\|_{L^2} \lesssim \|\nabla(\chi(P_{\leq 0} f))\|_{L^2} \lesssim \|f\|_{B^{1/2}_{1,2}}. \tag{2-19}$$

Next, by Hölder's inequality in space,

$$\|P_{\leq 0}(\chi(P_{\leq 0} f))\|_{B^{1/2}_{1,2}} \lesssim \|P_{\leq 0} f\|_{L^6} \lesssim \|f\|_{B^{1/2}_{1,2}}. \tag{2-20}$$

Finally,

$$\|P_{\leq 0}(\chi(P_{\geq 0} f))\|_{B^{1/2}_{1,2}} \lesssim \sum_{k\geq 0} 2^{-k/2}\|P_k f\|_{\dot{H}^{1/2}} \lesssim \|f\|_{B^{1/2}_{1,2}}. \tag{2-21}$$

Combining (2-18)–(2-21), we have proved

$$\|\chi f\|_{B^{1/2}_{1,2}(\mathbb{R}^3)} \lesssim \|f\|_{B^{1/2}_{1,2}(\mathbb{R}^3)}. \tag{2-22}$$

Also observe that (2-18)–(2-21) imply

$$\|\chi f\|_{B^{1/2}_{\infty,2}} \lesssim \|f\|_{B^{1/2}_{\infty,2}}, \tag{2-23}$$

and therefore by duality

$$\|\chi g\|_{B^{-1/2}_{1,2}} \lesssim \|g\|_{B^{-1/2}_{1,2}}. \tag{2-24}$$

To prove (2-13) observe that $B^{1/2}_{1,2} \times B^{-1/2}_{1,2}$ is invariant under the scaling (1-3); that is,

$$\begin{aligned} &\left\|\left(1-\chi\left(\frac{x}{R}\right)\right)u_0\right\|_{B^{1/2}_{1,2}} + \left\|\left(1-\chi\left(\frac{x}{R}\right)\right)u_1\right\|_{B^{-1/2}_{1,2}} \\ &\qquad = R\|(1-\chi(x))u_0(Rx)\|_{B^{1/2}_{1,2}} + R^2\|(1-\chi(x))u_1(Rx)\|_{B^{-1/2}_{1,2}}. \end{aligned} \tag{2-25}$$

The dominated convergence theorem, (2-19), and (2-20) imply

$$\lim_{R\to\infty} R\|(1-\chi(x))P_{\leq 0}(u_0(Rx))\|_{B^{1/2}_{1,2}} + R^2\|(1-\chi(x))P_{\leq 0}(u_1(Rx))\|_{B^{-1/2}_{1,2}} = 0. \tag{2-26}$$

Meanwhile, (2-18), (2-21), and the dominated convergence theorem imply

$$\begin{aligned} &\lim_{R\to\infty} R\|(1-\chi(x))P_{\geq 0}(u_0(Rx))\|_{B^{1/2}_{1,2}} + R^2\|(1-\chi(x))P_{\geq 0}(u_1(Rx))\|_{B^{-1/2}_{1,2}} \\ &= \lim_{R\to\infty} R\sum_{j\geq 0} 2^{j/2}\|(1-\chi(x))P_j(u_0(Rx))\|_{L^2} + R^2\sum_{j\geq 0} 2^{-j/2}\|(1-\chi(x))P_j(u_1(Rx))\|_{L^2} = 0, \end{aligned} \tag{2-27}$$

completing the proof. ◻

Theorem 2.3 (radial Sobolev embedding theorem). *For any j,*

$$\| |x| P_j f \|_{L^\infty(\mathbb{R}^3)} \lesssim \| P_j f \|_{\dot{H}^{1/2}(\mathbb{R}^3)}. \tag{2-28}$$

Proof. Since f is radial, assume without loss of generality $x = (0, 0, |x|)$. Writing ξ in polar coordinates, $\xi = (r \cos\varphi \cos\theta, r \sin\varphi \cos\theta, r \sin\theta)$, $x \cdot \xi = |x| r \sin\theta$, so by the Fourier inversion formula,

$$f(x) = (2\pi)^{-1/2} \int_0^\infty \hat{f}(r) r^2 \int_{-\pi/2}^{\pi/2} e^{i|x|r \sin\theta} \cos\theta \, d\theta \, dr. \tag{2-29}$$

Then by a change of variables,

$$\begin{aligned} f(x) &= (2\pi)^{-1/2} \int_0^\infty r^2 \hat{f}(r) \int_{-1}^{1} e^{i|x|ru} \, du \, dr \\ &= (2\pi)^{-1/2} \frac{1}{i|x|} \int_0^\infty \hat{f}(r) r [e^{i|x|r} - e^{-i|x|r}] \, dr. \end{aligned} \tag{2-30}$$

Replacing f by $P_j f$,

$$i|x| P_j f(x) = (2\pi)^{-1/2} \int_0^\infty \widehat{P_j f}(r) r [e^{i|x|r} - e^{-i|x|r}] \, dr. \tag{2-31}$$

By Plancherel's theorem,

$$\int_0^\infty |\widehat{P_j f}(r)|^2 r^3 \, dr \sim \| f \|^2_{\dot{H}^{1/2}(\mathbb{R}^3)}, \tag{2-32}$$

so by Hölder's inequality and the support of P_j,

$$\text{(2-31)} \lesssim \| P_j f \|^2_{\dot{H}^{1/2}(\mathbb{R}^3)}, \tag{2-33}$$

completing the proof. □

Now observe that the solution to the free wave equation

$$u_{tt} - \Delta u = 0, \qquad u(0, x) = f(x), \quad u_t(0, x) = g(x), \tag{2-34}$$

is given by the Fourier multiplier

$$u(t, x) = \mathcal{F}^{-1}(\cos(t|\xi|) \hat{f}(\xi) + \frac{\sin(t|\xi|)}{|\xi|} \hat{g}(\xi)) = S(t)(f, g). \tag{2-35}$$

Then the solution to

$$u_{tt} - \Delta u = F, \qquad u(0, x) = f(x), \quad u_t(0, x) = g(x), \tag{2-36}$$

is given by

$$S(t)(f, g) + \int_0^t S(t - \tau)(0, F) \, d\tau. \tag{2-37}$$

Remark. Sometimes, if $u = S(t)(f, g)$ it is convenient to write

$$(u(t), \partial_t u(t)) = S(t)(f, g). \tag{2-38}$$

By standard stationary phase calculations:

Theorem 2.4 (dispersive estimate).

$$\|S(t)(f,g)\|_{L^\infty(\mathbb{R}^3)} \lesssim \frac{1}{t}[\|\nabla^2 f\|_{L^1(\mathbb{R}^3)} + \|\nabla g\|_{L^1(\mathbb{R}^3)}]. \tag{2-39}$$

This has been proved in many textbooks. See for example [Evans 2010].

The dispersive estimates can be used to prove Strichartz estimates.

Theorem 2.5. *Let $I \subset \mathbb{R}$, $t_0 \in I$, be an interval and let u solve the linear wave equation*

$$u_{tt} - \Delta u = F, \qquad u(t_0) = u_0, \quad u_t(t_0) = u_1. \tag{2-40}$$

Then we have the estimates

$$\begin{aligned}\|u\|_{L_t^p L_x^q(I\times\mathbb{R}^3)} + \|u\|_{L_t^\infty \dot H^s(I\times\mathbb{R}^3)} + \|u_t\|_{L_t^\infty \dot H^{s-1}(I\times\mathbb{R}^3)} \\ \lesssim_{p,q,s,\tilde p,\tilde q} \|u_0\|_{\dot H^s(\mathbb{R}^3)} + \|u_1\|_{\dot H^{s-1}(\mathbb{R}^3)} + \|F\|_{L_t^{\tilde p'} L_x^{\tilde q'}(I\times\mathbb{R}^3)},\end{aligned} \tag{2-41}$$

whenever $s \geq 0$, $2 \leq p, \tilde p \leq \infty$, $2 \leq q, \tilde q < \infty$, and

$$\frac{1}{p} + \frac{1}{q} \leq \frac{1}{2}, \qquad \frac{1}{\tilde p} + \frac{1}{\tilde q} \leq \frac{1}{2}. \tag{2-42}$$

Proof. See for example [Tao 2006a]. □

Remark. This theorem can easily be combined with the Christ–Kiselev lemma, see [Smith and Sogge 2000], and the fact that $|\nabla|$ commutes with the operator $(\partial_{tt} - \Delta)$ to prove many additional estimates.

Lemma 2.6 (perturbation lemma). *Let $I \subset \mathbb{R}$ be a time interval. Let $t_0 \in I$, $(u_0, u_1) \in \dot H^{1/2} \times \dot H^{-1/2}$ and M, A, A' be positive constants. Let $\tilde u$ solve the equation*

$$(\partial_{tt} - \Delta)\tilde u = F(\tilde u) = e \tag{2-43}$$

on $I \times \mathbb{R}^3$, and also suppose $\sup_{t\in I} \|(\tilde u(t), \partial_t \tilde u(t))\|_{\dot H^{1/2}\times \dot H^{-1/2}} \leq A$, $\|\tilde u\|_{L^4_{t,x}(I\times\mathbb{R}^3)} \leq M$,

$$\|(u_0 - \tilde u(t_0), u_1 - \partial_t \tilde u(t_0))\|_{\dot H^{1/2}\times \dot H^{-1/2}} \leq A', \tag{2-44}$$

and

$$\|e\|_{L^{4/3}_{t,x}(I\times\mathbb{R}^3)} + \|S(t-t_0)(u_0 - \tilde u(t_0), u_1 - \partial_t \tilde u(t_0))\|_{L^4_{t,x}(I\times\mathbb{R}^3)} \leq \epsilon. \tag{2-45}$$

Then there exists $\epsilon_0(M, A, A')$ such that if $0 < \epsilon < \epsilon_0$ then there exists a solution to (1-1) on I with $(u(t_0), \partial_t u(t_0)) = (u_0, u_1)$, $\|u\|_{L^4_{t,x}(I\times\mathbb{R}^3)} \leq C(M, A, A')$, and for all $t \in I$,

$$\|(u(t), \partial_t u(t)) - (\tilde u(t), \partial_t \tilde u(t))\|_{\dot H^{1/2}\times \dot H^{-1/2}} \leq C(A, A', M)(A' + \epsilon). \tag{2-46}$$

Proof. The method of proof is by now fairly well known. See for example Lemma 2.20 of [Kenig and Merle 2008]. □

3. Local theory

By the dominated convergence theorem, for any $u_0 \in B^2_{1,1}$, $u_1 \in B^1_{1,1}$, and $\delta > 0$ there exists some $j_0(\delta) < \infty$ such that

$$\sum_{j \geq j_0} 2^{2j} \|P_j u_0\|_{L^1(\mathbb{R}^3)} + \sum_{j \geq j_0} 2^j \|P_j u_1\|_{L^1(\mathbb{R}^3)} < \delta. \tag{3-1}$$

Then by the rescaling (1-3) with $\lambda = 2^{-j_0}$, if $u_{0,\lambda}(x) = \lambda u_0(\lambda x)$ and $u_{1,\lambda}(x) = \lambda^2 u_1(\lambda x)$,

$$\sum_{j \geq 0} 2^{2j} \|P_j u_{0,\lambda}\|_{L^1(\mathbb{R}^3)} + \sum_{j \geq 0} 2^j \|P_j u_{1,\lambda}\|_{L^1(\mathbb{R}^3)} < \delta. \tag{3-2}$$

To simplify notation let u_0 and u_1 refer to the $u_{0,\lambda}$ and $u_{1,\lambda}$ such that (3-2) holds.

Lemma 3.1. *Fix $\epsilon_0 > 0$ small. There exists some $\delta(\epsilon, \|u_0\|_{B^2_{1,1}}, \|u_1\|_{B^1_{1,1}}) > 0$ such that if (3-2) holds then*

$$\|u\|_{L^4_{t,x}([-\delta,\delta]\times\mathbb{R}^3)} \lesssim \epsilon, \tag{3-3}$$

$$\|u\|_{L^\infty_t B^{1/2}_{1,2}([-\delta,\delta]\times\mathbb{R}^3)} \lesssim \|u_0\|_{B^2_{1,1}(\mathbb{R}^3)} + \|u_1\|_{B^1_{1,1}(\mathbb{R}^3)}. \tag{3-4}$$

Proof. Assume that (3-2) holds for some $\delta \ll \epsilon_0$. By the Sobolev embedding theorem and the definition of Besov spaces (see page 1027),

$$\|S(t)(P_{\leq 0}u_0, P_{\leq 0}u_1)\|_{L^4_x(\mathbb{R}^3)} \lesssim \|u_0\|_{B^2_{1,1}(\mathbb{R}^3)} + \|u_1\|_{B^1_{1,1}(\mathbb{R}^3)}, \tag{3-5}$$

while by Theorem 2.5, (3-1), and (3-2),

$$\|S(t)(P_{\geq 0}u_0, P_{\geq 0}u_1)\|_{L^4_{t,x}(\mathbb{R}\times\mathbb{R}^3)} \lesssim \delta. \tag{3-6}$$

Taking $\delta > 0$ sufficiently small, (3-5) and (3-6) imply

$$\|S(t)(u_0, u_1)\|_{L^4_{t,x}([-\delta,\delta]\times\mathbb{R}^3)} \lesssim \epsilon_0. \tag{3-7}$$

Then by the contraction mapping principle and Theorem 2.5,

$$\|u\|_{L^4_{t,x}([-\delta,\delta]\times\mathbb{R}^3)} \lesssim \|S(t)(u_0, u_1)\|_{L^4_{t,x}([-\delta,\delta]\times\mathbb{R}^3)} + \|u\|^3_{L^4_{t,x}([-\delta,\delta]\times\mathbb{R}^3)}, \tag{3-8}$$

which when $\epsilon_0 > 0$ is sufficiently small implies

$$\|u\|_{L^4_{t,x}([-\delta,\delta]\times\mathbb{R}^3)} \lesssim \epsilon_0. \tag{3-9}$$

Next observe that by Theorem 2.5 we also have

$$\||\nabla|^{1/4} u\|_{L^8_t L^{8/3}_x([-\delta,\delta]\times\mathbb{R}^3)} + \||\nabla|^{-1/4} u\|_{L^{8/3}_t L^8_x([-\delta,\delta]\times\mathbb{R}^3)} \lesssim \epsilon_0 \tag{3-10}$$

and

$$\|P_j u\|_{L^4_{t,x}([-\delta,\delta]\times\mathbb{R}^3)}+\|P_j u\|_{L^\infty_t \dot{H}^{1/2}([-\delta,\delta]\times\mathbb{R}^3)}$$
$$\lesssim \|P_j u_0\|_{\dot{H}^{1/2}(\mathbb{R}^3)}+\|P_j u_1\|_{\dot{H}^{-1/2}(\mathbb{R}^3)}$$
$$+2^{-j/2}\sum_{j_1\le j_2\le j-5}\sum_{j-3\le j_3\le j+3}\|P_{j_1}\|_{L^{8/3}_t L^8_x([-\delta,\delta]\times\mathbb{R}^3)}\|P_{j_2}u\|_{L^{8/3}_t L^8_x([-\delta,\delta]\times\mathbb{R}^3)}\|P_{j_3}u\|_{L^4_{t,x}([-\delta,\delta]\times\mathbb{R}^3)}$$
$$+2^{j/4}\sum_{j-5\le j_1\le j_2\le j_3}\|P_{j_1}u\|_{L^4_{t,x}([-\delta,\delta]\times\mathbb{R}^3)}\|P_{j_2}u\|_{L^4_{t,x}([-\delta,\delta]\times\mathbb{R}^3)}\|P_{j_3}u\|_{L^8_t L^{8/3}_x([-\delta,\delta]\times\mathbb{R}^3)}. \tag{3-11}$$

Then by (3-9) and (3-10),

$$\sum_j \|P_j u\|_{L^4_{t,x}([-\delta,\delta]\times\mathbb{R}^3)}+\|P_j u\|_{L^\infty_t \dot{H}^{1/2}([-\delta,\delta]\times\mathbb{R}^3)}$$
$$\lesssim \sum_j \|P_j u_0\|_{\dot{H}^{1/2}(\mathbb{R}^3)}+\|P_j u_1\|_{\dot{H}^{-1/2}(\mathbb{R}^3)}+\epsilon_0^2\sum_j \|P_j u\|_{L^4_{t,x}([-\delta,\delta]\times\mathbb{R}^3)}, \tag{3-12}$$

which also implies

$$\|u\|_{L^\infty_t B^{1/2}_{1,2}([-\delta,\delta]\times\mathbb{R}^3)}\lesssim \|u_0\|_{B^2_{1,1}(\mathbb{R}^3)}+\|u_1\|_{B^1_{1,1}(\mathbb{R}^3)}, \tag{3-13}$$

completing the proof. □

Next suppose $\chi(x)$ is a smooth function that is supported on $|x|\le 1$ and is equal to 1 on $|x|\le \frac{1}{2}$. By Lemma 2.2 there exists some $R(u_0,u_1,\epsilon)$ such that

$$\left\|\left(1-\chi\left(\frac{x}{R}\right)\right)u_0\right\|_{\dot{H}^{1/2}(\mathbb{R}^3)}+\left\|\left(1-\chi\left(\frac{x}{R}\right)\right)u_1\right\|_{\dot{H}^{-1/2}(\mathbb{R}^3)}\le\epsilon. \tag{3-14}$$

Remark. Notice that R depends on u_0 and u_1, not just their size. This dependence will be removed upon making a profile decomposition.

Again applying the scaling symmetry (1-3), this time with $\lambda=2R$, setting

$$u_{0,\lambda}(x)=\lambda u_0(\lambda x),\quad u_{1,\lambda}(x)=\lambda^2 u_1(\lambda x), \tag{3-15}$$

and letting u_λ denote the solution to (1-1) with initial data $(u_{0,\lambda},u_{1,\lambda})$ yields

$$\|P_{>2R}u_{0,\lambda}\|_{\dot{H}^{1/2}(\mathbb{R}^3)}+\|P_{>2R}u_{1,\lambda}\|_{\dot{H}^{-1/2}(\mathbb{R}^3)}\le\epsilon, \tag{3-16}$$

$$\|(1-\chi(2x))u_{0,\lambda}\|_{\dot{H}^{1/2}(\mathbb{R}^3)}+\|(1-\chi(2x))u_{1,\lambda}\|_{\dot{H}^{-1/2}(\mathbb{R}^3)}\le\epsilon, \tag{3-17}$$

$$\|u_\lambda\|_{L^4_{t,x}([-\frac{\delta}{2R},\frac{\delta}{2R}]\times\mathbb{R}^3)}\lesssim\epsilon_0, \tag{3-18}$$

and finally

$$\|u_\lambda\|_{L^\infty_t B^{1/2}_{1,2}([-\frac{\delta}{2R},\frac{\delta}{2R}]\times\mathbb{R}^3)}\lesssim \|u_0\|_{B^2_{1,1}(\mathbb{R}^3)}+\|u_1\|_{B^1_{1,1}(\mathbb{R}^3)}. \tag{3-19}$$

The next step is to show that this local solution has a singularity that is isolated in a suitable sense. We will once again abuse notation and use u in place of u_λ in (3-18).

Observe that the dispersive estimates imply that the linear wave equation $u_{tt}-\Delta u=0$ with initial data (u_0,u_1) lies in L^∞ when $t>0$. Indeed, by (2-39),

$$\begin{aligned}\|S(t)(u_0,u_1)\|_{L^\infty} &\lesssim \frac{1}{t}\sum_j [2^{2j}\|P_j u_0\|_{L^1(\mathbb{R}^3)} + 2^j\|P_j u_1\|_{L^1(\mathbb{R}^3)}] \\ &\lesssim \frac{1}{t}[\|u_0\|_{B^2_{1,1}(\mathbb{R}^3)} + \|u_1\|_{B^1_{1,1}(\mathbb{R}^3)}].\end{aligned} \tag{3-20}$$

Interpolating (3-20) with Bernstein's inequality, for any j,

$$\|S(t)(P_j u_0, P_j u_1)\|_{L^6(\mathbb{R}^3)} \lesssim \frac{2^{-j/6}}{t^{2/3}}[2^{2j}\|P_j u_0\|_{L^1(\mathbb{R}^3)} + 2^j\|P_j u_1\|_{L^1(\mathbb{R}^3)}], \tag{3-21}$$

while by the Sobolev embedding theorem $\dot{H}^1(\mathbb{R}^3) \hookrightarrow L^6(\mathbb{R}^3)$,

$$\|S(t)P_j(u_0,u_1)\|_{L^6(\mathbb{R}^3)} \lesssim 2^{j/2}[2^{2j}\|P_j u_0\|_{L^1(\mathbb{R}^3)} + 2^j\|P_j u_1\|_{L^1(\mathbb{R}^3)}]. \tag{3-22}$$

Thus by direct computation

$$\sup_{t>0} t^{1/2}\|S(t)(u_0,u_1)\|_{L^6(\mathbb{R}^3)} + \|S(t)(u_0,u_1)\|_{L^2_t L^6_x(\mathbb{R}\times\mathbb{R}^3)} \lesssim \|u_0\|_{B^2_{1,1}(\mathbb{R}^3)} + \|u_1\|_{B^1_{1,1}(\mathbb{R}^3)}. \tag{3-23}$$

Lemma 3.2. *If $\delta > 0$ is given by the local result in Lemma 3.1 for some $\epsilon_0 > 0$, then*

$$\sup_{-\frac{\delta}{2R}<t<\frac{\delta}{2R}} t^{1/2}\|u\|_{L^6_x(\mathbb{R}^3)} + \|u\|_{L^2_t L^6_x([-\frac{\delta}{2R},\frac{\delta}{2R}]\times\mathbb{R}^3)} \lesssim \|u_0\|_{B^2_{1,1}(\mathbb{R}^3)} + \|u_1\|_{B^1_{1,1}(\mathbb{R}^3)}. \tag{3-24}$$

Proof. By the dispersive estimates (Theorem 2.4), the Hardy–Littlewood–Sobolev inequality, and interpolation

$$\begin{aligned}\Big\|\int_0^t S(t-\tau)F(u(\tau))\,d\tau\Big\|_{L^2_t L^6_x([0,\frac{\delta}{2R}]\times\mathbb{R}^3)} &\lesssim \||\nabla|^{1/3}F(u)\|_{L^{6/5}_{t,x}([0,\frac{\delta}{2R}]\times\mathbb{R}^3)} \\ &\lesssim \||\nabla|^{1/2}u\|^{2/3}_{L^\infty_t L^2_x([0,\frac{\delta}{2R}]\times\mathbb{R}^3)}\|u\|^{4/3}_{L^4_{t,x}([0,\frac{\delta}{2R}]\times\mathbb{R}^3)}\|u\|_{L^2_t L^6_x([0,\frac{\delta}{2R}]\times\mathbb{R}^3)} \\ &\lesssim \epsilon_0^{4/3}\|u\|_{L^2_t L^6_x([0,\frac{\delta}{2R}]\times\mathbb{R}^3)}.\end{aligned} \tag{3-25}$$

Combining (3-25) with (3-23) proves

$$\|u\|_{L^2_t L^6_x([0,\frac{\delta}{2R}]\times\mathbb{R}^3)} \lesssim \|u_0\|_{B^2_{1,1}(\mathbb{R}^3)} + \|u_1\|_{B^1_{1,1}(\mathbb{R}^3)}. \tag{3-26}$$

Next let $c > 0$ be a small constant to be determined later. Again by Theorem 2.4, the Hardy–Littlewood–Sobolev inequality, and interpolation,

$$\begin{aligned}\sup_{t\in[0,\frac{\delta}{2R}]} t^{1/2}\Big\|\int_0^{(1-c)t} S(t-\tau)F(u(\tau))\,d\tau\Big\|_{L^6(\mathbb{R}^3)} &\lesssim \frac{1}{c^{1/2}}\||\nabla|^{1/3}F(u)\|_{L^{6/5}_{t,x}([0,\frac{\delta}{2R}]\times\mathbb{R}^3)} \\ &\lesssim \frac{1}{c^{1/2}}\||\nabla|^{1/2}u\|^{2/3}_{L^\infty_t L^2_x([0,\frac{\delta}{2R}]\times\mathbb{R}^3)}\|u\|^{4/3}_{L^4_{t,x}([0,\frac{\delta}{2R}]\times\mathbb{R}^3)}\|u\|_{L^2_t L^6_x([0,\frac{\delta}{2R}]\times\mathbb{R}^3)} \\ &\lesssim \frac{\epsilon_0^{4/3}}{c^{1/2}}(\|u_0\|_{B^2_{1,1}(\mathbb{R}^3)} + \|u_1\|_{B^1_{1,1}(\mathbb{R}^3)}).\end{aligned} \tag{3-27}$$

Also for any $t \in [0, \delta/(2R)]$, by Theorem 2.4,

$$t^{1/2}\left\| \int_{(1-c)t}^{t} S(t-\tau)F(u(\tau))\,d\tau \right\|_{L^6(\mathbb{R}^3)}$$
$$\lesssim \Big(\sup_{t\in[0,\frac{\delta}{2R}]} t^{1/2}\|u(t)\|_{L^6(\mathbb{R}^3)}\Big)^{5/3} \||\nabla|^{1/2}u\|^{2/3}_{L_t^\infty L_x^2([0,\frac{\delta}{2R}]\times\mathbb{R}^3)}$$
$$\times \|u\|^{2/3}_{L_t^\infty L_x^3([0,\frac{\delta}{2R}]\times\mathbb{R}^3)} \cdot \int_{(1-c)t}^{t} \frac{1}{(t-\tau)^{2/3}}\frac{1}{t^{1/3}}\,d\tau$$
$$\lesssim c^{1/3}\Big(\sup_{t\in[0,\frac{\delta}{2R}]} t^{1/2}\|u(t)\|_{L^6(\mathbb{R}^3)}\Big)^{5/3}. \tag{3-28}$$

Therefore, choosing $c > 0$ small and $\epsilon_0(c) > 0$ small,

$$\|u\|_{L_t^2 L_x^6([0,\frac{\delta}{2R}]\times\mathbb{R}^3)} + \sup_{0<t<\delta} t^{1/2}\|u(t)\|_{L^6} \lesssim \|u_0\|_{B^2_{1,1}(\mathbb{R}^3)} + \|u_1\|_{B^1_{1,1}(\mathbb{R}^3)}. \tag{3-29}$$

Then by time reversal symmetry the proof of Lemma 3.2 is complete. □

Next, we show that a local solution may be written as a sum of a term with bounded energy and a term with good dispersive properties. To simplify notation let $\delta_1 = \delta/(2R)$. By energy inequalities, Strichartz estimates (Theorem 2.5), and Lemma 3.2,

$$\left\| \int_{\frac{\delta_1}{10}}^{\delta_1} S(t-\tau)F(u(\tau))\,d\tau \right\|_{\dot H^1\times L^2(\mathbb{R}^3)} \lesssim \|u\|^3_{L_t^3 L_x^6([\frac{\delta_1}{10},\delta_1]\times\mathbb{R}^3)} \lesssim \frac{1}{\delta_1^{1/2}}. \tag{3-30}$$

Next, by the radial Sobolev embedding theorem (Theorem 2.3) and (3-13), if $\chi \in C_0^\infty(\mathbb{R}^3)$ is supported on $|x| \le 1$, $\chi(x) = 1$ on $|x| \le \frac{1}{2}$, then

$$\left\| \left(1-\chi\Big(\frac{10x}{\delta_1}\Big)\right)F(u) \right\|_{L_t^1 L_x^2([0,\frac{\delta_1}{10}]\times\mathbb{R}^3)} \lesssim \delta_1^{1/2} \left\| \left(1-\chi\Big(\frac{10x}{\delta_1}\Big)\right)u \right\|_{L^\infty_{t,x}([0,\frac{\delta_1}{10}]\times\mathbb{R}^3)} \|u\|^2_{L^4_{t,x}([0,\frac{\delta_1}{10}]\times\mathbb{R}^3)}$$
$$\lesssim \frac{1}{\delta_1^{1/2}}. \tag{3-31}$$

Now for $t > \delta_1$ let

$$v(t) = S(t)\chi\Big(\frac{10x}{\delta_1}\Big)(u_0,u_1) + \int_0^{\frac{\delta_1}{10}} S(t-\tau)\chi\Big(\frac{10x}{\delta_1}\Big)F(u(\tau))\,d\tau. \tag{3-32}$$

Combining Lemma 2.2 with

$$\|[P_j,\chi]F(u)\|_{L_t^1 \dot H^{-1/2}([-\frac{\delta_1}{10},\frac{\delta_1}{10}]\times\mathbb{R}^3)} \lesssim 2^{-j}\delta_1^{-1}\|F(u)\|_{L_t^1 L_x^{3/2}([-\frac{\delta_1}{10},\frac{\delta_1}{10}]\times\mathbb{R}^3)}, \tag{3-33}$$

Lemma 3.2,

$$\left\| P_{\le 0}\chi\Big(\frac{10x}{\delta_1}\Big)F(u) \right\|_{\dot H^{-1/2}(\mathbb{R}^3)} \lesssim \|u\|^3_{L_x^3(\mathbb{R}^3)}, \tag{3-34}$$

(3-10)–(3-13), the sharp Huygens principle, which implies v is supported on $\{(x,t) : ||x|-t| \leq \frac{1}{2}\delta_1\}$, and the radial Sobolev embedding theorem (Theorem 2.3), we have

$$\|v(t)\|_{L^\infty(\mathbb{R}^3)} \lesssim \frac{1}{t}[\|u_0\|_{B^2_{1,1}(\mathbb{R}^3)} + \|u_1\|_{B^1_{1,1}(\mathbb{R}^3)}]. \tag{3-35}$$

This implies good properties of $S(t-\delta_1)(v(\delta_1), v_t(\delta_1))$.

Lemma 3.3. *Let $w(\delta_1) + v(\delta_1) = u(\delta_1)$ and let*

$$\begin{aligned} w(\delta_1) = S(\delta_1)\left(1-\chi\left(\frac{10x}{\delta_1}\right)\right)(u_0,u_1) \\ + \int_0^{\frac{\delta_1}{10}} S(\delta_1-\tau)\left(1-\chi\left(\frac{10x}{\delta_1}\right)\right)F(u(\tau))\,d\tau + \int_{\frac{\delta_1}{10}}^{\delta_1} S(\delta_1-\tau)F(u(\tau))\,d\tau. \end{aligned} \tag{3-36}$$

Then

$$\|w(\delta_1)\|_{\dot{H}^1\times L^2(\mathbb{R}^3)} \lesssim \delta_1^{-1/2}. \tag{3-37}$$

Proof. By (3-31) and (3-32) it only remains to compute

$$\left\|\left(1-\chi\left(\frac{10x}{\delta_1}\right)\right)u_0\right\|_{\dot{H}^1(\mathbb{R}^3)} + \left\|\left(1-\chi\left(\frac{10x}{\delta_1}\right)\right)u_1\right\|_{L^2(\mathbb{R}^3)}. \tag{3-38}$$

First,

$$|u_1(0,r)| \lesssim \int_r^\infty |\partial_r u_1(0,s)|\,ds \lesssim \frac{1}{r^2}, \tag{3-39}$$

so

$$\int_{\frac{\delta_1}{10}}^\infty |u_1(r,0)|^2 r^2\,dr \lesssim \int_{\frac{\delta_1}{10}}^\infty |u_1(r,0)|\,dr \lesssim \frac{1}{\delta_1}. \tag{3-40}$$

Next, for any j, Bernstein's inequality implies $\|P_j u_0\|_{L^\infty} \lesssim 2^{3j}\|P_j u_0\|_{L^1}$, so

$$\int_0^{2^{-j}} \frac{r^2}{r}|\partial_r(P_j u_0)|\,dr \lesssim 2^{2j}\|P_j u_0\|_{L^1(\mathbb{R}^3)}, \tag{3-41}$$

while by Bernstein's inequality

$$\int_{2^{-j}}^\infty \frac{r^2}{r}|\partial_r(P_j u_0)|\,dr \lesssim 2^{2j}\|P_j u_0\|_{L^1(\mathbb{R}^3)}. \tag{3-42}$$

Thus $\|(1/r)\partial_r u_0\|_{L^1(\mathbb{R}^3)} \lesssim \|u_0\|_{B^2_{1,1}(\mathbb{R}^3)}$, and since u_0 is radially symmetric $\Delta u_0 = (\partial_{rr} + (2/r)\partial_r)u_0$, so $\|\partial_{rr}u_0\|_{L^1(\mathbb{R}^3)} \lesssim \|u_0\|_{B^2_{1,1}(\mathbb{R}^3)}$. By the fundamental theorem of calculus,

$$|u_r(r)| \leq \int_r^\infty |u_{rr}(s)|\,ds \lesssim \frac{1}{r^2}. \tag{3-43}$$

Therefore,

$$\int_{\frac{\delta_1}{10}}^\infty |u_r(r)|^2 r^2\,dr \lesssim \int_{\frac{\delta_1}{10}}^\infty |u_r(r)|\,dr \lesssim \frac{1}{\delta_1}, \tag{3-44}$$

and
$$\int_0^\infty (u_r(r))^2 r^3\,dr \le \int_0^\infty \left(\int_r^\infty |u_{rr}(s)|\,ds\right) r\,dr \lesssim \int_0^\infty |u_{rr}(s)| s^2\,ds < \infty, \tag{3-45}$$
completing the proof. □

4. Proof of global well-posedness

In this section we extend local well-posedness to global well-posedness, proving:

Theorem 4.1. *Equation (1-1) is globally well-posed, and for any compact interval $J \subset \mathbb{R}$,*
$$\|u\|_{L^4_{t,x}(J\times\mathbb{R}^3)} < \infty. \tag{4-1}$$

Proof. By time reversal symmetry, to prove this it suffices to show that the local well-posedness result of Lemma 3.1 can be extended to all times $t > \delta_1$. Throughout the proof the implicit constant depends on δ_1 and $\|u_0\|_{B^2_{1,1}} + \|u_1\|_{B^1_{1,1}}$.

For $t > \delta_1$ let
$$u(t) = w(t) + v(t), \tag{4-2}$$
where $v(t)$ is given by (3-32) and w solves
$$w_{tt} - \Delta w = -u^3. \tag{4-3}$$

By Strichartz estimates, (2-11), and (3-9),
$$\|v\|_{L^4_{t,x}([\delta_1,\infty)\times\mathbb{R}^3)} \lesssim \|u_0\|_{B^2_{1,1}} + \|u_1\|_{B^1_{1,1}}. \tag{4-4}$$

Thus to prove (4-1) it suffices to prove that $w \in L^4_x$ for all $t \in [\delta_1, \infty)$.

Copying (1-2) let $E(w(t))$ be the energy of w,
$$E(w(t)) = \frac{1}{2}\int |\nabla w(t,x)|^2\,dx + \frac{1}{2}\int (w_t(t,x))^2\,dx + \frac{1}{4}\int (w(t,x))^4\,dx. \tag{4-5}$$
By (3-19), (3-37), and the Sobolev embedding theorem, $w \in L^3 \cap L^6$, so
$$E(w(\delta_1)) \lesssim 1. \tag{4-6}$$
Next,
$$\begin{aligned}\frac{d}{dt}E(w(t)) &= \int ((w(t,x))^3 - (u(t,x))^3) w_t(t,x)\,dx \\ &= -\int w_t(t,x)[(v(t,x))^3 + 3v(t,x)^2 w(t,x) + 3v(t,x)w(t,x)^2]\,dx.\end{aligned} \tag{4-7}$$
It suffices to show that (3-35) and (4-4) give good bounds on the growth of $E(w(t))$. Indeed,
$$\int w_t(t,x)(w(t,x))^2 v(t,x)\,dx \lesssim \|v(t)\|_{L^\infty(\mathbb{R}^3)} \|w(t)\|^2_{L^4(\mathbb{R}^3)} \|w_t(t)\|_{L^2(\mathbb{R}^3)} \lesssim \frac{1}{t} E(w(t)), \tag{4-8}$$

and

$$\int w_t(t,x)(v(t,x))^3\,dx \lesssim \|w_t(t)\|_{L^2(\mathbb{R}^3)}\|v(t)\|_{L^\infty(\mathbb{R}^3)}\|v(t)\|^2_{L^4(\mathbb{R}^3)} \lesssim \frac{1}{t}E(w(t))^{1/2}\|v(t)\|^2_{L^4(\mathbb{R}^3)}. \tag{4-9}$$

Finally

$$\int w_t(t,x)v(t,x)^2 w(t,x)\,dx \lesssim \|w_t(t)\|_{L^2(\mathbb{R}^3)}\|w(t)\|_{L^4(\mathbb{R}^3)}\|v(t)\|_{L^4(\mathbb{R}^3)}\|v(t)\|_{L^\infty(\mathbb{R}^3)} \lesssim \frac{1}{t}E(w(t))^{3/4}\|v(t)\|_{L^4(\mathbb{R}^3)}, \tag{4-10}$$

so by interpolation

$$\frac{d}{dt}E(w(t)) \lesssim \frac{1}{t}E(w(t)) + \frac{1}{t}\|v(t)\|^4_{L^4(\mathbb{R}^3)}. \tag{4-11}$$

Then by Gronwall's inequality and time reversal symmetry there exist constants $C_1(\|u_0\|_{B^2_{1,1}}, \|u_1\|_{B^1_{1,1}}, \delta_1)$ and $C_2(\|u_0\|_{B^2_{1,1}}, \|u_1\|_{B^1_{1,1}}, \delta_1)$ such that

$$E(w(t)) \lesssim C_1(1+|t|)^{C_2}. \tag{4-12}$$

This combined with (4-2) implies that u is global. □

5. Hyperbolic coordinates

Having shown that the solution to (1-1) is globally well-posed, the next step is to show that the solution scatters. It is possible to prove this by utilizing hyperbolic coordinates and the fact that by Theorem 4.1, $u(t,x)$ is well-defined for all $(t,x) \in \mathbb{R}^{1+3}$.

Theorem 5.1. *The global solution given in Theorem 4.1 scatters both forward and backward in time.*

Proof. By time reversal symmetry and (1-14), it suffices to show that

$$\int_0^\infty \int_0^\infty u(t,r)^4 r^2\,dr\,dt < \infty. \tag{5-1}$$

To begin setting up hyperbolic coordinates, translate $t_0 = 0$ in time to $t_0 = 1-\delta_1$. Let $\tilde{u}$ refer to the time-translated solution in Theorem 4.1,

$$\tilde{u}(t,r) = u(t-(1-\delta_1), r). \tag{5-2}$$

Once again, we make an abuse of notation and let u refer to $\tilde{u}$. Inequality (4-1) implies that after time translation

$$\int_0^1 \int_0^\infty u(t,r)^4 r^2\,dr\,dt < \infty. \tag{5-3}$$

Next, by small-data arguments, see for example [Lindblad and Sogge 1995], the solution to (1-1) with initial data given by (3-14), has finite $L^4_{t,x}$ norm. Finite propagation speed and (3-14) imply

$$\int_1^\infty \int_{r>\frac{1}{2}+t} u(t,r)^4 r^2\,dr\,dt \lesssim \epsilon_0. \tag{5-4}$$

Therefore, it only remains to prove

$$\int_1^\infty \int_{r \leq \frac{1}{2}+t} u(t,r)^4 r^2 \, dr \, dt < \infty. \tag{5-5}$$

Such x and t fall within the domain, which may be described by hyperbolic coordinates. Let

$$\tilde{w}(\tau,s) = \frac{e^\tau \sinh s}{s} w(e^\tau \cosh s, e^\tau \sinh s), \tag{5-6}$$

$$\tilde{v}(\tau,s) = \frac{e^\tau \sinh s}{s} v(e^\tau \cosh s, e^\tau \sinh s), \tag{5-7}$$

$$\tilde{u}(\tau,s) = \frac{e^\tau \sinh s}{s} u(e^\tau \cosh s, e^\tau \sinh s). \tag{5-8}$$

Now by a change of variables

$$\begin{aligned}\int_0^\infty \int_0^\infty \tilde{u}(\tau,s)^4 s^2 \left(\frac{s}{\sinh s}\right)^2 ds \, d\tau &= \int_0^\infty \int_0^\infty \frac{e^{4\tau}(\sinh s)^4}{s^4} \left(\frac{s^4}{(\sinh s)^2}\right) u(e^\tau \cosh s, e^\tau \sinh s)^4 \, ds \, d\tau \\ &= \iint_{t^2-r^2 \geq 1} u(t,r)^4 r^2 \, dr \, dt. \end{aligned} \tag{5-9}$$

The analogous estimate also holds for v and w. Since (4-1) implies

$$\int_1^2 \int u(t,r)^4 r^2 \, dr \, dt < \infty, \tag{5-10}$$

proving (5-5) is equivalent to proving

$$\int_0^\infty \int_0^\infty \tilde{u}(\tau,s)^4 \left(\frac{s}{\sinh s}\right)^2 s^2 \, ds \, d\tau < \infty. \tag{5-11}$$

Also, since v is a solution to the linear wave equation with data in $\dot{H}^{1/2} \times \dot{H}^{-1/2}$, it suffices to prove

$$\int_0^\infty \int_0^\infty \tilde{w}(\tau,s)^4 \left(\frac{s}{\sinh s}\right)^2 s^2 \, ds \, d\tau < \infty. \tag{5-12}$$

Direct computation and (4-3) imply that $\tilde{w}$ solves

$$\partial_{\tau\tau}\tilde{w} - \partial_{ss}\tilde{w} - \frac{2}{s}\partial_s \tilde{w} = -\left(\frac{s}{\sinh s}\right)^2 \tilde{u}^3. \tag{5-13}$$

Moreover, $\tilde{w}$ has bounded hyperbolic energy

$$\frac{1}{2}\int_0^\infty \tilde{w}_s(\tau,s)^2 s^2 \, ds + \frac{1}{2}\int_0^\infty \tilde{w}_\tau(\tau,s)^2 s^2 \, ds + \frac{1}{4}\int_0^\infty \left(\frac{s}{\sinh s}\right)^2 \tilde{w}(\tau,s)^4 s^2 \, ds. \tag{5-14}$$

Indeed:

Lemma 5.2. *There exists some $0 < \tau < \delta_1$ such that*

$$E(\tilde{w}(\tau)) < \infty. \tag{5-15}$$

Proof. To prove (5-15) it suffices to show

$$\frac{1}{\delta_1}\int_0^{\delta_1} E(\tilde{w}(\tau))\,d\tau < \infty. \tag{5-16}$$

By the hyperbolic Pythagorean theorem,

$$e^{\tau}(\cosh(s)-\sinh(s)) = e^{\tau}\frac{\cosh^2(s)-\sinh^2(s)}{\cosh(s)+\sinh(s)} = \frac{e^{\tau}}{\cosh(s)+\sinh(s)}. \tag{5-17}$$

Therefore, for any fixed $\tau > 0$,

$$\lim_{s\nearrow\infty} e^{\tau}(\cosh(s)-\sinh(s)) = 0. \tag{5-18}$$

Combining (3-32) with the fact that $t_0 = 0$ was translated to $t_0 = 1-\delta_1$, for any $t \geq 1$,

$$v(t) = S(t-1+\delta_1)\chi\left(\frac{10x}{\delta_1}\right)(u_0,u_1) + \int_0^{\frac{\delta_1}{10}} S(t-1+\delta_1-\tau)\chi\left(\frac{10x}{\delta_1}\right)F(u(\tau))\,d\tau. \tag{5-19}$$

Therefore, by finite propagation speed there exists some s_0 such that, for any $0 < \tau < \delta_1$,

$$\int_{s_0}^{\infty} s^2\tilde{w}_s(\tau,s)^2\,ds + \int_{s_0}^{\infty} s^2\tilde{w}_\tau(\tau,s)\,ds = \int_{s_0}^{\infty} s^2\tilde{u}_s(\tau,s)^2\,ds + \int_{s_0}^{\infty} s^2\tilde{u}_\tau(\tau,s)\,ds. \tag{5-20}$$

Now, if u is a radial solution to (1-1), then by (5-8),

$$\begin{aligned} s\tilde{u}(\tau,s) &= e^{\tau}(\sinh s)u(e^{\tau}\cosh s, e^{\tau}\sinh s) \\ &= \frac{1}{2}(e^{\tau+s}-(1-\delta_1))u_0(e^{\tau+s}-(1-\delta_1)) + \frac{1}{2}(1-\delta_1-e^{\tau-s})u_0(1-\delta_1-e^{\tau-s}) \\ &\quad + \frac{1}{2}\int_{e^{\tau-s}+(1-\delta_1)}^{e^{\tau+s}-(1-\delta_1)} u_1(r)r\,dr + \frac{1}{2}\int_{1-\delta_1}^{e^{\tau}\cosh s}\int_{-e^{\tau-s}+t}^{e^{\tau+s}-t} ru^3(t,r)\,dr\,dt. \end{aligned} \tag{5-21}$$

Now by direct computation,

$$\begin{aligned} \partial_{\tau+s}(s\tilde{u}(\tau,s))|_{\tau=0} &= \frac{e^s}{2}u_0(e^s-(1-\delta_1)) + \frac{e^s}{2}(e^s-(1-\delta_1))u_0'(e^s-(1-\delta_1)) \\ &\quad + \frac{e^s}{2}(e^s-(1-\delta_1))u_1(e^s-(1-\delta_1)) + \frac{e^s}{2}\int_{1-\delta_1}^{\cosh s}(e^s-t)u^3(t,e^s-t)\,dt, \end{aligned} \tag{5-22}$$

and

$$\begin{aligned} &\partial_{\tau-s}(s\tilde{u}(\tau,s))|_{\tau=0} \\ &\quad = \frac{e^{-s}}{2}u_0((1-\delta_1)-e^{-s}) + \frac{e^{-s}}{2}((1-\delta_1)-e^{-s})u_0'((1-\delta_1)-e^{-s}) \\ &\qquad + \frac{e^{-s}}{2}((1-\delta_1)-e^{-s})u_1((1-\delta_1)-e^{-s}) + \frac{e^{-s}}{2}\int_{1-\delta_1}^{e^{\tau}\cosh s}(t-e^{-s})u^3(t,t-e^{-s})\,dt. \end{aligned} \tag{5-23}$$

First, making a change of variables and using (3-39)–(3-45),

$$\int_{s_0}^{\infty} e^{2s} u_0(e^s-(1-\delta_1))^2\,ds \lesssim \int_0^{\infty} u_0(r)^2 r\,dr < \infty, \tag{5-24}$$

$$\int_{s_0}^{\infty} e^{2s}(e^s-(1-\delta_1))^2 u_0'(e^s-(1-\delta_1))^2\,ds \lesssim \int_0^{\infty} (\partial_r u_0(r))^2 r^3\,dr < \infty, \tag{5-25}$$

$$\int_{s_0}^{\infty} e^{2s}(e^s-(1-\delta_1))^2 u_1(e^s-(1-\delta_1))^2\,ds \lesssim \int_0^{\infty} r^3 u_1(r)^2\,dr < \infty. \tag{5-26}$$

Also, by (3-39)–(3-45), the fact that $e^{\tau}(\cosh s-\sinh s) \le \frac{1}{4}$ when $s \ge s_0$ and $0<\tau<\delta_1$, $|ru(t,r)| \lesssim 1$ for $t-r \le \frac{1}{2}$, (5-3), and (5-4), we have

$$\begin{aligned}\int_{s_0}^{\infty} e^{2s}\Big(\int_{1-\delta_1}^{\cosh s}(e^s-t)u^3(t,e^s-t)\,dt\Big)^2\,ds &\lesssim \int_{s_0}^{\infty}\int_{1-\delta_1}^{\cosh s} e^{3s}(e^s-t)^2 u^6(t,e^s-t)\,dt\,ds \\ &\lesssim \int_{s_0}^{\infty}\int_{1-\delta_1}^{\cosh(s)} e^{3s} u^4(t,e^s-t)\,dt\,ds < \infty. \end{aligned}\tag{5-27}$$

Additionally,

$$\int_{s_0}^{\infty} e^{-2s} u_0((1-\delta_1)-e^{-s})^2\,ds \lesssim \int_{s_0}^{\infty} e^{-2s}\,ds < \infty, \tag{5-28}$$

$$\int_{s_0}^{\infty} e^{-2s}((1-\delta_1)-e^{-s})^2(u_0'((1-\delta_1)-e^{-s}))^2\,ds \lesssim \int_{s_0}^{\infty} e^{-2s}\,ds < \infty, \tag{5-29}$$

$$\int_{s_0}^{\infty} e^{-2s}((1-\delta_1)-e^{-s})^2 u_1((1-\delta_1)-e^{-s})^2\,ds \lesssim \int_{s_0}^{\infty} e^{-2s}\,ds < \infty. \tag{5-30}$$

Also by the fact that $|u(t,r)|r$ is uniformly bounded for $t-r \le \frac{1}{2}$,

$$\int_{s_0}^{\infty} e^{-2s}\Big(\int_{1-\delta_1}^{e^{\tau}\cosh s}(t-e^{-s})u^3(t,t-e^{-s})\,dt\Big)^2\,ds \lesssim \int_{s_0}^{\infty} e^{-2s}\,ds < \infty. \tag{5-31}$$

In fact the above computations could be made for any $0<\tau<\delta_1$ with some uniform s_0. Now then, integrating by parts,

$$\int_{s_0}^{\infty} \partial_s(s\tilde{w}(\tau,s))^2\,ds = \int_{s_0}^{\infty} s^2\tilde{w}_s(\tau,s)^2\,ds - s_0\tilde{w}(\tau,s_0)^2. \tag{5-32}$$

Remark. It is straightforward to verify that by (5-18),

$$\lim_{s\nearrow\infty} s|\tilde{w}(\tau,s)|^2 = 0, \tag{5-33}$$

so (5-32) is well-defined. Therefore,

$$\sup_{0<\tau<\delta_1}\int_{s_0}^{\infty} \tilde{w}_{\tau}(s,\tau)^2 s^2\,ds + \int_{s_0}^{\infty} \tilde{w}_{\tau}(s,\tau)^2 s^2\,ds < \infty. \tag{5-34}$$

So by (5-16) it suffices to show that

$$\int_0^{\delta_1}\int_0^{s_0} s^2\tilde{w}_s(\tau,s)^2\,ds\,d\tau + \int_0^{\delta_1}\int_0^{s_0} s^2\tilde{w}_\tau(\tau,s)\,ds\,d\tau < \infty. \tag{5-35}$$

This fact is an immediate consequence of (5-6), Theorem 4.1, and the fact that $e^\tau \sinh s$ and $e^\tau \cosh s$ are uniformly bounded when $s \le s_0$ and $\tau \le \delta_1$. Thus, for some $0 < \tau_0 < \delta_1$,

$$\int_0^{\infty} s^2\tilde{w}_s(\tau_0,s)^2\,ds + \int_0^{\infty} s^2\tilde{w}_\tau(\tau_0,s)\,ds < \infty. \tag{5-36}$$

An application of the Sobolev embedding theorem $\dot{H}^1 \hookrightarrow L^6$ combined with the fact that $s/\sinh s \in L^1 \cap L^\infty(s^2\,ds)$ completes the proof of Lemma 5.2. □

Next we compute

$$\frac{d}{d\tau}E(\tilde{w}(\tau)) = \int \tilde{w}_\tau[\tilde{u}^3 - \tilde{w}^3]\left(\frac{s}{\sinh s}\right)^2 s^2\,ds. \tag{5-37}$$

Because $v(t,r)$ is supported on $t-r = 1+O(\delta_1)$, (5-17) implies $1/\sinh s \lesssim e^{-\tau}$ on the support of $\tilde{v}(\tau,s)$. Therefore, the radial Sobolev embedding theorem implies $\|s\tilde{v}(\tau,s)\|_{L^\infty} < \infty$, so

$$\|\tilde{w}_\tau(\tau)\|_{L^2}\left\|\tilde{v}(\tau,s)^2\left(\frac{s}{\sinh s}\right)\right\|_{L^2}\left\|\tilde{v}(\tau,s)\left(\frac{s}{\sinh s}\right)\right\|_{L^\infty} \lesssim e^{-\tau}E(\tilde{w}(\tau))^{1/2}\left\|\tilde{v}(\tau,s)^2\left(\frac{s}{\sinh s}\right)\right\|_{L^2}. \tag{5-38}$$

Meanwhile,

$$\|\tilde{w}_\tau(\tau)\|_{L^2}\left\|\tilde{v}(\tau,s)\left(\frac{s}{\sinh s}\right)\right\|_{L^\infty}\left\|\tilde{w}(\tau,s)^2\left(\frac{s}{\sinh s}\right)\right\|_{L^2} \lesssim e^{-\tau}E(\tilde{w}(\tau)), \tag{5-39}$$

and

$$\begin{aligned}\|\tilde{w}_\tau(\tau)\|_{L^2}\left\|\tilde{v}(\tau,s)\left(\frac{s}{\sinh s}\right)\right\|_{L^\infty}\left\|\tilde{w}(\tau,s)\left(\frac{s}{\sinh s}\right)^{1/2}\right\|_{L^4}\left\|\tilde{v}(\tau,s)\left(\frac{s}{\sinh s}\right)^{1/2}\right\|_{L^4}\\ \lesssim e^{-\tau}E(\tilde{w}(\tau))^{3/4}\left\|\tilde{v}(\tau,s)\left(\frac{s}{\sinh s}\right)^{1/2}\right\|_{L^4}.\end{aligned} \tag{5-40}$$

Now by this, $\|v\|_{L^4_{t,x}} < \infty$, (5-9), and Gronwall's inequality, we know $E(\tilde{w}(\tau))$ is uniformly bounded on $\mathbb{R}$.

Next we prove the Morawetz estimate.

Theorem 5.3.

$$\iint \tilde{w}(s,\tau)^4\left(\frac{s}{\sinh s}\right)^2 s^2\,ds\,d\tau < \infty. \tag{5-41}$$

Proof. We have

$$M(\tau) = \int \tilde{w}_\tau\left(\frac{x}{|x|}\cdot\nabla\tilde{w}\right)dx. \tag{5-42}$$

Then

$$\frac{d}{d\tau}M(\tau) = \int\left(\frac{\cosh s}{\sinh s}\right)\left(\frac{s}{\sinh s}\right)^2\tilde{w}^4 s^2\,ds + \iint \frac{x}{|x|}\cdot(\nabla\tilde{w})(\tilde{u}^3-\tilde{w}^3)s^2\,ds\,d\tau. \tag{5-43}$$

As in the bounded energy computations,

$$\|\nabla\tilde{w}\|_{L^2}\left\|\tilde{v}^2\left(\frac{s}{\sinh s}\right)\right\|_{L^2}\left\|\tilde{v}\left(\frac{s}{\sinh s}\right)\right\|_{L^\infty}\lesssim e^{-\tau}E(w(\tau))^{1/2}\left\|\tilde{v}^2\left(\frac{s}{\sinh s}\right)\right\|_{L^2}, \tag{5-44}$$

$$\|\nabla\tilde{w}\|_{L^2}\left\|\tilde{v}\left(\frac{s}{\sinh s}\right)\right\|_{L^\infty}\left\|\tilde{w}^2\left(\frac{s}{\sinh s}\right)\right\|_{L^2}\lesssim e^{-\tau}E(\tilde{w}), \tag{5-45}$$

and

$$\|\nabla\tilde{w}\|_{L^\infty}\left\|\tilde{v}\left(\frac{s}{\sinh s}\right)\right\|_{L^\infty}\left\|\tilde{w}\left(\frac{s}{\sinh s}\right)^{1/2}\right\|_{L^4}\left\|\tilde{v}\left(\frac{s}{\sinh s}\right)^{1/2}\right\|_{L^4}$$
$$\lesssim e^{-\tau}E(\tilde{w})^{3/4}\left\|\tilde{v}\left(\frac{s}{\sinh s}\right)^{1/2}\right\|_{L^4}. \tag{5-46}$$

Therefore, by the fundamental theorem of calculus, the fact that the energy is uniformly bounded, and (5-43),

$$\int_0^\infty\int_0^\infty \tilde{w}(s,\tau)^4\left(\frac{s}{\sinh s}\right)^2 s^2\,ds\,d\tau<\infty, \tag{5-47}$$

completing the proof of Theorem 5.3. □

Since $(\cosh s/\sinh s)\geq 1$, Theorem 5.3 directly implies (5-12), which completes the proof of Theorem 5.1. □

Remark. Notice that Theorem 5.1 implies

$$\int_0^\infty\int_0^\infty u(t,r)^4 r^2\,dr\,dt\leq C(\|u_0\|_{B^2_{1,1}},\|u_1\|_{B^1_{1,1}},\delta_1)<\infty. \tag{5-48}$$

Thus Theorem 5.1 is not equivalent to Theorem 1.4. This $\delta_1>0$ depends on the support of u_0 and u_1 in space (3-14) and in frequency (3-2). To remove this requirement, it is necessary make a profile decomposition, the subject of the final section of this paper.

6. Profile decomposition

Now, to prove Theorem 1.4 from Theorem 5.1, it only remains to show that if (u_0^n,u_1^n) is a bounded sequence in $B^2_{1,1}\times B^1_{1,1}$ and $u^n(t)$ is the corresponding solution to (1-1) with initial data (u_0^n,u_1^n), then

$$\|u^n(t)\|_{L^4_{t,x}(\mathbb{R}\times\mathbb{R}^3)} \tag{6-1}$$

is uniformly bounded. This may be accomplished by proving that (u_0^n,u_1^n) must converge to a maximizer in $B^2_{1,1}\times B^1_{1,1}$. An argument of this type was used in [Gérard 1998] to prove the existence of a maximizer of the Sobolev embedding, and for many other maximizer problems. See [Bahouri and Gérard 1999] for an early application of the profile decomposition to the nonlinear wave equation.

The intuition behind this argument may be summarized as follows. The uncertainty principle implies that when most of $B^2_{1,1}\times B^1_{1,1}$ lies below frequency 1, $R\gtrsim 1$, where R is defined in (3-14). On the other hand, if $R \gg 1$ and (u_0, u_1) is a radial function supported on the annulus $R \le r \le 2R$, then the $\dot{H}^{1/2}$ norm on balls of radius cR for some $c > 0$ small would actually be fairly small.

Combining the small-data arguments of [Lindblad and Sogge 1995] with finite propagation speed, one can show

$$\|u\|_{L^4_{t,x}([-cR,cR]\times\mathbb{R}^3)} \ll 1. \tag{6-2}$$

This provides substantial improvement over the frequency scale arguments used in the proof of Lemma 3.1.

Theorem 6.1 (profile decomposition). *Suppose that there is a uniformly bounded, radially symmetric sequence*

$$\|u_0^n\|_{\dot{H}^{1/2}(\mathbb{R}^3)} + \|u_1^n\|_{\dot{H}^{-1/2}(\mathbb{R}^3)} \le C_0 < \infty. \tag{6-3}$$

Then there exists a subsequence, also denoted by $(u_0^n, u_1^n) \subset \dot{H}^{1/2} \times \dot{H}^{-1/2}$, such that for any $N < \infty$

$$S(t)(u_0^n, u_1^n) = \sum_{j=1}^{N} \Gamma_n^j S(t)(\phi_0^j, \phi_1^j) + S(t)(R_{0,n}^N, R_{1,n}^N), \tag{6-4}$$

with

$$\lim_{N\to\infty} \limsup_{n\to\infty} \|S(t)(R_{0,n}^N, R_{1,n}^N)\|_{L^4_{t,x}(\mathbb{R}\times\mathbb{R}^3)} = 0. \tag{6-5}$$

$\Gamma_n^j = (\lambda_n^j, t_n^j)$ belongs to the group $(0,\infty)\times\mathbb{R}$, which acts by

$$\Gamma_n^j F(t,x) = \lambda_n^j F(\lambda_n^j(t - t_n^j), \lambda_n^j x). \tag{6-6}$$

The Γ_n^j are pairwise orthogonal; that is, for every $j \ne k$,

$$\lim_{n\to\infty} \frac{\lambda_n^j}{\lambda_n^k} + \frac{\lambda_n^k}{\lambda_n^j} + (\lambda_n^j)^{1/2}(\lambda_n^k)^{1/2}|t_n^j - t_n^k| = \infty. \tag{6-7}$$

Furthermore, for every $N \ge 1$,

$$\|(u_{0,n}, u_{1,n})\|^2_{\dot{H}^{1/2}\times\dot{H}^{-1/2}} = \sum_{j=1}^{N} \|(\phi_0^j, \phi_0^k)\|^2_{\dot{H}^{1/2}\times\dot{H}^{-1/2}} + \|(R_{0,n}^N, R_{1,n}^N)\|^2_{\dot{H}^{1/2}\times\dot{H}^{-1/2}} + o_n(1). \tag{6-8}$$

Proof. Ramos [2012] proved this result for data which need not be radially symmetric. Such a result is substantially more difficult since it requires accounting for Lorentz transformations and translation in space.

Now take a uniformly bounded sequence

$$\|u_0^n\|_{B^2_{1,1}(\mathbb{R}^3)} + \|u_1^n\|_{B^1_{1,1}(\mathbb{R}^3)} \le C_0 < \infty \tag{6-9}$$

such that if $u^n(t)$ is the solution of (1-1) with initial data (u_0^n, u_1^n), then

$$\|u^n(t)\|_{L^4_{t,x}(\mathbb{R}\times\mathbb{R}^3)} \nearrow \sup\{\|u\|_{L^4_{t,x}(\mathbb{R}\times\mathbb{R}^3)} : \|u_0\|_{B^2_{1,1}(\mathbb{R}^3)} + \|u_1\|_{B^1_{1,1}(\mathbb{R}^3)} \le C_0\}. \tag{6-10}$$

By the Sobolev embedding theorem,

$$\|u_0^n\|_{\dot H^{1/2}(\mathbb{R}^3)} + \|u_1^n\|_{\dot H^{-1/2}(\mathbb{R}^3)} \lesssim \|u_0^n\|_{B^2_{1,1}(\mathbb{R}^3)} + \|u_1^n\|_{B^1_{1,1}(\mathbb{R}^3)} \le C_0 < \infty, \tag{6-11}$$

which by Theorem 6.1 gives a profile decomposition

$$S(t)(u_0^n, u_1^n) = \sum_{j=1}^{N} S(t - t_n^j)\big(\lambda_n^j \phi_0^j(\lambda_n^j x), (\lambda_n^j)^2 \phi_1^j(\lambda_n^j x)\big) + S(t)(R_{0,n}^N, R_{1,n}^N). \tag{6-12}$$

In the course of proving Theorem 6.1, Ramos [2012] proved

$$S(\lambda_n^j t_n^j)\left(\frac{1}{\lambda_n^j} u_0^n\left(\frac{x}{\lambda_n^j}\right), \frac{1}{(\lambda_n^j)^2} u_1^n\left(\frac{x}{\lambda_n^j}\right)\right) \rightharpoonup \phi_0^j(x) \tag{6-13}$$

weakly in $\dot H^{1/2}(\mathbb{R}^3)$, and

$$\partial_t S(t + \lambda_n^j t_n^j)\left(\frac{1}{\lambda_n^j} u_0^n\left(\frac{x}{\lambda_n^j}\right), \frac{1}{(\lambda_n^j)^2} u_1^n\left(\frac{x}{\lambda_n^j}\right)\right)\bigg|_{t=0} \rightharpoonup \phi_1^j(x) \tag{6-14}$$

weakly in $\dot H^{-1/2}(\mathbb{R}^3)$. The fact that (u_0^n, u_1^n) is uniformly bounded in $B^2_{1,1} \times B^1_{1,1}$ prevents t_n^j from going off to $-\infty$ or $+\infty$.

Lemma 6.2. *For each j, t_n^j is uniformly bounded.*

Proof. The proof of this fact utilizes dispersive estimates and Lemma 4.1 from [Ramos 2012]:

Lemma 6.3.

$$(u_0^n, u_1^n) \rightharpoonup (\phi_0, \phi_1) \tag{6-15}$$

weakly in $\dot H^{1/2}(\mathbb{R}^3) \times \dot H^{-1/2}(\mathbb{R}^3)$ is equivalent to

$$S(t)(u_0^n, u_1^n) \rightharpoonup S(t)(\phi_0, \phi_1) \tag{6-16}$$

weakly in $L^4_{t,x}(\mathbb{R} \times \mathbb{R}^3)$.

Now observe that for any j,

$$S(t + \lambda_n^j t_n^j)\left(\frac{1}{\lambda_n^j} u_0^n(x), \frac{1}{(\lambda_n^j)^2} u_1^n(x)\right) \rightharpoonup S(t)(\phi_0^j, \phi_1^j) \tag{6-17}$$

weakly in $L^4_{t,x}$.

Now, by the dispersive estimates (Theorem 2.4), for any $l \in \mathbb{Z}$,

$$\begin{aligned}\left\| P_l S(t + \lambda_n^j t_n^j)\left(\frac{1}{\lambda_n^j} u_0^n(x), \frac{1}{(\lambda_n^j)^2} u_1^n(x)\right)\right\|_{L^\infty(\mathbb{R}^3)} \\ \lesssim \frac{1}{|t + \lambda_n^j t_n^j|}\left[2^{2l}\left\| P_l\left(\frac{1}{\lambda_n^j} u_0^n\left(\frac{x}{\lambda_n^j}\right)\right)\right\|_{L^1(\mathbb{R}^3)} + 2^l \left\| P_l\left(\frac{1}{(\lambda_n^j)^2} u_1^n\left(\frac{x}{\lambda_n^j}\right)\right)\right\|_{L^1(\mathbb{R}^3)}\right].\end{aligned} \tag{6-18}$$

Meanwhile, by Bernstein's inequality and the Sobolev embedding theorem

$$\Big\| P_l S(t+\lambda_n^j t_n^j)\Big(\frac{1}{\lambda_n^j}u_0^n(x), \frac{1}{(\lambda_n^j)^2}u_1^n(x)\Big)\Big\|_{L^2(\mathbb{R}^3)} \lesssim 2^{-l/2}\Big[2^{2l}\Big\|P_l\Big(\frac{1}{\lambda_n^j}u_0^n\Big(\frac{x}{\lambda_n^j}\Big)\Big)\Big\|_{L^1(\mathbb{R}^3)} + 2^l\Big\|P_l\Big(\frac{1}{(\lambda_n^j)^2}u_1^n\Big(\frac{x}{\lambda_n^j}\Big)\Big)\Big\|_{L^1(\mathbb{R}^3)}\Big]. \tag{6-19}$$

Then by interpolation, for any $l \in \mathbb{Z}$,

$$\Big\| P_l S(t+\lambda_n^j t_n^j)\Big(\frac{1}{\lambda_n^j}u_0^n(x), \frac{1}{(\lambda_n^j)^2}u_1^n(x)\Big)\Big\|_{L^4_{t,x}(\{|t+\lambda_n^j t_n^j|>C2^{-l}\}\times\mathbb{R}^3)} \lesssim \frac{1}{C^{1/4}}\Big[2^{2l}\Big\|P_l\Big(\frac{1}{\lambda_n^j}u_0^n\Big(\frac{x}{\lambda_n^j}\Big)\Big)\Big\|_{L^1(\mathbb{R}^3)} + 2^l\Big\|P_l\Big(\frac{1}{(\lambda_n^j)^2}u_1^n\Big(\frac{x}{\lambda_n^j}\Big)\Big)\Big\|_{L^1(\mathbb{R}^3)}\Big]. \tag{6-20}$$

Thus if $\limsup_{n\to\infty} \lambda_n^j |t_n^j| = \infty$, then

$$S(t+\lambda_n^j t_n^j)\Big(\frac{1}{\lambda_n^j}u_0^n(x), \frac{1}{(\lambda_n^j)^2}u_1^n(x)\Big) \rightharpoonup 0 \tag{6-21}$$

weakly in $L^4_{t,x}(\mathbb{R}\times\mathbb{R}^3)$. Therefore, $(\phi_0^j, \phi_1^j) = (0,0)$, completing the proof of Lemma 6.3. □

Thus $\lambda_n^j |t_n^j|$ is uniformly bounded, so after passing to a subsequence, $\lambda_n^j t_n^j$ converges to some $t_0^j \in \mathbb{R}$. Therefore,

$$S(\lambda_n^j t_n^j)(\phi_0^j, \phi_1^j) \to S(t_0^j)(\phi_0^j, \phi_1^j) \tag{6-22}$$

strongly in $\dot H^{1/2}(\mathbb{R}^3)\times \dot H^{-1/2}(\mathbb{R}^3)$. Absorbing the error into $(R_{0,n}^N, R_{1,n}^N)$ and taking

$$(\tilde\phi_0^j, \tilde\phi_1^j) = S(t_0^j)(\phi_0, \phi_1), \tag{6-23}$$

assume $t_n^j \equiv 0$. Therefore,

$$(u_0^n, u_1^n) = \sum_{j=1}^N \big(\lambda_n^j \phi_0^j(\lambda_n^j x), (\lambda_n^j)^2\phi_1^j(\lambda_n^j x)\big) + (R_{0,n}^N, R_{1,n}^N), \tag{6-24}$$

and

$$\lim_{n\to\infty} \frac{\lambda_j^n}{\lambda_k^n} + \frac{\lambda_k^n}{\lambda_j^n} = \infty. \tag{6-25}$$

But then

$$\|u_0^n\|_{B^2_{1,1}(\mathbb{R}^3)} + \|u_1^n\|_{B^1_{1,1}(\mathbb{R}^3)} \le C_0 < \infty \tag{6-26}$$

combined with Lemma 6.3, (6-24), and (6-25) implies that for any j

$$\|\phi_0^j\|_{B^2_{1,1}(\mathbb{R}^3)} + \|\phi_1^j\|_{B^1_{1,1}(\mathbb{R}^3)} \le C_0. \tag{6-27}$$

Possibly reordering j, (6-8) implies that there exists $N_0(\epsilon, C_0)$ such that if $j \ge N_0(\epsilon)$,

$$\|(\phi_0^j, \phi_1^j)\|_{\dot H^{1/2}\times\dot H^{-1/2}} < \epsilon. \tag{6-28}$$

Now for each j let $v^j(t,x)$ be the solution of (1-1) with initial data (ϕ_0^j, ϕ_1^j). By the small-data arguments of [Lindblad and Sogge 1995], when $j \geq N_0(\epsilon)$,

$$\|v^j\|_{L^4_{t,x}(\mathbb{R}\times\mathbb{R}^3)} \lesssim \|\phi_0^j\|_{\dot{H}^{1/2}(\mathbb{R}^3)} + \|\phi_1^j\|_{\dot{H}^{-1/2}(\mathbb{R}^3)}. \tag{6-29}$$

Meanwhile, by Theorem 5.1 combined with (6-27), when $j \leq N_0(\epsilon)$,

$$\|v^j\|_{L^4_{t,x}(\mathbb{R}\times\mathbb{R}^3)} \lesssim_{j,C_0} 1. \tag{6-30}$$

Also by (6-25), for any $j \neq k$, the Lebesgue dominated convergence theorem implies

$$\lim_{n\to\infty} \iint |\lambda_n^j v^j(\lambda_n^j t, \lambda_n^j x)|^2 \, |\lambda_n^k v^k(\lambda_n^k t, \lambda_n^k x)|^2 \, dx\, dt = 0. \tag{6-31}$$

Therefore,

$$\lim_{n\to\infty} \Big\| \sum_{1\leq j\leq N} \lambda_n^j v^j(\lambda_n^j t, \lambda_n^j x) \Big\|_{L^4_{t,x}(\mathbb{R}\times\mathbb{R}^3)} \tag{6-32}$$

is uniformly bounded, independent of N. Also,

$$F\Big(\sum_{j=1}^N \lambda_n^j v^j(\lambda_n^j t, \lambda_n^j x)\Big) - \sum_{j=1}^N F(\lambda_n^j v^j(\lambda_n^j t, \lambda_n^j x)) \\ = \sum_{1\leq j\neq k\leq N} O(|\lambda_n^j v^j(\lambda_n^j t, \lambda_n^j x)|\,|\lambda_n^k v^k(\lambda_n^k t, \lambda_n^k x)|^2), \tag{6-33}$$

so by (6-30), (6-31), and (6-32),

$$\lim_{n\to\infty} \Big\| F\Big(\sum_{j=1}^N \lambda_n^j v^j(\lambda_n^j t, \lambda_n^j x)\Big) - \sum_{j=1}^N F(\lambda_n^j v^j(\lambda_n^j t, \lambda_n^j x)) \Big\|_{L^{4/3}_{t,x}(\mathbb{R}\times\mathbb{R}^3)} = 0. \tag{6-34}$$

Therefore, by Lemma 2.6, the solution $u_N^n(t,x)$ to (1-1) with initial data

$$\sum_{j=1}^N (\lambda_n^j \phi_0^j(\lambda_n^j x), (\lambda_n^j)^2 \phi_1^j(\lambda_n^j x)) \tag{6-35}$$

has

$$\lim_{n\to\infty} \|u_N^n(t)\|_{L^4_{t,x}(\mathbb{R}\times\mathbb{R}^3)} \tag{6-36}$$

bounded uniformly in N. By another application of Lemma 2.6 combined with

$$\lim_{N\to\infty} \limsup_{n\to\infty} \|S(t)(R_{0,n}^N, R_{1,n}^N)\|_{L^4_{t,x}(\mathbb{R}\times\mathbb{R}^3)} = 0, \tag{6-37}$$

if $u^n(t)$ is the solution to (1-1) with initial data (u_0^n, u_1^n) satisfying (6-3), then

$$\|u^n(t)\|_{L^4_{t,x}(\mathbb{R}\times\mathbb{R}^3)} \tag{6-38}$$

is uniformly bounded. This proves Theorem 1.4. □

Acknowledgements

The author was partially supported by NSF grant DMS-1500424 during the writing of this paper. The author was also hosted at the IHES in Paris when working on this paper. The author also thanks the anonymous referee for many helpful comments and suggestions.

References

[Bahouri and Gérard 1999] H. Bahouri and P. Gérard, "High frequency approximation of solutions to critical nonlinear wave equations", *Amer. J. Math.* **121**:1 (1999), 131–175. MR Zbl

[Bahouri and Shatah 1998] H. Bahouri and J. Shatah, "Decay estimates for the critical semilinear wave equation", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **15**:6 (1998), 783–789. MR Zbl

[Christ et al. 2003] M. Christ, J. Colliander, and T. Tao, "Asymptotics, frequency modulation, and low regularity ill-posedness for canonical defocusing equations", *Amer. J. Math.* **125**:6 (2003), 1235–1293. MR Zbl

[Dodson 2016] B. Dodson, "Global well-posedness and scattering for the radial, defocusing, cubic wave equation with almost sharp initial data", preprint, 2016. arXiv

[Dodson and Lawrie 2015] B. Dodson and A. Lawrie, "Scattering for the radial 3D cubic wave equation", *Anal. PDE* **8**:2 (2015), 467–497. MR Zbl

[Evans 2010] L. C. Evans, *Partial differential equations*, 2nd ed., Graduate Studies in Mathematics **19**, American Mathematical Society, Providence, RI, 2010. MR Zbl

[Georgiev et al. 1997] V. Georgiev, H. Lindblad, and C. D. Sogge, "Weighted Strichartz estimates and global existence for semilinear wave equations", *Amer. J. Math.* **119**:6 (1997), 1291–1319. MR Zbl

[Gérard 1998] P. Gérard, "Description du défaut de compacité de l'injection de Sobolev", *ESAIM Control Optim. Calc. Var.* **3** (1998), 213–233. MR Zbl

[Ginibre and Velo 1995] J. Ginibre and G. Velo, "Generalized Strichartz inequalities for the wave equation", *J. Funct. Anal.* **133**:1 (1995), 50–68. MR Zbl

[Ginibre et al. 1992] J. Ginibre, A. Soffer, and G. Velo, "The global Cauchy problem for the critical nonlinear wave equation", *J. Funct. Anal.* **110**:1 (1992), 96–130. MR Zbl

[Grillakis 1990] M. G. Grillakis, "Regularity and asymptotic behaviour of the wave equation with a critical nonlinearity", *Ann. of Math.* (2) **132**:3 (1990), 485–509. MR Zbl

[Kenig 2015] C. E. Kenig, *Lectures on the energy critical nonlinear wave equation*, CBMS Regional Conference Series in Mathematics **122**, American Mathematical Society, Providence, RI, 2015. MR Zbl

[Kenig and Merle 2008] C. E. Kenig and F. Merle, "Global well-posedness, scattering and blow-up for the energy-critical focusing non-linear wave equation", *Acta Math.* **201**:2 (2008), 147–212. MR Zbl

[Lindblad and Sogge 1995] H. Lindblad and C. D. Sogge, "On existence and scattering with minimal regularity for semilinear wave equations", *J. Funct. Anal.* **130**:2 (1995), 357–426. MR Zbl

[Miao et al. 2018] C. Miao, J. Yang, and T. Zhao, "The global well-posedness and scattering for the 5D defocusing conformal invariant NLW with radial initial data in a critical Besov space", preprint, 2018. arXiv

[Nakanishi 1999] K. Nakanishi, "Unique global existence and asymptotic behaviour of solutions for wave equations with non-coercive critical nonlinearity", *Comm. Partial Differential Equations* **24**:1-2 (1999), 185–221. MR Zbl

[Ramos 2012] J. Ramos, "A refinement of the Strichartz inequality for the wave equation with applications", *Adv. Math.* **230**:2 (2012), 649–698. MR Zbl

[Shatah and Struwe 1993] J. Shatah and M. Struwe, "Regularity results for nonlinear wave equations", *Ann. of Math.* (2) **138**:3 (1993), 503–518. MR Zbl

[Shen 2017] R. Shen, "Scattering of solutions to the defocusing energy subcritical semi-linear wave equation in 3D", *Comm. Partial Differential Equations* **42**:4 (2017), 495–518. MR Zbl

[Smith and Sogge 2000] H. F. Smith and C. D. Sogge, "Global Strichartz estimates for nontrapping perturbations of the Laplacian", *Comm. Partial Differential Equations* **25**:11-12 (2000), 2171–2183. MR Zbl

[Strichartz 1977] R. S. Strichartz, "Restrictions of Fourier transforms to quadratic surfaces and decay of solutions of wave equations", *Duke Math. J.* **44**:3 (1977), 705–714. MR Zbl

[Struwe 1988] M. Struwe, "Globally regular solutions to the u^5 Klein–Gordon equation", *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **15**:3 (1988), 495–513. MR Zbl

[Tao 2006a] T. Tao, *Nonlinear dispersive equations: local and global analysis*, CBMS Regional Conference Series in Mathematics **106**, American Mathematical Society, Providence, RI, 2006. MR Zbl

[Tao 2006b] T. Tao, "Spacetime bounds for the energy-critical nonlinear wave equation in three spatial dimensions", *Dyn. Partial Differ. Equ.* **3**:2 (2006), 93–110. MR Zbl

[Tataru 2001] D. Tataru, "Strichartz estimates in the hyperbolic space and global existence for the semilinear wave equation", *Trans. Amer. Math. Soc.* **353**:2 (2001), 795–807. MR Zbl

Received 10 Jul 2017. Revised 2 Apr 2018. Accepted 5 Jul 2018.

BENJAMIN DODSON: bdodson4@jhu.edu
Department of Mathematics, Johns Hopkins University, Baltimore, MD, United States

ANALYSIS AND PDE
Vol. 12, No. 4, 2019
dx.doi.org/10.2140/apde.2019.12.1049

NONEXISTENCE OF WENTE'S L^∞ ESTIMATE FOR THE NEUMANN PROBLEM

JONAS HIRSCH

We provide a counterexample of Wente's inequality in the context of Neumann boundary conditions. We will also show that Wente's estimate fails for general boundary conditions of Robin type.

1. Introduction

Wente's L^∞ estimate is a fundamental example of a "gain" of regularity due to the special structure of Jacobian determinants. It concerns the Dirichlet problem

$$\begin{cases} -\Delta u = f & \text{in } D, \\ u = 0 & \text{on } \partial D \end{cases} \tag{1-1}$$

for the specific choice of $f = \det(\nabla V)$ with $V \in H^1(D, \mathbb{R}^2)$. Wente's theorem states:

Theorem 1.1. *Let $\Omega \subset \mathbb{R}^2$ be the disc and $f \in \mathcal{H}^1(D)$. Then if u is the unique solution in $W_0^{1,1}(\Omega, \mathbb{R})$ to (1-1), we have the estimate*

$$\|u\|_{L^\infty(D)} + \|\nabla u\|_{L^2(D)} \leq C \|\nabla V\|^2_{L^2(D)}.$$

The proof can be found in the original article [Wente 1971]. Later on it was proved that Wente's inequality holds true under the slightly weaker assumption that $f \in \mathcal{H}^1(D)$, where $\mathcal{H}^1(D)$ is the local Hardy space; see [Semmes 1994, Definition 1.90]. Proofs can be found for instance in [Hélein 2002; Topping 1997]. This estimate found many applications; an incomplete list includes [Rivière 2008; Colding and Minicozzi 2008; Lamm and Lin 2013].

It is natural to ask whether a similar estimate holds true for the Neumann problem

$$\begin{cases} -\Delta u = f & \text{in } D, \\ \dfrac{\partial u}{\partial \nu} = \dfrac{1}{2\pi} \displaystyle\int_D f & \text{on } \partial D, \end{cases} \tag{1-2}$$

again for the specific choice of $f = \det(\nabla V)$ with $V \in H^1(D, \mathbb{R}^2)$.

The aim of this note is to show that Wente's L^∞ estimate fails for the Neumann problem.

MSC2010: primary 35J05; secondary 35J25.
Keywords: compensated compactness, Jacobian determinants.

Theorem 1.2. *There exists a sequence $V_n = (a_n, b_n) \in C^\infty(\bar{D}, \mathbb{R}^2)$, $\|\nabla V_n\|_{L^{2,1}(D)} \leq C$, $\int \det(\nabla V_n) = 0$ for all n with the property that if $u_n \in W^{1,1}(D)$ are the solutions to (1-2) with $f_n = \det(\nabla V_n)$ one has*

$$\left\| u_n - \fint_D u_n \right\|_{L^\infty(D)}, \|\nabla u_n\|_{L^2(D)} \to +\infty \quad \text{as } n \to \infty.$$

Additionally we can extend the above example to more general boundary conditions. Namely we have the following:

Theorem 1.3. *Let $E \subset \partial D$ be a nonempty union of open intervals, with $0 < \mathcal{H}^1(E) < 2\pi$ and $\alpha, \beta, \gamma \in \mathbb{R}$ given, with $\alpha > 0$, $\gamma \geq 0$. There exists a sequence $V_n = (a_n, b_n) \in C^\infty(\bar{D}, \mathbb{R}^2)$, with $\|\nabla V_n\|_{L^{2,1}(D)} < C$, with the property that if $u_n \in W^{1,1}(D)$ is the solution to*

$$\begin{cases} -\Delta u_n = \det(\nabla V_n) & \text{in } D, \\ \alpha \dfrac{\partial u_n}{\partial \nu} + \beta \dfrac{\partial u_n}{\partial \tau} + \gamma u_n = 0 & \text{on } E, \\ u = 0 & \text{on } \partial D \setminus E, \end{cases} \tag{1-3}$$

one has

$$\|\nabla u_n\|_{L^2(D)} \to \infty \quad \text{as } n \to \infty.$$

The paper is organized as follows. In Section 3 we collect some known results and a priori estimates. In Section 4 we give the proof of Theorem 1.2 and in Section 5 its extension to mixed Robin boundary conditions.

While finishing this paper the author became aware that a similar example has been found independently by Francesca Da Lio and Francesco Palmurella [2017].

2. Some remarks on the conformal invariance of the problem

Let $m : U \to D$ be a smooth conformal map from a domain U with Lipschitz continuous boundary to the disc (i.e., up to conjugation m corresponds to holomorphic map on U). If u is a solution of the Dirichlet problem (1-1) then $u \circ m$ is a solution of

$$\begin{cases} -\Delta(u \circ m) = \left(\tfrac{1}{2}|\nabla m|^2\right) f \circ m & \text{in } U, \\ u \circ m = 0 & \text{on } \partial U. \end{cases}$$

In particular in the case $f = \det(\nabla V)$ we have $\left(\frac{1}{2}|\nabla m|^2\right) f \circ m = \det(\nabla(V \circ m))$. Additionally one notes that Wente's estimate in Theorem 1.1 is as well conformally invariant since for any function w one has

$$\|w \circ m\|_{L^\infty(U)} = \|w\|_{L^\infty(D)}, \quad \|\nabla(w \circ m)\|_{L^2(U)} = \|\nabla w\|_{L^2(D)}.$$

In the case of the Neumann problem one has to be a bit more careful. If u is a solution to (1-2) then $u \circ m$ solves

$$\begin{cases} -\Delta(u \circ m) = \left(\tfrac{1}{2}|\nabla m|^2\right) f \circ m & \text{in } U, \\ \dfrac{\partial(u \circ m)}{\partial \nu} = \left(\tfrac{1}{2}|\nabla m|^2\right)^{1/2} \dfrac{1}{2\pi} \displaystyle\int_D f & \text{on } \partial U. \end{cases}$$

Although we have

$$\frac{1}{2\pi}\int_D f = \frac{1}{2\pi}\int_U \big(\tfrac{1}{2}|\nabla m|^2\big) f\circ m,$$

the problem is only conformally "invariant" if $\int_D f = 0$ since $|\nabla m| = 1$ on ∂U implies that m is a rigid motion. Furthermore one should note that even in the case $\int f = 0$, in general one has

$$\fint_U u\circ m \neq \fint_D u.$$

Nonetheless we can forget about the additional condition $\int_D \det(\nabla V_n) = 0$ in the proofs of Theorems 1.2 and 1.3 by the following procedure. Consider a sequence V_n as stated, but not necessarily satisfying $\alpha_n := \int \det(\nabla V_n) = 0$, that is compactly supported in some ball $B_{r_0}(p)$ for some $0 < r_0 < \frac{1}{4}$ and $p \in \partial D$. Let us fix two smooth functions $\hat{a}, \hat{b}$ supported in $B_{2r_0}(p)\setminus B_{r_0}(p)$ satisfying

$$\int_D d\hat{a}\wedge d\hat{b} = 1.$$

For instance take $\hat{a} = \varphi_1(z)$ and $\hat{b} = \varphi_2(z)\theta$, where φ_i are two bump functions such that $\operatorname{spt}(\varphi_1) \subset \{\varphi_2 = 1\}$,

$$\int_D d\hat{a}\wedge d\hat{b} = \int_{\partial D} \hat{a}\nabla_\theta \hat{b} = \int_{\partial D}\varphi_1 = 1.$$

Let $\hat{u}$ be the smooth unique solution to (1-2) with $\fint_D \hat{u} = 1$, $f = \det(\nabla\widehat{V})$ and $\widehat{V} = (\hat{a}, \hat{b})$. Since $|\alpha_n| \le \frac{1}{2}\|\nabla V_n\|^2_{L^2(D)}$ and $\operatorname{spt}(V_n)\cap\operatorname{spt}(\widehat{V}) = \varnothing$ for all n we can pass to $\tilde{u}_n = u_n - \alpha_n\hat{u}$, which solves the Neumann problem (1-2) with right-hand side

$$\det(\nabla V_n - \alpha\nabla\widehat{V}) = \det(\nabla V_n) - \alpha\det(\nabla\widehat{V}).$$

Since $\int_D \det(\nabla V_n - \alpha\nabla\widehat{V}) = 0$ we have $\partial\tilde{u}_n/\partial\nu = 0$ on ∂D. By the uniform boundedness of α_n we still have

$$\Big\|\tilde{u}_n - \fint_D \tilde{u}_n\Big\|_{L^\infty(D)}, \|\nabla\tilde{u}_n\|_{L^2(D)} \to +\infty \quad \text{as } n\to\infty$$

and we obtain the full strength of the theorems.

3. Some known results

Classical solutions to (1-1) and (1-2) have to be understood in the distributional sense.

Definition 3.1. A function u is called a solution of the Dirichlet problem if $u \in W_0^{1,1}(D,\mathbb{R})$ and

$$\int_D \nabla u\cdot\nabla\psi - f\psi = 0 \quad \text{for all } \psi \in C_0^1(D). \tag{3-1}$$

A function u is called a solution of the Neumann problem if $u \in W^{1,1}(D,\mathbb{R})$ and

$$\frac{1}{2\pi}\int_D f\int_{\partial D}\psi = \int_D \nabla u\cdot\nabla\psi - f\psi \quad \text{for } \psi \in C_0^\infty(\mathbb{R}^2) \text{ for all } \psi \in C^1(\overline{D}). \tag{3-2}$$

The Green's functions for both problems are explicit. For the Dirichlet problem it is

$$G_D(x,y)=\frac{1}{2\pi}\ln(|x-y|)-\frac{1}{2\pi}\ln(|y|\,|x-y^*|),\quad \text{with } y^*=\frac{y}{|y|^2}, \tag{3-3}$$

and for Neumann problem it is

$$G_N(x,y)=\frac{1}{2\pi}\ln(|x-y|)+\frac{1}{2\pi}\ln(|y|\,|x-y^*|)-\tfrac{1}{4}|x|^2-\tfrac{1}{4}|y|^2. \tag{3-4}$$

Using G_N one has the representation formula

$$u(y)-\int_D u=-\int_{\partial D} G_N(x,y)\frac{\partial u}{\partial \nu}+\int_D G(x,y)\Delta u \quad \text{for } u\in C^2(\bar{D}).$$

In terms of existence and uniqueness one has:

Lemma 3.2. *For every $f\in L^1(D)$ there exists a solution u_D/u_N to the Dirichlet/ Neumann problem in the sense of Definition 3.1. Furthermore the solutions belong to $W^{1,p}(D,\mathbb{R})$ for every $p<2$, are unique* (*up to constant in the Neumann problem*) *and satisfy the estimate*

$$\|Du\|_{L^p(D)}\leq C_p\|f\|_{L^1(D)}. \tag{3-5}$$

Proof. There are several proofs in the literature treating the case of uniqueness and a priori estimates; see for instance [Littman et al. 1963; Ancona 2009, Appendix A]. In our case existence and the a priori estimate (3-5) can be obtained by using the Green's functions G_D, G_N. Uniqueness for the Dirichlet problem can be obtained by antisymmetric reflection: Let u be a distributional solution of (3-1) with $f=0$. One checks that

$$\hat{u}(x):=\begin{cases} u(x) & \text{for } x\in D,\\ -u(x^*) & \text{for } x\notin D \text{ with } x^*=x/|x|^2\end{cases}$$

solves

$$\int_{\mathbb{R}^2}\nabla\hat{u}\cdot\nabla\psi=\int_D \nabla u\cdot\nabla(\psi(x)-\psi(x^*))\quad \text{for all } \psi\in C^1_c(\mathbb{R}^2).$$

But since $\psi(x)-\psi(x^*)\in C^{0,1}_0(D)$ we deduce that $\hat{u}$ is harmonic and therefore smooth in $\mathbb{R}^2$. Now the maximum principle applies since u takes the boundary values in the strong sense.

Similarly we deduce the uniqueness in the Neumann problem using the symmetric reflection: Let v be a distributional solution of (3-2) with $f=0$. As before one checks that

$$\hat{v}(x):=\begin{cases} v(x) & \text{for } x\in D,\\ v(x^*) & \text{for } x\notin D\end{cases}$$

solves

$$\int_{\mathbb{R}^2}\nabla\hat{v}\cdot\nabla\psi=\int_D \nabla v\cdot\nabla(\psi(x)+\psi(x^*))\quad \text{for all } \psi\in C^1_c(\mathbb{R}^2).$$

But since $\psi(x)+\psi(x^*)\in C^{0,1}(\bar{D})$ we deduce that $\hat{v}$ is harmonic and therefore smooth in $\mathbb{R}^2$. Now the maximum principle implies that $v=\text{constant}$. ☐

4. Proof of Theorem 1.2

In the following we will always identify $\mathbb{R}^2$ with the complex plane $\mathbb{C}$, i.e., $i = e_2$.

Proof of Theorem 1.2. The main step of the proof consists in the following claim: For every $r_0 > 0$ there exists a sequence $(a_n, b_n) \in C^\infty(\bar{D}, \mathbb{R}^2)$ with the properties that

$$\operatorname{spt}(a_n) \cup \operatorname{spt}(b_n) \subset B_{r_0}(-e_2), \tag{4-1a}$$

$$a_n, b_n \rightharpoonup 0 \quad \text{in } H^1(D), \tag{4-1b}$$

$$\|a_n\|_{L^\infty(D)} + \|\nabla a_n\|_{L^{2,1}(D)}, \qquad \|b_n\|_{L^\infty(D)} + \|\nabla b_n\|_{L^{2,1}(D)} \le C, \tag{4-1c}$$

$$\|da_n \wedge db_n\|_{H^{-1}(D)} \to \infty \quad \text{as } n \to \infty. \tag{4-2}$$

Given such a sequence we can conclude the theorem. Let u_n be the unique solution to the Dirichlet problem (1-1) with right-hand side $f_n = da_n \wedge db_n$ and h_n be the unique harmonic function satisfying

$$\frac{\partial h_n}{\partial \nu} = \frac{\partial u_n}{\partial \nu} - \frac{1}{2\pi} \int_{\partial D} \frac{\partial u_n}{\partial \nu} \quad \text{on } \partial D.$$

Such a harmonic function exists since

$$\int_{\partial D} \left(\frac{\partial u_n}{\partial \nu} - \frac{1}{2\pi} \int_{\partial D} \frac{\partial u_n}{\partial \nu} \right) = 0.$$

It is straightforward to check that

$$v_n := u_n - h_n$$

is the unique solution to the Neumann problem (1-2). Observe that v_n is a Cauchy sequence in $W^{1,p}(D)$ for all $p < 2$ converging to $v \in W^{1,p}(D)$, the unique solution of (1-2) with $f = da \wedge db$. By Wente's theorem we have

$$\|\nabla v_n\|_{L^2(D)} \ge \|\nabla h_n\|_{L^2(D)} - \|\nabla u_n\|_{L^2(D)} \ge \|\nabla h_n\|_{L^2(D)} - C\|\nabla a_n\|_{L^2(D)} \|\nabla b_n\|_{L^2(D)}.$$

The theorem follows by showing that

$$\|\nabla h_n\|_{L^2(D)} \to \infty. \tag{4-3}$$

To do so we will use the Dirichlet-to-Neumann map in the following formulation: Let

$$X_0 := \left\{ h \in H^1(D) : \Delta h = 0 \text{ in } D \text{ and } ⨍_D h = 0 \right\},$$
$$Y_0 := \left\{ u \in H^1(D) : ⨍_D u = 0 \right\}.$$

Endowed with the L^2 inner product $\langle u, v \rangle = \int_D \nabla u \cdot \nabla v$, we obtain Hilbert spaces satisfying $X_0 \subset Y_0$. If we set $Z_0^* := \{ l \in Y_0^* : l(\psi) = 0 \text{ for all } \psi \in H_0^1(D) \cap Y_0 \}$ then classical results concerning Dirichlet-to-Neumann operators imply that the operator

$$A : X_0 \to Z_0^*, \quad \text{with } Ah := \frac{\partial h}{\partial \nu},$$

is continuous and onto; i.e., it has a continuous inverse A^{-1}.

Next we identify

$$\frac{\partial u_n}{\partial \nu} - \frac{1}{2\pi}\int_{\partial D} \frac{\partial u_n}{\partial \nu}$$

with a linear functional $l_n \in Y_0^*$; i.e.,

$$l_n(\psi) := \int_{\partial D}\left(\frac{\partial u_n}{\partial \nu} - \frac{1}{2\pi}\int_{\partial D} \frac{\partial u_n}{\partial \nu}\right)\psi.$$

We will show that they are elements of Z_0^* with the property that $\|l_n\|_{H^{-1}(D)} \to +\infty$. The normal derivative of a solution $u \in W^{1,1}(D)$ to the Dirichlet problem (1-1), with $f \in L^1(D)$, is given in the sense of distributions by

$$\int_{\partial D} \frac{\partial u}{\partial \nu}\psi := \int_D \nabla u \cdot \nabla \psi - f\psi \quad \text{for } \psi \in C^1(\bar{D}). \tag{4-4}$$

The distribution is supported on ∂D since given $\psi_1, \psi_2 \in C^\infty(\bar{D})$ with $\psi_1 = \psi_2$ on ∂D we have $\varphi = \psi_1 - \psi_2 \in C_0^1(\bar{D})$ with $\varphi = 0$ on ∂D and so by (3-1) we have

$$\int_{\partial D} \frac{\partial u}{\partial \nu}\varphi = \int_D \nabla u \cdot \nabla \varphi - f\varphi = 0.$$

By density of $C_c^\infty(D)$ in $H_0^1(D)$ we conclude $l_n(\psi) = 0$ for all $\psi \in H_0^1(D)$. Furthermore it is straightforward to check that l_n vanishes on the constant functions and hence l_n is a well-defined element of Y_0^*, since $l_n(\psi) = l_n(\psi - ⨍\psi)$. Thus we conclude that $l_n \in Z_0^*$ for all n. The first part of (4-4) and the second part in the definition of l_n are uniformly bounded by Wente's theorem (Theorem 1.1) because

$$\int_D \nabla u_n \cdot \nabla \psi \le \|\nabla u_n\|_{L^2(D)} \|\nabla \psi\|_{L^2(D)}$$
$$\left|\frac{1}{2\pi}\int_{\partial D} \frac{\partial u_n}{\partial \nu}\right| = \left|\frac{1}{2\pi}\int_D f_n\right| \le \frac{1}{2\pi}\|\nabla a_n\|_{L^2(D)} \|\nabla b_n\|_{L^2(D)}.$$

Hence $\|l_n\|_{H^{-1}(D)} \to \infty$ by (4-2). Since $h_n = A^{-1}(l_n)$ and A^{-1} is continuous, we conclude (4-3).

It remains to construct the sequence (a_n, b_n) with the properties (4-1)–(4-2). Performing a translation we can consider the translated disc $D' := D + i$; i.e., $D' \subset H := \mathbb{C} \cap \{y \ge 0\} = \{re^{i\theta} : 0 < \theta < \pi\}$. Furthermore one readily checks that if $\Re(h)$ and $\Im(h)$ are the real and imaginary parts of a holomorphic function h then we have pointwise

$$d\Re(h) \wedge d\Im(h) = |h'(z)|^2 dx \wedge dy \quad \text{and} \quad |d\Re(h)|^2 = |d\Im(h)|^2 = |h'(z)|^2. \tag{4-5}$$

We will construct our contradicting sequence (a_n, b_n) as the real and imaginary parts of a sequence of holomorphic functions h_n on H multiplied by a truncation function φ.

Indeed consider the family of Möbius transformations of the complex plane $\mathbb{C}$

$$m_\epsilon(z) := \frac{z - i\epsilon}{z + i\epsilon}.$$

We observe that m_ϵ maps the upper half-space H onto the disc D for every $\epsilon > 0$. Furthermore one readily calculates

$$m'_\epsilon(z) = \frac{2i\epsilon}{(z+i\epsilon)^2}, \quad m_\epsilon^{-1}(z) = i\epsilon\frac{z+1}{1-z}. \tag{4-6}$$

We note that for every $\delta > 0$ one has $m'_\epsilon(z) \to 0$ and $m_\epsilon(z) \to 1$ uniformly on $\mathbb{C} \setminus D_\delta$ for $\epsilon \to 0$. Furthermore $m_\epsilon^{-1}(z) \to 0$ uniformly on $\mathbb{C} \setminus D_\delta(1)$. Thus we can conclude that $l_\epsilon := |m'_\epsilon(z)|^2\, dx \wedge dy \to \pi\delta_0$ in the sense of distributions; i.e., given $\psi \in C_c^0(\mathbb{C})$ arbitrary one has

$$\int_H \psi(z)|m'_\epsilon(z)|^2\, dx \wedge dy = \int_D \psi \circ m_\epsilon^{-1}(z)\, dx \wedge dy \to \psi(0)\pi.$$

Furthermore we conclude that if φ is any cutoff function with $\varphi = 1$ in a neighborhood of 0 we still have $l_\epsilon \lfloor \varphi \to \pi\delta_0$. Since $\pi\delta_0 \notin H^{-1}(H)$ we conclude that $\|l_\epsilon \lfloor \varphi\|_{H^{-1}(D)} \to \infty$ as $\epsilon \to 0$. Fixing a sequence $\epsilon_n \to 0$, we have

$$a_n := \varphi\, \Re(m_{\epsilon_n} - 1) \quad \text{and} \quad b_n := \varphi\, \Im(m_{\epsilon_n} - 1)$$

satisfy $a_n, b_n \in C^\infty(H)$ and $a_n, b_n \to 0$ uniformly in C^1 on $\overline{H} \setminus D_\delta$ for any $\delta > 0$. Hence for an appropriate choice of φ the first two parts of (4-1) follow.

We calculate

$$da_n \wedge db_n = l_\epsilon \lfloor \varphi^2 + \varphi d\varphi \wedge \big(\Re(m_{\epsilon_n}) d\Im(m_{\epsilon_n}) - \Im(m_{\epsilon_n}) d\Re(m_{\epsilon_n})\big) = l_\epsilon \lfloor \varphi^2 + \varphi d\varphi \wedge w_\epsilon.$$

Since we have $\operatorname{spt}(d\varphi) \subset \mathbb{C} \setminus D_\delta$ for some $\delta > 0$ and $|w_\epsilon| \to 0$ uniformly on $\mathbb{C} \setminus D_\delta$ we conclude that $\|\varphi d\varphi \wedge w_\epsilon\|_{H^{-1}} \to 0$ as $n \to \infty$. Hence $da_n \wedge db_n \to \pi\delta_0$ in the sense of distributions and therefore $\|da_n \wedge db_n\|_{H^{-1}(H)} \to \infty$ as $n \to \infty$; i.e., (4-2) holds.

It remains to show that $|da_n|$, $|db_n|$ are uniformly bounded in $L^{2,1}$. By (4-6) we have

$$\{z \in H : |m'_\epsilon(z)| \geq t\} = B_{r(t)}(-i\epsilon) \cap H, \quad \text{with } \frac{2\epsilon}{r(t)^2} = t$$

and $|m'_\epsilon|(z) \leq 2/\epsilon$ for all $z \in H$. Hence we may estimate

$$\mu(t) := |\{z \in H : |m'_\epsilon(z)| \geq t\}| \leq \pi r(t)^2 = \frac{2\epsilon}{t}\pi.$$

Recall that the $L^{2,1}$ norm can be written as

$$\|f\|_{L^{2,1}(H)} = 2\int_0^\infty \mu_f(t)^{1/2}\, dt.$$

Here $\mu_f(t) = |\{z \in H : |f(z)| > t\}|$ is the distribution function; see [Grafakos 2014, Proposition 1.4.9]. Using the estimates above we obtain

$$\||m'_\epsilon|\|_{L^{2,1}(H)} \leq 2\sqrt{2\pi\epsilon}\int_0^{2/\epsilon} \frac{1}{\sqrt{t}}\, dt \leq 8\sqrt{\pi},$$

which is uniformly bounded in ϵ, proving the last part of (4-1). ☐

Remark 4.1. Observe that if the solution to the Neumann problem is not in $H^1(D)$ then it can neither be in L^∞ nor in $W^{2,1}(D)$. Indeed $u \in W^{2,1}(D)$ would imply $u \in L^\infty$ since $W^{2,1}(D)$ embeds in L^∞ in two dimensions; see for instance Theorem 3.3.10 combined with Theorem 3.3.4 in [Hélein 2002]. If u were in $L^\infty(D)$ then we could take $u_\epsilon \in C^\infty(\bar{D})$ with $u_\epsilon \to u$ in $W^{1,1}(D)$ and uniformly bounded in $L^\infty(D)$. Testing (3-2) with u_ϵ would give

$$\int_D \nabla u \cdot \nabla u_\epsilon = \int_D f u_\epsilon + \frac{1}{2\pi} \int_D f \int_{\partial D} u_\epsilon \leq 2\|f\|_{L^1} \|u_\epsilon\|_{L^\infty}.$$

The right-hand side is bounded independent of ϵ so we conclude that $u \in H^1(D)$, a contradiction.

By using more or less an abstract functional analytic argument we are able to obtain the following corollary. Its proof is presented in the Appendix.

Corollary 4.2. *There exists $a, b \in H^1(D)$ with the additional properties $a, b \in L^\infty(D)$ and $da, db \in L^{2,1}(D)$ such that if $u \in W^{1,1}(D)$ denotes the solution to the Neumann problem (1-2) with $f = da \wedge db$ then $u \notin H^1(D)$.*

5. More general boundary conditions

Our construction of the counterexample relies mainly on the continuity of the Dirichlet-to-Neumann map D_0. The extension to more general boundary conditions of Robin type follows by finding a replacement of the Dirichlet-to-Neumann map. The replacement is constructed as follows:

$$X := \{h \in H^1(D) : \Delta h = 0 \text{ in } D \text{ and } h = 0 \text{ on } \partial D \setminus E\},$$
$$Y := \{u \in H^1(D) : u = 0 \text{ on } \partial D \setminus E\}.$$

Since by assumption $\mathcal{H}^1(\partial D \setminus E) > 0$ we can endow X, Y with the norm $\|u\| = \|\nabla u\|_{L^2(D)}$. Finally we define the closed subset $Z^* \subset Y^*$ by

$$Z^* := \{l \in Y^* : l(u) = 0 \text{ for all } u \in H^1_0(D)\}.$$

Obviously one has the inclusion $X \subset Y$ and $Z^* \subset Y^*$.

Lemma 5.1. *The operator $B : X \to Z^*$ defined by*

$$\langle Bh, \psi \rangle = \int_{\partial D} \left(\alpha \frac{\partial h}{\partial \nu} + \beta \frac{\partial h}{\partial \tau} + \gamma h \right) \psi := \alpha \int_D \nabla h \cdot \nabla \psi + \beta \int_{\partial D} \frac{\partial h}{\partial \tau} \psi + \gamma \int_{\partial D} h \psi$$

is continuous and onto, with continuous inverse $B^{-1} : Z^* \to X$.

Proof. Instead of B itself we consider the family of operators $B_s : X \to Z^*$ for $s \in [0, 1]$. B_s is defined as B with $s\beta, s\gamma$ replacing β, γ. Since h is harmonic in D we have $\langle B_s h, \psi \rangle = 0$ for all $\psi \in H^1_0(D)$ by density of $C^\infty_c(D)$ in $H^1_0(D)$. Furthermore we have the estimate

$$\begin{aligned}\langle B_s h, \psi \rangle &\leq \alpha \|\nabla h\|_{L^2(D)} + |s\beta| \left\| \frac{\partial h}{\partial \tau} \right\|_{H^{-1/2}\partial D} \|\psi\|_{H^{1/2}\partial D} + s\gamma \|h\|_{L^2(\partial D)} \|\psi\|_{L^2(\partial D)} \\ &\leq (\alpha + C|\beta| + C\gamma) \|\nabla h\|_{L^2(D)} \|\nabla \psi\|_{L^2(D)}.\end{aligned}$$

In the last line we used that for harmonic functions we have

$$\left\|\frac{\partial h}{\partial \tau}\right\|_{H^{-1/2}(\partial D)} = \left\|\frac{\partial h}{\partial \nu}\right\|_{H^{-1/2}(\partial D)} = \|\nabla h\|_{L^2(D)}$$

and the trace theorem for Sobolev functions.

This shows that B_s is a family of uniformly bounded operators taking values in Z^*. Since $X \subset Y$ we have the lower bound

$$\begin{aligned}\langle B_s h, h\rangle &= \alpha \int_D \nabla h \cdot \nabla h + s\beta \frac{1}{2}\int_{\partial D} \frac{\partial h^2}{\partial \tau} + s\gamma \int_{\partial D} h^2 \\ &= \alpha \int_D \nabla h \cdot \nabla h + s\gamma \int_{\partial D} h^2 \geq \alpha \|\nabla h\|^2_{L^2(D)}.\end{aligned}$$

Finally since $B_s = (1-s)B_0 + sB$, the *method of continuity*, see, e.g., [Gilbarg and Trudinger 1998, Theorem 5.2], applies and $B = B_1$ is onto if and only if B_0 is onto. By construction we have $B_0 h = \alpha(\partial h/\partial \nu)$, the classical normal derivative on E, which is known to be onto by the Dirichlet-to-Neumann map. □

Now we are able to complete the proof of the theorem.

Proof of Theorem 1.3. The construction is now essentially the same as in the proof of Theorem 1.2. After a rotation we may assume that $-i = -e_2 \in E$. Fix $r_0 > 0$ such that $\partial D \cap B_{r_0}(-i) \subset E$. Let $a_n, b_n, u_n \in C^\infty(\overline{D})$ be the sequences constructed in the proof of Theorem 1.2. By the choice of $r_0 > 0$ we have ensured that

$$\operatorname{spt}(a_n) \cup \operatorname{spt}(b_n) \subset B_{r_0}(-i).$$

Observe that

$$l_n := \alpha \frac{\partial u_n}{\partial \nu} + \beta \frac{\partial u_n}{\partial \tau} + \gamma u_n \in Z^*$$

because

$$\langle B u_n, \psi\rangle = \alpha \int_{\partial D} \frac{\partial u_n}{\partial \nu}\psi = \alpha \int_D \nabla u_n \cdot \nabla \psi - \alpha \int_D da_n \wedge db_n \psi$$

and the discussion below (4-4) applies. Furthermore we have

$$\|l_n\|_{Z^*} \geq \alpha \|da_n \wedge db_n\|_{H^{-1}(D)} - \alpha \|\nabla u_n\|_{L^2(D)}.$$

By Wente's theorem (Theorem 1.1), $\|\nabla u_n\|_{L^2(D)}$ is uniformly bounded and so the application of Lemma 5.1 gives for $h_n := B^{-1}(l_n)$ that

$$\|\nabla h_n\|_{L^2(D)} \to \infty \quad \text{as } n \to \infty.$$

We conclude by observing that $v_n := u_n - h_n$ satisfies the boundary value problem (1-3) because $u_n = h_n = 0$ on $\partial D \setminus E$ and

$$\begin{cases} -\Delta v_n = -\Delta u_n = da_n \wedge db_n & \text{in } D, \\ \alpha \dfrac{\partial v_n}{\partial \nu} + \beta \dfrac{\partial v_n}{\partial \tau} + \gamma v_n = l_n - B(h_n) = 0 & \text{on } E. \end{cases}$$

The blow-up of the H^1 norm now follows since

$$\|\nabla v_n\|_{L^2(D)} \geq \|\nabla h_n\|_{L^2(D)} - \|\nabla u_n\|_{L^2(D)} \to \infty.$$ □

As before we obtain as a consequence of Theorem 1.3 the following:

Corollary 5.2. *There exists $a, b \in H^1(D)$ with the additional properties $a, b \in L^\infty(D)$ and $da, db \in L^{2,1}(D)$ such that if $u \in W^{1,1}(D)$ denotes the solution to the problem (1-3) with $f = da \wedge db$ then $u \notin H^1(D)$.*

Its combined proof with Corollary 4.2 can be found in the Appendix.

Appendix: Abstract functional analytic argument

Now we want to present the abstract functional analytic argument that leads to Corollaries 4.2 and 5.2. We will first proof an "easier" version where every embedding of the involved spaces is linear. Thereafter we show how the same idea translates to our setting.

Lemma A.1. *Consider Banach spaces $E_1 \subset E_2$ and $F_1 \subset F_2$ such that the inclusion $\subset$ corresponds to a continuous embedding. Let $A : E_2 \to F_2$ be a continuous linear operator. Suppose that F_1 is a Hilbert space and there is a sequence $\{x_n\}_{n\in\mathbb{N}}$ with the properties that*

(a) *$Ax_n \in F_1$ and $\|x_n\|_{E_1} \leq 1$ for all $n \in \mathbb{N}$;*

(b) *$\limsup_{n\to\infty} \|Ax_n\|_{F_1} = \infty$;*

(c) *$f \in F_1 \mapsto \langle Ax_n, f\rangle$ extends to a linear functional l_n on F_2 for each n.*

Then there exists $x \in E_1$ such that $Ax \in F_2 \setminus F_1$ in the sense that there is a sequence $l_n \in F_2^*$ with $\|l_n\|_{F_1^*} \leq 1$ but

$$l_n(Ax) \to \infty.$$

Proof. Passing to a subsequence we may assume that the lim sup in (b) is actually a limit.

In a first step we show by induction that there exists $\{y_1, \dots, y_n\} \in E_1$ with the properties

(i) $\|y_i\|_{E_1} \leq 1$ for all i;

(ii) $\langle Ay_i, Ay_j\rangle = 0$ if $i \neq j$;

(iii) $\|Ay_i\|_{F_1} \geq 2^{2i}$ for all i.

By (b) there exists $m_1 \in \mathbb{N}$ such that $\|Ax_{m_1}\| \geq 4$. Hence we may set $y_1 := x_{m_1}$.

Now suppose $\{y_1, \dots, y_n\}$ have been chosen. We define the linear continuous operator $P_n : F_1 \to F_1$ by

$$P_n := \sum_{i=1}^{n} \frac{Ay_i \otimes Ay_i}{\|Ay_i\|^2}.$$

It is obvious that $P_n = P_n^t$ and (ii) implies that $P_n^2 = P_n$; i.e., P_n is the orthogonal projection onto the finite-dimensional space $V_n := \operatorname{span}\{Ay_1, \dots, Ay_n\}$. Hence $(P_nA) : E_1 \to V_n$ is a continuous linear operator onto a finite-dimensional vector space. Let $(P_nA)^{-1} : V_n \to \operatorname{span}\{y_1, \dots, y_n\}$ denote the inverse of the operator (P_nA) restricted to the finite-dimensional space $\operatorname{span}\{y_1, \dots, y_n\}$. We may define now the operator

$$Q_n : E_1 \to E_1, \quad Q_n := (P_nA)^{-1} \circ (P_nA).$$

We note that Q_n is continuous and $Q_n^2 = Q_n$; hence Q_n is a projection operator. As a direct consequence we have as well that $(I - Q_n)$ is a continuous projection operator; here I denotes the identity map on E_2.

By construction we have

$$P_n\, A\,(I - Q_n) = 0. \tag{A-1}$$

The range of Q_n is finite and $(A Q_n)$ is a continuous operator and therefore

$$\limsup_{m\to\infty} \|(A Q_n) x_m\|_{F_1} < \infty.$$

Hence we have

$$\lim_{m\to\infty} \|A(I - Q_n)x_m\|_{F_1} \geq \lim_{m\to\infty} \|A x_m\|_{F_1} - \limsup_{m\to\infty} \|(A Q_n)x_m\|_{F_1} = \infty.$$

Thus there exists $m_{n+1} \in \mathbb{N}$ such that

$$\|A(I - Q_n)x_{m_{n+1}}\|_{F_1} > 2^{2(n+1)} \|I - Q_n\|.$$

We define $y_{n+1} = (I - Q_n)x_{m_{n+1}} / \|I - Q_n\|$. Clearly we have $\|y_{n+1}\|_{E_1} \leq 1$ and (iii) holds by the choice of m_{n+1}. Finally (ii) follows using that P_n is a orthogonal projection, that Q_n is a projection and (A-1):

$$\langle A y_i, A y_{n+1}\rangle = \langle P_n A y_i, A(I - Q_n) y_{n+1}\rangle = \langle P_n A y_i, (P_n A(I - Q_n)) y_{n+1}\rangle = 0.$$

Having the sequence $\{y_i\}_{i\in\mathbb{N}}$ at our disposal we obtain x as follows: For each n we define the elements $z_n \in E_1$ and $f_n \in F_1$ by

$$z_n := \sum_{i=1}^{n} 2^{-i} y_i \quad \text{and} \quad f_n := \sum_{i=1}^{n} 2^{-i} \frac{A y_i}{\|A y_i\|_{F_1}}.$$

Since E_1, F_1 are Banach spaces we have that their limits exist: $z = \lim_{n\to\infty} z_n = \sum_{i=1}^{\infty} 2^{-i} y_i \in E_1$ and

$$f = \lim_{n\to\infty} f_n = \sum_{i=1}^{\infty} 2^{-i} \frac{A y_i}{\|A y_i\|_{F_1}}.$$

Assumption (c) implies that for each $i \in \mathbb{N}$ the map

$$f \in F_1 \mapsto \left\langle \frac{A y_i}{\|A y_i\|_{F_1}}, f \right\rangle$$

extends to a continuous linear functional $l_i \in F_1^*$. Therefore the continuous linear functional $L_n := \sum_{i=1}^{n} 2^{-i} l_i$ has the desired properties using (i)–(iii) since

$$\begin{aligned} L_n(Az) &= \lim_{m\to\infty} L_n(A z_m) = \lim_{m\to\infty} \langle f_n, A z_m\rangle \\ &= \lim_{m\to\infty} \sum_{i=1}^{n} \sum_{j=1}^{m} 2^{-i-j} \left\langle \frac{A y_i}{\|A y_i\|_{F_1}}, A y_j \right\rangle = \sum_{i=1}^{n} 2^{-2i} \|A y_i\|_{F_1} \geq n. \end{aligned}$$

□

Observe that we could directly apply the above result with the following choice of spaces: let $E_1 = \mathcal{H}^1_{\mathrm{loc}}(D)$ be the local Hardy space of the disk, $E_2 = L^1(D)$, $F_1 = \left\{ f \in H^1(D) : \fint_D f = 0 \right\}$ and $F_2 = W^{1,1}(D)$. But this would not give single elements $a, b \in H^1(D)$ as stated in the Corollaries 4.2 and 5.2.

Proof of Corollaries 4.2 and 5.2. We introduce the space

$$X := \left\{h \in H^1(D) : ⨍_D h = 0 \text{ and } dh \in L^{2,1}(D)\right\}.$$

It becomes a complete Banach space with respect to the norm $\|h\|_X := \|dh\|_{L^{2,1}}$. Furthermore as suggested before we set $E_2 := L^1(D)$, $F_1 := H^1(D)$, $F_2 = W^{1,1}(D)$. Observe that we have a "bilinear" linear embedding of $X \times X \hookrightarrow E_2$ by $(h,k) \mapsto dh \wedge dk$ with $\|dh \wedge dk\|_{L^1} \leq \|dh\|_{L^{2,1}} \|dk\|_{L^{2,1}}$.

The construction of (a,b) out of the contradicting sequence is the same in the case of a Neumann or Robin-type boundary condition. Hence we will give a simultaneous proof for both. We denote by $A : L^1(D) \to W^{1,1}(D)$ the solution operator to problem (1-2) or problem (1-3). Recall that by classical elliptic theory there is a constant $C_A > 0$ such that $\|Ax\|_{H^1} \leq C_A \|x\|_{L^2}$.

Let $(a_n, b_n) \in C^\infty(\bar{D}, \mathbb{R}^2)$ be the corresponding contradicting sequence of Theorem 1.2 or Theorem 1.3. Without loss of generality we may assume that $⨍ a_n = 0 = ⨍ b_n$ for all n; hence $a_n, b_n \in X$. From now on we do not have to distinguish the cases anymore.

We will now proceed approximately as in Lemma A.1. By induction we show the existence of a sequence $\{y_1, y_2, \dots, y_n\} \in L^1(D) \cap C^\infty(\bar{D})$ with the properties

(i) $\|y_i\|_{L^1} \leq 1$ for all i;

(ii) $\langle Ay_i, Ay_j \rangle = 0$ if $i \neq j$;

(iii) $\|Ay_i\|_{F_1} \geq 2^{3i}$ for all i.

Simultaneously we will construct a sequence of tuples $(h_i, k_i) \in X \cap C^\infty(\bar{D}) \times X \cap C^\infty(\bar{D})$, $i = 1, \dots, n$, such that

(1) $\|h_i\|_{L^\infty} + \|dh_i\|_{L^{2,1}} + \|k_i\|_{L^\infty} + \|dk_i\|_{L^{2,1}} \leq 1$;

(2) $dh_i \wedge dk_i = y_i + R_i$ with $\|R_i\|_{L^2} \leq 1$;

(3) $\|dh_i\|_{L^2} + \|dk_i\|_{L^2} \leq \left(1 + \sum_{j<i} \|dh_j\|_{L^\infty} + \|dk_j\|_{L^\infty}\right)^{-1}$.

We start the induction by choosing (a_1, b_1) in the contradicting sequence such that $\|A(da_1 \wedge db_1)\| > 2^2$. We set $y_1 = da_1 \wedge db_1$ and $(h_1, k_1) = (a_1, b_1)$. All properties are clearly satisfied ($R_1 = 0$).

Now suppose that we have chosen $y_i, (h_i, k_i)$ for $i = 1, \dots, n$. We want to construct y_{n+1} and the tuple (h_{n+1}, k_{n+1}). As in the previous lemma we define the projection operators

$$P_n := \sum_{i=1}^n \frac{Ay_i \otimes Ay_i}{\|Ay_i\|^2}, \qquad Q_n := (P_n A)^{-1}(P_n A).$$

Here $(P_n A)^{-1}$ denotes as before the inverse of $(P_n A)$ if restricted to the space $\operatorname{span}\{y_1, \dots, y_n\}$. Hence for all $x \in L^1(D)$ we have $Q_n x = \sum_{i=1}^n \alpha_i y_i$ and the existence of a constant $C_n > 0$ such that $\sum_{i=1}^n |\alpha_i| \leq C_n$ for all $x \in L^1(D)$ with $\|x\|_{L^1} \leq 1$. Furthermore due to the properties of the contradicting sequence, there exists $m \in \mathbb{N}$ such that

$$\|A(I - Q_n) da_m \wedge db_m\|_{H^1} \geq 2^{3(n+1)} C_n^2 \left(n + 3 + \sum_{j \leq n} \|dh_j\|_{L^\infty} + \|dk_j\|_{L^\infty}\right)^2.$$

Let $Q_n da_m \wedge db_m = \sum_{i=1}^n \alpha_i y_i$, and define the elements

$$\tilde{y}_{n+1} := (I - Q_n) da_m \wedge db_m, \quad \tilde{h}_{n+1} := a_m - \sum_{i=1}^n \alpha_i h_i, \quad \tilde{k}_{n+1} := b_m + \sum_{i=1}^n k_i.$$

We calculate

$$\begin{aligned} d\tilde{h}_{n+1} \wedge d\tilde{k}_{n+1} &= da_m \wedge db_m - \sum_{i=1}^n (\alpha_i dh_i \wedge dk_i) \\ &\quad + \underbrace{d\Big(-\sum_{i=1}^n \alpha_i h_i\Big) \wedge db_m}_{(I)} + \underbrace{da_m \wedge d\Big(\sum_{i=1}^n k_i\Big)}_{(II)} \underbrace{- \sum_{i<j} (\alpha_i dh_i \wedge dk_j + \alpha_j dh_j \wedge dk_i)}_{(III)} \\ &\overset{(2)}{=} da_m \wedge db_m - \sum_{i=1}^n \alpha_i y_i - \sum_{i=1}^n \alpha_i R_i + (I) + (II) + (III). \end{aligned}$$

We estimate the size of the remainder terms in $L^2(D)$: Due to (2), we have $\left\|\sum_{i=1}^n \alpha_i R_i\right\|_{L^2} \leq C_n$. The terms (I), (II) are similarly estimated by

$$\left\| d\Big(-\sum_{i=1}^n \alpha_i h_i\Big) \wedge db_m \right\|_{L^2} \leq \Big(\sum_{i=1}^n |\alpha_i| \|dh_i\|_{L^\infty}\Big) \|db_m\|_{L^2},$$

$$\left\| da_m \wedge d\Big(\sum_{i=1}^n k_i\Big) \right\|_{L^2} \leq \Big(\sum_{i=1}^n \|dk_i\|_{L^\infty}\Big) \|da_m\|_{L^2}.$$

Adding both we obtain $\|(I)\|_{L^2} + \|(II)\|_{L^2} \leq C_n\big(1 + \sum_{j\leq n} \|dh_j\|_{L^\infty} + \|dk_j\|_{L^\infty}\big)$. The last term can be estimated using only property (3) by

$$\begin{aligned} \|(III)\|_{L^2} &\leq \sum_{i=1}^n |\alpha_i|\, \|dh_i\|_{L^2} \Big(\sum_{j<i} \|dk_j\|_{L^\infty}\Big) + \|dk_i\|_{L^2} \Big(\sum_{j<i} |\alpha_j|\, \|dh_j\|_{L^\infty}\Big) \\ &\leq \Big(\sum_{i=1}^n |\alpha_i|\Big) + \sup_{j\leq n} |\alpha_j| n \leq (n+1) C_n. \end{aligned}$$

Hence we found that $\|\widetilde{R}_{n+1}\|_{L^2} \leq C_n\big(n+3+\sum_{j\leq n} \|dh_j\|_{L^\infty} + \|dk_j\|_{L^\infty}\big)$, where $\widetilde{R}_{n+1} = -\sum_{i=1}^n \alpha_i R_i + (I) + (II) + (III)$ and

$$d\tilde{h}_{n+1} \wedge d\tilde{k}_{n+1} = (I - Q_n) da_m \wedge db_m + \widetilde{R}_{n+1} = \tilde{y}_{n+1} + \widetilde{R}_{n+1}.$$

The desired functions are now simply

$$y_{n+1} = \frac{\tilde{y}_{n+1}}{\lambda_n}, \quad h_{n+1} = \frac{\tilde{h}_{n+1}}{\lambda_n}, \quad k_{n+1} = \frac{\tilde{k}_{n+1}}{\lambda_n}, \quad \text{with } \lambda_n = C_n\Big(n+3+\sum_{j\leq n} \|dh_j\|_{L^\infty} + \|dk_j\|_{L^\infty}\Big).$$

Having established the existence of the sequences y_i, h_i, k_i with the claimed properties we construct $a, b \in X$ and a sequence $f_n \in H^1(D) = F_1$ very much as in the proof of Lemma A.1: Due to (1) and the fact that X is a complete Banach space we can define elements

$$a := \sum_{i=1}^{\infty} 2^{-i} h_i, \quad b := \sum_{i=1}^{\infty} 2^{-i} k_i.$$

Furthermore for each $n \in \mathbb{N}$ let

$$f_n := \sum_{i=1}^{n} 2^{-i} \frac{Ay_i}{\|Ay_i\|_{H^1}}.$$

Observe that f_n is a finite sum of C^1-functions; hence it is C^1 and can therefore be considered as an element of $(L^1)^* = L^\infty$. It remains to check that $\lim_{n\to\infty} \int_D f_n A(da \wedge db) = +\infty$. We have

$$A(da \wedge db) = \lim_{m\to\infty} \sum_{i=1}^{m} 2^{-2i} A(dh_i \wedge dk_i) + \sum_{i<j}^{m} 2^{-i-j} A(dh_i \wedge dk_j + dh_j \wedge dk_i).$$

Using (2) we estimate

$$\left\langle \frac{Ay_k}{\|Ay_k\|_{H^1}}, A(dh_i \wedge dk_i) \right\rangle = \left\langle \frac{Ay_k}{\|Ay_k\|_{H^1}}, Ay_i + AR_i \right\rangle \geq \delta_{ki} \|Ay_i\|_{H^1} - C_A \|R_i\|_{L^2} \geq \delta_{ki} \|Ay_i\|_{H^1} - C_A.$$

Hence

$$\sum_{i=1}^{m} 2^{-2i} \left\langle \frac{Ay_k}{\|Ay_k\|_{H^1}}, A(dh_i \wedge dk_i) \right\rangle \geq 2^{-2k} \|Ay_k\|_{H^1} - \lim_{m\to\infty} \sum_{i=1}^{m} 2^{-2i} C_A \geq 2^k - C_A.$$

Using (3) we get

$$\begin{aligned} \sum_{i<j}^{m} 2^{-i-j} \|A(dh_i \wedge dk_j + dh_j \wedge dk_i)\|_{H^1} &\leq C_A \sum_{i<j}^{m} 2^{-i-j} (\|dh_i\|_{L^2} \|dk_j\|_{L^\infty} + \|dh_j\|_{L^\infty} \|dk_i\|_{L^2}) \\ &\leq C_A \sum_{i=1}^{m} 2^{-i} 2 \leq 2C_A. \end{aligned}$$

Finally combining both we obtain

$$\left\langle \frac{Ay_k}{\|Ay_k\|_{H^1}}, A(da \wedge db) \right\rangle \geq 2^k - 3C_A.$$

This completes the estimate since

$$\int_D f_n A(da \wedge db) = \sum_{k=1}^{n} 2^{-k} \left\langle \frac{Ay_k}{\|Ay_k\|_{H^1}}, A(da \wedge db) \right\rangle \geq n - 3C_A. \qquad \square$$

Acknowledgments

First of all the author thanks Vincent Millot for proposing the consideration of Möbius transformations and in this way improving the result. Furthermore he thanks Guido De Philippis for various useful discussions and he wants to thank the anonymous referees for carefully reading an earlier version of this manuscript, and for their valuable comments and suggestions. The author is supported by the MIUR SIR-grant Geometric Variational Problems (RBSI14RVEZ).

References

[Ancona 2009] A. Ancona, "Elliptic operators, conormal derivatives and positive parts of functions", *J. Funct. Anal.* **257**:7 (2009), 2124–2158. MR Zbl

[Colding and Minicozzi 2008] T. H. Colding and W. P. Minicozzi, II, "Width and finite extinction time of Ricci flow", *Geom. Topol.* **12**:5 (2008), 2537–2586. MR Zbl

[Da Lio and Palmurella 2017] F. Da Lio and F. Palmurella, "Remarks on Neumann boundary problems involving Jacobians", *Comm. Partial Differential Equations* **42**:10 (2017), 1497–1509. MR Zbl

[Gilbarg and Trudinger 1998] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, Grundlehren der Mathematischen Wissenschaften **224**, Springer, 1998.

[Grafakos 2014] L. Grafakos, *Classical Fourier analysis*, 3rd ed., Graduate Texts in Mathematics **249**, Springer, 2014. MR Zbl

[Hélein 2002] F. Hélein, *Harmonic maps, conservation laws and moving frames*, 2nd ed., Cambridge Tracts in Mathematics **150**, Cambridge University Press, 2002. MR Zbl

[Lamm and Lin 2013] T. Lamm and L. Lin, "Estimates for the energy density of critical points of a class of conformally invariant variational problems", *Adv. Calc. Var.* **6**:4 (2013), 391–413. MR Zbl

[Littman et al. 1963] W. Littman, G. Stampacchia, and H. F. Weinberger, "Regular points for elliptic equations with discontinuous coefficients", *Ann. Scuola Norm. Sup. Pisa* (3) **17** (1963), 43–77. MR Zbl

[Rivière 2008] T. Rivière, "Analysis aspects of Willmore surfaces", *Invent. Math.* **174**:1 (2008), 1–45. MR Zbl

[Semmes 1994] S. Semmes, "A primer on Hardy spaces, and some remarks on a theorem of Evans and Müller", *Comm. Partial Differential Equations* **19**:1-2 (1994), 277–319. MR Zbl

[Topping 1997] P. Topping, "The optimal constant in Wente's L^∞ estimate", *Comment. Math. Helv.* **72**:2 (1997), 316–328. MR Zbl

[Wente 1971] H. C. Wente, "An existence theorem for surfaces of constant mean curvature", *Bull. Amer. Math. Soc.* **77** (1971), 200–202. MR Zbl

Received 19 Jul 2017. Revised 15 Apr 2018. Accepted 23 Jul 2018.

JONAS HIRSCH: jonas.hirsch@sissa.it
Scuola Internazionale Superiore di Studi Avanzati, Trieste, Italy

ANALYSIS AND PDE
Vol. 12, No. 4, 2019
dx.doi.org/10.2140/apde.2019.12.1065

GLOBAL GEOMETRY AND C^1 CONVEX EXTENSIONS OF 1-JETS

DANIEL AZAGRA AND CARLOS MUDARRA

Let E be an arbitrary subset of $\mathbb{R}^n$ (not necessarily bounded) and $f : E \to \mathbb{R}$, $G : E \to \mathbb{R}^n$ be functions. We provide necessary and sufficient conditions for the 1-jet (f, G) to have an extension $(F, \nabla F)$ with $F : \mathbb{R}^n \to \mathbb{R}$ convex and C^1. Additionally, if G is bounded we can take F so that $\mathrm{Lip}(F) \lesssim \|G\|_\infty$. As an application we also solve a similar problem about finding convex hypersurfaces of class C^1 with prescribed normals at the points of an arbitrary subset of $\mathbb{R}^n$.

1. Introduction and main results

This paper concerns the following problem.

Problem 1.1. *Given $\mathcal{C}$ a differentiability class in $\mathbb{R}^n$, E a subset of $\mathbb{R}^n$, and functions $f : E \to \mathbb{R}$ and $G : E \to \mathbb{R}^n$, how can we decide whether there is a* ***convex*** *function $F \in \mathcal{C}$ such that $F(x) = f(x)$ and $\nabla F(x) = G(x)$ for all $x \in E$?*

This is a natural question which we solved in [Azagra and Mudarra 2017] in the case that $\mathcal{C} = C^{1,\omega}(\mathbb{R}^n)$, where $\omega : [0, \infty) \to [0, \infty)$ is a (strictly increasing and concave) modulus of continuity. A necessary and sufficient condition is that there exists a constant $M > 0$ such that

$$f(x) \geq f(y) + \langle G(y), x - y\rangle + |G(x) - G(y)|\omega^{-1}\Big(\frac{1}{2M}|G(x) - G(y)|\Big) \quad \text{for all } x, y \in E. \qquad (CW^{1,\omega})$$

Very recently, some explicit formulas for such extensions were found in [Daniilidis et al. 2018] for the $C^{1,1}$ case, and more generally in [Azagra et al. 2018] for the $C^{1,\omega}$ case when ω is a modulus of continuity with the additional property that $\omega(\infty) = \infty$; in particular this includes all the Hölder differentiability classes $C^{1,\alpha}$ with $\alpha \in (0, 1]$. Moreover, it can be arranged that

$$\sup_{x,y\in\mathbb{R}^n,\, x\neq y} \frac{|\nabla F(x) - \nabla F(y)|}{\omega(|x - y|)} \leq 8M$$

(or even $\mathrm{Lip}(\nabla F) \leq M$ in the $C^{1,1}$ case, that is to say, when $\omega(t) = t$).

Besides the very basic character of Problem 1.1, there are other reasons for wanting to solve this kind of problem, as extension techniques for convex functions have natural applications in analysis, differential geometry, PDE theory (in particular Monge–Ampère equations), economics, and quantum computing. See

Azagra was partially supported by Ministerio de Educación, Cultura y Deporte, Programa Estatal de Promoción del Talento y su Empleabilidad en I+D+i, Subprograma Estatal de Movilidad. Mudarra was supported by Programa Internacional de Doctorado Fundación La Caixa–Severo Ochoa. Both authors were partially supported by grant MTM2015-65825-P.
MSC2010: 26B05, 26B25, 52A20.
Keywords: convex function, C^1 function, Whitney extension theorem, global differential geometry, differentiable function.

the introductions of [Azagra and Mudarra 2017; Ghomi 2001; Yan 2014] for background about convex extensions problems, and see [Brudnyi and Shvartsman 2001; Fefferman 2005; 2006; 2009; Fefferman et al. 2016; 2017; Glaeser 1958; Jiménez-Sevilla and Sánchez-González 2013; Le Gruyer 2009] for information about general Whitney extension problems.

Let $C^1_{\text{conv}}(\mathbb{R}^n)$ stand for the set of all functions $f : \mathbb{R}^n \to \mathbb{R}$ which are convex and of class C^1. In [Azagra and Mudarra 2017], and for the class $\mathcal{C} = C^1(\mathbb{R}^n)$, we could only obtain a solution to Problem 1.1 in the particular case that E is a compact set. In this special situation the three necessary and sufficient conditions on (f, G) that we obtained for $C^1_{\text{conv}}(\mathbb{R}^n)$ extendibility are

$$G \text{ is continuous, } \quad \text{and} \quad \lim_{|z-y|\to 0^+} \frac{f(z) - f(y) - \langle G(y), z-y\rangle}{|z-y|} = 0 \text{ uniformly on } E \tag{W^1}$$

(which is equivalent to Whitney's classical condition for C^1 extendibility),

$$f(x) - f(y) \geq \langle G(y), x-y\rangle \quad \text{for all } x, y \in E \tag{C}$$

(which ensures convexity), and

$$f(x) - f(y) = \langle G(y), x-y\rangle \quad \Longrightarrow \quad G(x) = G(y), \qquad \text{for all } x, y \in E \tag{CW^1}$$

(which tells us that if two points of the graph of f lie on a line segment contained in a hyperplane which we want to be tangent to the graph of an extension at one of the points, then our putative tangent hyperplanes at both points must be the same). In fact, it is easy to see [Azagra and Mudarra 2017, Remark 1.9] that continuity of G plus conditions (C) and (CW^1) imply Whitney's condition (W^1).

In [Azagra and Mudarra 2017] we also gave examples showing that the above conditions are no longer sufficient when E is not compact (even if E is an unbounded convex body). The reasons for this insufficiency can be mainly classified into two kinds of difficulties that only arise if the set E is unbounded and G is not uniformly continuous on E:

(1) There may be no convex extension of the jet (f, G) to the whole of $\mathbb{R}^n$.

(2) Even when there are convex extensions of (f, G) defined on all of $\mathbb{R}^n$, and even when some of these extensions are differentiable in some neighborhood of E, there may be no $C^1(\mathbb{R}^n)$ convex extension of (f, G).

The aim of this paper is to show how one can overcome these difficulties by adding new necessary conditions to (W^1), (C), (CW^1) in order to obtain a complete solution to Problem 1.1 for the case that $\mathcal{C} = C^1(\mathbb{R}^n)$.

As is perhaps inevitable, our solution to Problem 1.1 contains several technical conditions which may be quite difficult to grasp at a first reading. For this reason we will reverse the logical order of the exposition: we will start by providing some corollaries and examples. Only after this will the main theorem be stated.

The first kind of complication we have mentioned is well understood thanks to [Schulz and Schwartz 1979], and is not difficult to deal with: the requirement that

$$\lim_{k\to\infty} \frac{\langle G(x_k), x_k\rangle - f(x_k)}{|G(x_k)|} = +\infty \quad \text{for every sequence } (x_k)_k \subset E \text{ with } \lim_{k\to\infty} |G(x_k)| = +\infty \tag{EX}$$

guarantees that there exist convex functions $\varphi : \mathbb{R}^n \to \mathbb{R}$ such that $\varphi_{|_E} = f$. In fact the extension φ can be determined as the minimal convex extension of the jets, i.e., $\varphi(z) = \sup_{x\in E}\{f(x) + \langle G(x), z - x\rangle\}$, and so that one has $G(x) \in \partial\varphi(x)$ (the subdifferential of φ at x). See Lemma 4.3 below.

The second kind of difficulty, however, is of a subtler geometrical character, and is related, on the one hand, to the rigid global behavior of convex functions (see Theorem 1.11 below) and, on the other hand, to the fact that a differentiable (or even real-analytic) convex function $f : \mathbb{R}^n \to \mathbb{R}$ may have what one can call *corners at infinity*. As we indicated in (2) above, there are examples of data (E, f, G) which satisfy (EX), (W^1), (C), and (CW^1) but which do not admit $C^1(\mathbb{R}^2)$ convex extensions. A prototypical instance is given in Example 1.9(4) below, or by the variant that we next formulate.

Example 1.2. Consider $E = \{(x, y) \in \mathbb{R}^2 : y \le \log|x|\} \cup \{(x, y) \in \mathbb{R}^2 : |x| \ge 1\}$, $f(x, y) = |x|$, $G(x, y) = (-1, 0)$ if $x < 0$, $G(x, y) = (1, 0)$ if $x > 0$. Since the convex function $(x, y) \mapsto |x|$ and its derivative are smooth in $\mathbb{R}^2 \setminus \{x = 0\}$ and extend the jet (f, G) from E to this region, it is clear that (f, G) satisfies (EX), (W^1), (C), and (CW^1) on E. We claim that there is no C^1 convex extension of (f, G) to $\mathbb{R}^2$. The quickest and easiest way to prove this claim is just to apply Theorem 1.8 below. An intuitive geometrical argument that can be made rigorous is that in order to allow the *corner at infinity* that this jet has along the line $x = 0$, any $C^1_{\mathrm{conv}}(\mathbb{R}^2)$ extension of this datum would have to be *essentially coercive* in the direction of this line, but the requirement that $f(x, y) = |x|$ for $|x| \ge 1$ precludes the existence of any such extension. However, it is interesting to note that if we replace E with $C = \{(x, y) \in \mathbb{R}^2 : y \le \log|x|\}$ then the situation changes completely: there are $C^1_{\mathrm{conv}}(\mathbb{R}^2)$ extensions of (f, G) from C to $\mathbb{R}^2$.

In a short while we will be giving a precise meaning to the expressions *corner at infinity* and *essentially coercive*, but let us first ask ourselves this question: what would appear to be a natural generalization of condition (CW^1) to the noncompact setting? In the absence of compactness it is natural to try to replace points by sequences, so as a first guess one is tempted to consider the following condition: if $(x_k)_k$, $(z_k)_k$ are sequences in E then

$$\lim_{k\to\infty} (f(x_k) - f(z_k) - \langle G(z_k), x_k - z_k\rangle) = 0 \quad\implies\quad \lim_{k\to\infty} |G(x_k) - G(z_k)| = 0. \tag{1-1}$$

Unfortunately, if E is unbounded, this condition is not necessary for the existence of a convex function $F \in C^1(\mathbb{R}^n)$ such that $(F, \nabla F) = (f, G)$ on E, as the following example shows.

Example 1.3. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x, y) = \sqrt{x^2 + e^{-2y}}$. This is a real-analytic strictly convex function on $\mathbb{R}^2$ (one can easily check that the Hessian $D^2 f$ is strictly positive everywhere). We have

$$\nabla f(x, y) = \left(\frac{x}{\sqrt{x^2 + e^{-2y}}}, -\frac{e^{-2y}}{\sqrt{x^2 + e^{-2y}}}\right),$$

and by considering the sequences

$$z_k = \left(\frac{1}{k}, k\right), \qquad x_k = (0, k),$$

one easily sees that

$$\lim_{k\to\infty} (f(x_k) - f(z_k) - \langle \nabla f(z_k), x_k - z_k\rangle) = 0,$$

and yet we have

$$\lim_{k\to\infty} |\nabla f(x_k) - \nabla f(z_k)| = 1 \neq 0$$

(according to Definition 1.5 below this means that the jet $(f, \nabla f)$ has a *corner at infinity* along the line $x = 0$). So our first guess turned out to be wrong, and we have to be more careful. In view of the above example, and at least if we are looking for extensions $(F, \nabla F)$ with $F \in C^1(\mathbb{R}^n)$ convex and essentially coercive (that is, C^1 convex extensions $F(x)$ which, up to a linear perturbation, tend to ∞ as $|x|$ goes to infinity), it could make sense to restrict condition (1-1) to sequences $(x_k)_k$ which are bounded. On the other hand, if $(G(z_k))_k$ is not bounded as well, then by using condition (EX), up to extracting a subsequence, we would have

$$\lim_{k\to\infty} \frac{\langle G(z_k), z_k\rangle - f(z_k)}{|G(z_k)|} = \infty;$$

hence

$$\langle G(z_k), z_k\rangle - f(z_k) = M_k|G(z_k)|, \quad \text{with } \lim_{k\to\infty} M_k = \infty,$$

and it follows that

$$\begin{aligned} f(x_k) - f(z_k) - \langle G(z_k), x_k - z_k\rangle &= f(x_k) - f(z_k) + \langle G(z_k), z_k\rangle - \langle G(z_k), x_k\rangle \\ &\geq f(x_k) + (M_k - |x_k|)|G(z_k)| \to \infty \end{aligned}$$

(because $(f(x_k))_k$ and $(x_k)_k$ are bounded and $M_k \to \infty$). Thus we have learned that we cannot have

$$\lim_{k\to\infty} (f(x_k) - f(z_k) - \langle G(z_k), x_k - z_k\rangle) = 0$$

unless $(G(x_k))_k$ is bounded. An educated guess for a good substitute of (CW^1) could then be to require that

$$\begin{aligned} &\lim_{k\to\infty} (f(x_k) - f(z_k) - \langle G(z_k), x_k - z_k\rangle) = 0 \quad \Longrightarrow \quad \lim_{k\to\infty} |G(x_k) - G(z_k)| = 0 \\ &\text{for all sequences } (x_k)_k \text{ and } (z_k)_k \text{ in } E \text{ such that } (x_k)_k \text{ and } (G(z_k))_k \text{ are bounded.} \end{aligned} \tag{1-2}$$

This new condition can be checked to be necessary for the existence of a function F which solves our problem. Now, if we add (1-2) to (EX) and (C), will this new set of conditions be sufficient as well? The answer to this question depends on how large the set $\operatorname{span}\{G(x) - G(y) : x, y \in E\}$ is. If this set coincides with $\mathbb{R}^n$ then those conditions are sufficient: this is the content of the following easy[1] consequence of the main result of this paper. On the other hand, if we do not have $\operatorname{span}\{G(x) - G(y) : x, y \in E\} = \mathbb{R}^n$ then we already know by Example 1.2 that our problem will be more difficult to tackle.

Corollary 1.4. *Fix an arbitrary subset E of $\mathbb{R}^n$ and two functions $f : E \to \mathbb{R}$, $G : E \to \mathbb{R}^n$. Suppose that* $\operatorname{span}\{G(x) - G(y) : x, y \in E\} = \mathbb{R}^n$. *Then there exists a convex function $F : \mathbb{R}^n \to \mathbb{R}$ of class C^1 with $F_{|_E} = f$ and $(\nabla F)_{|_E} = G$ if and only if the following conditions are satisfied*:

(i) *G is continuous and $f(x) \geq f(y) + \langle G(y), x - y\rangle$ for all $x, y \in E$.*

[1] But especially useful, as for generic initial data (E, f, G) one has $\operatorname{span}\{G(x) - G(y) : x, y \in E\} = \mathbb{R}^n$.

(ii) *If* $(x_k)_k \subset E$ *is a sequence for which* $\lim_{k\to\infty} |G(x_k)| = +\infty$, *then*

$$\lim_{k\to\infty} \frac{\langle G(x_k), x_k\rangle - f(x_k)}{|G(x_k)|} = +\infty.$$

(iii) *If* $(x_k)_k$, $(z_k)_k$ *are sequences in* E *such that* $(x_k)_k$ *and* $(G(z_k))_k$ *are bounded, and*

$$\lim_{k\to\infty} \left(f(x_k) - f(z_k) - \langle G(z_k), x_k - z_k\rangle\right) = 0,$$

then $\lim_{k\to\infty} |G(x_k) - G(z_k)| = 0$.

Moreover, whenever these conditions are satisfied, the extension F can be taken to be ***essentially coercive***.

Here, by saying that F is *essentially coercive* we mean that there exists a linear function $\ell : \mathbb{R}^n \to \mathbb{R}$ such that

$$\lim_{|x|\to\infty} \left(F(x) - \ell(x)\right) = \infty.$$

Let us mention that the above corollary is applied in [Azagra and Hajłasz 2017] to show that a convex function $f : \mathbb{R}^n \to \mathbb{R}$ has a Lusin property of type $C^1_{\text{conv}}(\mathbb{R}^n)$ (meaning that for every $\varepsilon > 0$ there exists a convex function $g \in C^1(\mathbb{R}^n)$ such that $\mathcal{L}^n(\{x \in \mathbb{R}^n : f(x) \neq g(x)\}) < \varepsilon$, where $\mathcal{L}^n$ denotes Lebesgue's measure) if and only if either f is essentially coercive or else f is already C^1 (in which case taking $g = f$ is the only possible option).

In order to quickly understand Corollary 1.4, instead of looking at the rather technical proof of Theorem 1.13 below we recommend reading the proof of [Azagra and Mudarra 2017, Theorem 1.10], which can be easily adapted to produce a simpler proof of Corollary 1.4.

By comparing Examples 1.2 and 1.3 with Corollary 1.4 we may arrive at a remarkable conclusion: our given jet (f, G) may well have some *corners at infinity* and, for C^1 convex extension purposes, that will not matter at all as long as (f, G) forces all possible convex extensions to be essentially coercive (equivalently, as long as $\operatorname{span}\{G(x) - G(y) : x, y \in E\} = \mathbb{R}^n$). But, if the given datum presents some corners at infinity and does not force essential coercive in the directions of those corners, then we will have to be more careful, as $C^1_{\text{conv}}(\mathbb{R}^n)$ extensions may not exist in this case.

Let us now explain what we mean by a jet having a corner at infinity.

Definition 1.5. Let X be a proper linear subspace of $\mathbb{R}^n$ and let us denote by $X^\perp$ its orthogonal complement. We say that a jet $(f, G) : E \subset \mathbb{R}^n \to \mathbb{R} \times \mathbb{R}^n$ has a corner at infinity in a direction of $X^\perp$ provided that there exist two sequences $(x_k)_k$, $(z_k)_k$ in E such that, if $P_X : \mathbb{R}^n \to X$ denotes the orthogonal projection, we have $(P_X(x_k))_k$ and $(G(z_k))_k$ are bounded, $\lim_{k\to\infty} |x_k| = \infty$,

$$\lim_{k\to\infty} \left(f(x_k) - f(z_k) - \langle G(z_k), x_k - z_k\rangle\right) = 0,$$

and yet

$$\limsup_{k\to\infty} |G(x_k) - G(z_k)| > 0.$$

We will also say that the jet (f, G) has a corner at infinity in the direction of the line $\{tv : t \in \mathbb{R}\}$ (where $v \in \mathbb{R}^n \setminus \{0\}$) provided that there exist sequences $(x_k)_k$, $(z_k)_k$ satisfying the above properties with P_X being the orthogonal projection onto the hyperplane X perpendicular to v.

For instance, the function $f(x, y)$ of Example 1.3, and its gradient, when restricted to the sequences $(x_k)_k$, $(z_k)_k$ defined there, give an instance of a jet that has a corner at infinity directed by the line $x = 0$. Of course, the pair $(f, \nabla f)$, unrestricted, provides another instance. In this case it is natural to say that the function f itself has a corner at infinity. More pathological examples can be given in higher dimensions: for instance, if $1 \leq k \leq n$ then

$$f(x_1, \ldots, x_n) = \sqrt{\sum_{j=1}^{k} x_j^2 + \sum_{j=k+1}^{n} e^{-2x_j}}, \quad x = (x_1, \ldots, x_n) \in \mathbb{R}^n, \tag{1-3}$$

is a convex function of class C^∞ with strictly positive Hessian at every point, which has a corner at infinity in the direction of e_j for every $j = k+1, \ldots, n$, and which is essentially coercive. On the other hand, if $n \geq 3$ and $2 \leq k < n$, then

$$g(x_1, \ldots, x_n) = \sqrt{x_1^2 + \sum_{j=2}^{k} e^{-2x_j}}, \quad x = (x_1, \ldots, x_n) \in \mathbb{R}^n, \tag{1-4}$$

is convex and of class C^∞, g has a corner at infinity in the direction of e_j for every $j = 2, \ldots, k$, and g is not essentially coercive. Nevertheless g is *essentially k-coercive* (meaning that g can be written as $g = c \circ P$, where P is the orthogonal projection onto a k-dimensional subspace of X of $\mathbb{R}^n$ and $c : X \to \mathbb{R}$ is essentially coercive).

In general it can be shown (and in fact this is a consequence of our main results) that the presence of a corner at infinity in the graph of a differentiable convex function $f : \mathbb{R}^n \to \mathbb{R}$ forces essential k-coercivity of f, for some $k \geq 2$, in a subspace of directions containing the directions of the corner.

We will not explicitly use the notion of corner at infinity in our proofs. Our reasons for introducing these objects are the facts that: (1) one way or another, corners at infinity will be to blame for most of the predicaments and technicalities involved in any attempt to solve Problem 1.1 for $\mathcal{C} = C^1(\mathbb{R}^n)$; and (2) we firmly believe that the reader will be more able to understand the statements and proofs of the following results once he has been acquainted with this notion. As a matter of fact, the most technical conditions of Theorems 1.8 and 1.13 below can be rephrased more intuitively in terms of corners at infinity and essential coercivity of data in the directions of those corners.

Unfortunately Corollary 1.4 does not provide a characterization of the 1-jets which admit essentially coercive C^1 convex extensions. This is due to the fact that a jet (f, G) defined on a set E may admit such an extension and yet $\operatorname{span}\{G(x) - G(y) : x, y \in E\} \neq \mathbb{R}^n$, as shown by the trivial example of the jet (f_0, G_0) with $E_0 = \{0\} \subset \mathbb{R}^2$, $f_0(0) = 0$, $G_0(0) = 0$, which admits a C^1 convex and coercive extension given by $(F_0, \nabla F_0)$, where $F_0(x, y) = x^2 + y^2$.

Of course, a C^1 convex extension problem for a given 1-jet (f, G) may have solutions which are not essentially coercive; in fact it may happen that none of its solutions are essentially coercive. A sister of Corollary 1.4 which provides a more general, but still partial solution to Problem 1.1 is the following.

Corollary 1.6. *Given an arbitrary subset E of $\mathbb{R}^n$ and two functions $f : E \to \mathbb{R}$, $G : E \to \mathbb{R}^n$, assume that the following conditions are satisfied*:

(i) *G is continuous and $f(x) \geq f(y) + \langle G(y), x-y\rangle$ for all $x, y \in E$.*

(ii) *If $(x_k)_k \subset E$ is a sequence for which $\lim_{k\to\infty} |G(x_k)| = +\infty$, then*
$$\lim_{k\to\infty} \frac{\langle G(x_k), x_k\rangle - f(x_k)}{|G(x_k)|} = +\infty.$$

(iii) *Let $P = P_Y : \mathbb{R}^n \to \mathbb{R}^n$ be the orthogonal projection onto $Y := \operatorname{span}\{G(x) - G(y) : x, y \in E\}$. If $(x_k)_k$, $(z_k)_k$ are sequences in E such that $(P(x_k))_k$ and $(G(z_k))_k$ are bounded and*
$$\lim_{k\to\infty} (f(x_k) - f(z_k) - \langle G(z_k), x_k - z_k\rangle) = 0,$$
then $\lim_{k\to\infty} |G(x_k) - G(z_k)| = 0$.

Then there exists a convex function $F : \mathbb{R}^n \to \mathbb{R}$ of class C^1 such that $F_{|_E} = f$ and $(\nabla F)_{|_E} = G$.

Condition (iii) of the above corollary can be intuitively rephrased by saying that: (1) our jet satisfies a natural generalization of condition (CW^1); and (2) (f, G) cannot have corners at infinity in any direction contained in the orthogonal complement of the subspace $Y = \operatorname{span}\{G(x) - G(y) : x, y \in E\}$.

It could be natural to hope for the conditions of Corollary 1.6 to be necessary as well, thus providing a nice characterization of those 1-jets which admit C^1 convex extensions. Unfortunately the solution to Problem 1.1 is necessarily more complicated, as the following example shows.

Example 1.7. Let $E_1 = \left\{(x, y) \in \mathbb{R}^2 : y = \log|x|,\ |x| \in \mathbb{N} \cup \left\{\frac{1}{n} : n \in \mathbb{N}\right\}\right\}$, $f_1(x, y) = |x|$, $G_1(x, y) = (-1, 0)$ if $x < 0$, $G_1(x, y) = (1, 0)$ if $x > 0$. In this case we have
$$Y_1 := \operatorname{span}\{G_1(x, y) - G_1(x', y') : (x, y), (x', y') \in E_1\} = \mathbb{R} \times \{0\},$$
and it is easily seen that condition (iii) is not satisfied. However, it is not difficult to check that, for $\varepsilon > 0$ small enough, if we set $E_1^* = E_1 \cup \{(0, 1)\}$, $f_1^* = f_1$ on E_1, $f_1^*(0, 1) = \varepsilon$, $G_1^* = G_1$ on E_1, and $G_1^*(0, 1) = (0, \varepsilon)$, then Corollary 1.4 implies that the problem of finding a C^1 convex extension of the jet (f_1^*, G_1^*) does have a solution, and therefore the same is true of the jet (f_1, G_1).

This example shows that in some cases the C^1 convex extension problem for a 1-jet (f, G) may be geometrically underdetermined in the sense that we may not have been given enough differential data so as to have condition (iii) of the above corollary satisfied with $Y = \operatorname{span}\{G(x) - G(y) : x, y \in E\}$, and yet it may be possible to find a few more jets (β_j, w_j) associated to finitely many points $p_j \in \mathbb{R}^n \setminus \bar{E}$, $j = 1, \dots, m$, so that, if we define $E^* = E \cup \{p_1, \dots, p_m\}$ and extend the functions f and G from E to E^* by setting
$$f(x_j) := \beta_j, \quad G(p_j) := w_j \quad \text{for } j = 1, \dots, m, \tag{1-5}$$
then the new extension problem for (f, G) defined on E^* does satisfy condition (iii) of Corollary 1.6. Notice that, the larger Y grows, the weaker condition (iii) of Corollary 1.6 becomes.

We are now prepared to state a first version of our main result.

Theorem 1.8. *Given an arbitrary subset E of $\mathbb{R}^n$ and two functions $f : E \to \mathbb{R}$, $G : E \to \mathbb{R}^n$, the following is true. There exists a convex function $F : \mathbb{R}^n \to \mathbb{R}$ of class C^1 such that $F_{|_E} = f$, and $(\nabla F)_{|_E} = G$ if and only if the following conditions are satisfied*:

(i) *G is continuous and $f(x) \geq f(y) + \langle G(y), x-y\rangle$ for all $x, y \in E$.*

(ii) *If $(x_k)_k \subset E$ is a sequence for which $\lim_{k\to\infty} |G(x_k)| = +\infty$, then*
$$\lim_{k\to\infty} \frac{\langle G(x_k), x_k\rangle - f(x_k)}{|G(x_k)|} = +\infty.$$

(iii) *Let $Y := \operatorname{span}\{G(x) - G(y) : x, y \in E\}$. There exists a linear subspace $X \supseteq Y$ such that, either $Y = X$, or else, if we define $k = \dim Y$ and $d = \dim X$, and $P_X : \mathbb{R}^n \to \mathbb{R}^n$ is the orthogonal projection from $\mathbb{R}^n$ onto X, there exist points $p_1, \dots, p_{d-k} \in \mathbb{R}^n \setminus \overline{E}$, numbers $\beta_1, \dots, \beta_{d-k} \in \mathbb{R}$, and vectors $w_1, \dots, w_{d-k} \in \mathbb{R}^n$ such that*
   (a) $X = \operatorname{span}(\{u - v : u, v \in G(E) \cup \{w_1, \dots, w_{d-k}\}\})$,
   (b) $\beta_j > \max_{1\leq i\neq j\leq d-k}\{\beta_i + \langle w_i, p_j - p_i\rangle\}$ *for all* $1 \leq j \leq d-k$,
   (c) $\beta_j > \sup_{z\in E,\, |G(z)|\leq N}\{f(z) + \langle G(z), p_j - z\rangle\}$ *for all* $1 \leq j \leq d-k$ *and* $N \in \mathbb{N}$,
   (d) $\inf_{x\in E,\, |P_X(x)|\leq N}\{f(x) - \max_{1\leq j\leq d-k}\{\beta_j + \langle w_j, x - p_j\rangle\}\} > 0$ *for all* $N \in \mathbb{N}$.

(iv) *If X and P_X are as in* (iii), *and $(x_k)_k$, $(z_k)_k$ are sequences in E such that $(P_X(x_k))_k$ and $(G(z_k))_k$ are bounded and*
$$\lim_{k\to\infty} (f(x_k) - f(z_k) - \langle G(z_k), x_k - z_k\rangle) = 0,$$
then $\lim_{k\to\infty} |G(x_k) - G(z_k)| = 0$.

As we see, the difference between Theorem 1.8 and Corollary 1.6 is in the technical condition (iii), which can be informally summed up by saying that, whenever the jets $(f(x), G(x))$, $x \in E$, do not provide us with enough differential data so that condition (iii) of Corollary 1.6 holds, there is enough room in $\mathbb{R}^n \setminus \overline{E}$ to add finitely many new jets (β_j, w_j), associated to new points p_j, $j = 1, \dots, d-k$, in such a way that the new extension problem does satisfy the conditions of Corollary 1.6. This condition also tells us that the new extension problem will be one for which, even though there may be corners at infinity, those corners will necessarily be directed by subspaces which are contained in the span of the putative derivatives, and the new data will force essential coercivity of all possible extensions in the directions of the corners.

Later on we will show that, in the particular case that G is bounded (and so we may expect to find an F with a bounded gradient), these complicated conditions about compatibility of the old and new data admit a much nicer geometrical reformulation; see Theorem 1.14 below.

Let us consider some examples that will hopefully offer further clarification of these comments.

Example 1.9. Consider the following 1-jets (f_j, G_j) defined on subsets E_j of $\mathbb{R}^n$:

(1) $E_1 = \left\{(x, y) \in \mathbb{R}^2 : y = \log|x|,\ |x| \in \mathbb{N} \cup \left\{\frac{1}{n} : n \in \mathbb{N}\right\}\right\}$, $f_1(x, y) = |x|$, $G_1(x, y) = (-1, 0)$ if $x < 0$, $G_1(x, y) = (1, 0)$ if $x > 0$.

(2) $E_2 = \left\{(x, y) \in \mathbb{R}^2 : y = \log|x|,\ |x| \in \mathbb{N} \cup \left\{\frac{1}{n} : n \in \mathbb{N}\right\}\right\}$, $f_2 = \varphi$, $G_2 = \nabla\varphi$, where $\varphi(x, y) = \sqrt{x^2 + e^{-2y}}$.

(3) $E_3 = \left\{(x, y, z) \in \mathbb{R}^3 : z = 0,\ y = \log|x|,\ |x| \in \mathbb{N} \cup \left\{\frac{1}{n} : n \in \mathbb{N}\right\}\right\}$, $f_3 = \varphi$, $G_3 = \nabla\varphi$, where $\varphi(x, y, z) = \sqrt{x^2 + e^{-2y}}$.

(4) $E_4 = E_1 \cup \{(x, y) \in \mathbb{R}^2 : |x| \geq 1\}$, $f_4(x, y) = |x|$, $G_4(x, y) = (-1, 0)$ if $x < 0$, $G_4(x, y) = (1, 0)$ if $x > 0$.

Then one can check that:

(i) For the jet (f_1, G_1), and with the notation of Theorem 1.8, we have $Y = \mathbb{R} \times \{0\}$, but the smallest possible X we can take is $X = \mathbb{R}^2$ (and all possible extensions F must be essentially coercive on $\mathbb{R}^2$).

(ii) For the jet (f_2, G_2) we have $Y = \mathbb{R}^2$, and all possible extensions F must be essentially coercive on $\mathbb{R}^2$.

(iii) For the jet (f_3, G_3) we have $Y = \mathbb{R}^2 \times \{0\}$, and we can take either $X = Y$ or $X = \mathbb{R}^3$.

(iv) For the jet (f_4, G_4) we have $Y = \mathbb{R} \times \{0\}$, but one cannot apply Theorem 1.8 with any X. There exists no $F \in C^1_{\mathrm{conv}}(\mathbb{R}^2)$ such that $(F, \nabla F)$ extends (f_4, G_4).

Even though Theorem 1.8 fully solves Problem 1.1, an important question[2] remains open: how can we characterize those 1-jets (f, G) such that there exists an essentially coercive convex function $F \in C^1(\mathbb{R}^n)$ so that $(F, \nabla F)$ extends (f, G)? The answer is: those jets are the jets which satisfy the conditions of Theorem 1.8 with $X = \mathbb{R}^n$. More generally, one could ask for C^1 convex extensions with prescribed global behavior (meaning extensions which are essentially coercive only in some directions, and affine in others). This ties in with a question which will be extremely important in our proofs: what is the global geometrical shape of the C^1 convex extension we are trying to build?

In this regard, it will be convenient for us to state a refinement of Theorem 1.8 which characterizes the set of 1-jets admitting C^1 convex extensions with a prescribed global behavior, and which requires introducing some definitions and notation.

Definition 1.10. Let Z be a real vector space and $P : Z \to X$ be the orthogonal projection onto a subspace $X \subseteq Z$. We will say that a function f defined on a subset E of Z is *essentially P-coercive* provided that there exists a linear function $\ell : Z \to \mathbb{R}$ such that for every sequence $(x_k)_k \subset E$ with $\lim_{k\to\infty} |P(x_k)| = \infty$ one has

$$\lim_{k\to\infty} (f - \ell)(x_k) = \infty.$$

We will say that f is *essentially coercive* whenever f is essentially I-coercive, where $I : Z \to Z$ is the identity mapping.

If X is a linear subspace of $\mathbb{R}^n$, we will denote by $P_X : \mathbb{R}^n \to X$ the orthogonal projection, and we will say that $f : E \to \mathbb{R}$ is *coercive in the direction of* X whenever f is P_X-coercive.

We will also denote by $X^\perp$ the orthogonal complement of X in $\mathbb{R}^n$. For a subset V of $\mathbb{R}^n$, $\mathrm{span}(V)$ will stand for the linear subspace spanned by the vectors of V.

In [Azagra 2013] essentially coercive convex functions were called *properly convex*, and some approximation results, which fail for general convex functions, were shown to be true for this class of functions. The following result was also implicitly proved in [Azagra 2013, Lemma 4.2]. Since this will be a very important tool in the statements and proofs of all the results of the present paper, and because we have introduced new terminology and added conclusions, we will provide a self-contained proof in Section 2 for the readers' convenience.

[2]Coercivity of a convex function may be relevant or even essential to a number of possible applications, e.g., in PDE theory.

Theorem 1.11. *For every convex function $f : \mathbb{R}^n \to \mathbb{R}$ there exist a unique linear subspace X_f of $\mathbb{R}^n$, a unique vector $v_f \in X_f^{\perp}$, and a unique essentially coercive function $c_f : X_f \to \mathbb{R}$ such that f can be written in the form*

$$f(x) = c_f(P_{X_f}(x)) + \langle v_f, x\rangle \quad \text{for all } x \in \mathbb{R}^n.$$

Moreover, if Y is a linear subspace of $\mathbb{R}^n$ such that f is essentially coercive in the direction of Y, then $Y \subseteq X_f$.

The following proposition shows that the directions X_f given by these decompositions are stable by approximation.

Proposition 1.12. *With the notation of the preceding theorem, if $f, g : \mathbb{R}^n \to \mathbb{R}$ are convex functions and A is a positive number such that $f(x) \leq g(x) + A$ for all $x \in \mathbb{R}^n$, then $X_f \subseteq X_g$.*

In particular, if $|f - g| \leq A$ then $X_f = X_g$.

Proof. The inequality $f(x) \leq g(x) + A$ and the essential coercivity of f in the direction X_f implies that g is essentially coercive in the direction X_f. Then $X_f \subseteq X_g$ by the last part of Theorem 1.11. □

We are finally ready to state the announced refinement of Theorem 1.8 which characterizes precisely which 1-jets (f, G) admit extensions $(F, \nabla F)$ such that $F \in C^1_{\mathrm{conv}}(\mathbb{R}^n)$ and X_F coincides with a prescribed linear subspace X of $\mathbb{R}^n$.

Theorem 1.13. *Given an arbitrary subset E of $\mathbb{R}^n$, a linear subspace $X \subset \mathbb{R}^n$, the orthogonal projection $P := P_X : \mathbb{R}^n \to X$, and two functions $f : E \to \mathbb{R}$, $G : E \to \mathbb{R}^n$, the following is true. There exists a convex function $F : \mathbb{R}^n \to \mathbb{R}$ of class C^1 such that $F_{|_E} = f$, $(\nabla F)_{|_E} = G$, and $X_F = X$ if and only if the following conditions are satisfied:*

(i) *G is continuous and $f(x) \geq f(y) + \langle G(y), x - y\rangle$ for all $x, y \in E$.*

(ii) *If $(x_k)_k \subset E$ is a sequence for which $\lim_{k\to\infty} |G(x_k)| = +\infty$, then*

$$\lim_{k\to\infty} \frac{\langle G(x_k), x_k\rangle - f(x_k)}{|G(x_k)|} = +\infty.$$

(iii) *$Y := \mathrm{span}(\{G(x) - G(y) : x, y \in E\}) \subseteq X$.*

(iv) *If $Y \neq X$ and we define $k = \dim Y$ and $d = \dim X$, there exist points $p_1, \dots, p_{d-k} \in \mathbb{R}^n \setminus \overline{E}$, numbers $\beta_1, \dots, \beta_{d-k} \in \mathbb{R}$, and vectors $w_1, \dots, w_{d-k} \in \mathbb{R}^n$ such that*

 (a) *$X = \mathrm{span}(\{u - v : u, v \in G(E) \cup \{w_1, \dots, w_{d-k}\}\})$,*

 (b) *$\beta_j > \max_{1\leq i\neq j\leq d-k}\{\beta_i + \langle w_i, p_j - p_i\rangle\}$ for all $1 \leq j \leq d-k$,*

 (c) *$\beta_j > \sup_{z\in E,\, |G(z)|\leq N}\{f(z) + \langle G(z), p_j - z\rangle\}$ for all $1 \leq j \leq d-k$ and $N \in \mathbb{N}$,*

 (d) *$\inf_{x\in E,\, |P(x)|\leq N}\{f(x) - \max_{1\leq j\leq d-k}\{\beta_j + \langle w_j, x - p_j\rangle\}\} > 0$ for all $N \in \mathbb{N}$.*

(v) *If $(x_k)_k$, $(z_k)_k$ are sequences in E such that $(P(x_k))_k$ and $(G(z_k))_k$ are bounded and*

$$\lim_{k\to\infty} (f(x_k) - f(z_k) - \langle G(z_k), x_k - z_k\rangle) = 0,$$

then $\lim_{k\to\infty} |G(x_k) - G(z_k)| = 0$.

In particular, by considering the case that $X = \mathbb{R}^n$, we obtain a characterization of the 1-jets which admit C^1 convex extensions which are essentially coercive in $\mathbb{R}^n$.

It is clear that Theorem 1.8 and Corollaries 1.4 and 1.6 are immediate consequences of the above theorem. The proof of Theorem 1.13 will be given in Sections 3 and 4.

In the special case that the function G of the above theorem is bounded, one should expect to find Lipschitz convex functions $F \in C^1(\mathbb{R}^n)$ such that $(F, \nabla F)$ extends (f, G) and $\mathrm{Lip}(F) \lesssim \|G\|_\infty$. Notice that this kind of control of $\mathrm{Lip}(F)$ in terms of $\sup_{y\in E} |G(y)|$ solely cannot be obtained, in general, for nonconvex jets, but it is possible in the convex case, at least when E is compact; see the comments after [Azagra and Mudarra 2017, Theorem 1.10]. The next result tells us that this is indeed feasible, and moreover shows that the technical conditions of (iv) in Theorem 1.13 can be replaced (just in this Lipschitz case) by a nicer geometric condition which tells us that the complement of the closure of E in $\mathbb{R}^n$ contains the union of a certain finite collection of cones.

Theorem 1.14. *Given an arbitrary subset E of $\mathbb{R}^n$, a linear subspace $X \subset \mathbb{R}^n$, the orthogonal projection $P := P_X : \mathbb{R}^n \to X$, and two functions $f : E \to \mathbb{R}$, $G : E \to \mathbb{R}^n$, the following is true. There exists a Lipschitz convex function $F : \mathbb{R}^n \to \mathbb{R}$ of class C^1 such that $F_{|E} = f$, $(\nabla F)_{|E} = G$, and $X_F = X$ if and only if the following conditions are satisfied*:

(i) *G is continuous and bounded and $f(x) \geq f(y) + \langle G(y), x - y\rangle$ for all $x, y \in E$.*

(ii) *$Y := \mathrm{span}(\{G(x) - G(y) : x, y \in E\}) \subseteq X$.*

(iii) *If $Y \neq X$ and we define $k = \dim Y$ and $d = \dim X$, there exist points $p_1, \dots, p_{d-k} \in \mathbb{R}^n \setminus \overline{E}$, a number $\varepsilon \in (0, 1)$, and linearly independent normalized vectors $w_1, \dots, w_{d-k} \in X \cap Y^\perp$ such that, for every $j = 1, \dots, d-k$, the cone $V_j := \{x \in \mathbb{R}^n : \varepsilon\langle w_j, x - p_j\rangle \geq |P_Y(x - p_j)|\}$ does not contain any point of $\overline{E}$. Here $P_Y : \mathbb{R}^n \to Y$ denotes the orthogonal projection onto Y.*

(iv) *If $(x_k)_k$, $(z_k)_k$ are sequences in E such that $(P_X(x_k))_k$ is bounded and*

$$\lim_{k\to\infty} (f(x_k) - f(z_k) - \langle G(z_k), x_k - z_k\rangle) = 0,$$

then $\lim_{k\to\infty} |G(x_k) - G(z_k)| = 0$.

Moreover, there exists a constant $C(n) > 0$ only depending on n such that, whenever these conditions are satisfied, the extension F can be taken so that

$$\mathrm{Lip}(F) = \sup_{x\in\mathbb{R}^n} |\nabla F(x)| \leq C(n) \sup_{y\in E} |G(y)|.$$

Finally, let us turn our attention to a geometrical problem which is closely related to our results.

Problem 1.15. *Given an arbitrary subset E of $\mathbb{R}^n$ and a unitary vector field $N : E \to \mathbb{R}^n$, what conditions will be necessary and sufficient in order to guarantee the existence of a convex hypersurface M of class C^1 with the properties that $E \subset M$ and $N(x)$ is normal to M at each $x \in E$?*

Our solution to this problem is as follows. We say that a subset W of $\mathbb{R}^n$ is a (possibly unbounded) convex body provided that W is closed and convex, with nonempty interior. Assuming, as we may, that

$0 \in \operatorname{int}(W)$, we will say that W is of class C^1 provided that its Minkowski functional

$$\mu_W(x) = \inf\left\{\lambda > 0 : \frac{1}{\lambda} x \in W\right\}$$

is of class C^1 on the open set $\mathbb{R}^n \setminus \mu_W^{-1}(0)$. This is equivalent to saying that W can be locally parametrized as a graph $(x_1, \ldots, x_{n-1}, g(x_1, \ldots, x_{n-1}))$ (coordinates taken with respect to an appropriate permutation of the canonical basis of $\mathbb{R}^n$), where g is of class C^1. We will denote by

$$n_W(x) = \frac{\nabla \mu_W(x)}{|\nabla \mu_W(x)|}, \quad x \in \partial W,$$

the outer normal to ∂W.

Theorem 1.16. *Let E be an arbitrary subset of $\mathbb{R}^n$, $N : E \to \mathbb{S}^{n-1}$ a continuous mapping, X a linear subspace of $\mathbb{R}^n$, and $P : \mathbb{R}^n \to X$ the orthogonal projection. Then there exists a (possibly unbounded) convex body W of class C^1 such that $E \subset \partial W$, $0 \in \operatorname{int}(W)$, $N(x) = n_W(x)$ for all $x \in E$, and $X = \operatorname{span}(n_W(\partial W))$ if and only if the following conditions are satisfied:*

(1) $\langle N(y), x - y\rangle \leq 0$ *for all* $x, y \in E$.

(2) *For all sequences $(x_k)_k$, $(z_k)_k$ contained in E with $(P(x_k))_k$ bounded, we have*

$$\lim_{k\to\infty} \langle N(z_k), x_k - z_k\rangle = 0 \quad \Longrightarrow \quad \lim_{k\to\infty} |N(z_k) - N(x_k)| = 0.$$

(3) $0 < \inf_{y\in E} \langle N(y), y\rangle$.

(4) *Defining $d = \dim(X)$, $Y = \operatorname{span}(N(E))$, $\ell = \dim(Y)$, we have $Y \subseteq X$, and if $Y \neq X$ and $P_Y : \mathbb{R}^n \to Y$ is the orthogonal projection then there exist linearly independent normalized vectors $w_1, \ldots, w_{d-k} \in X \cap Y^{\perp}$, points $p_1, \ldots, p_{d-\ell} \in \mathbb{R}^n$, and a number $\varepsilon \in (0, 1)$ such that*

$$(\bar{E} \cup \{0\}) \cap \left(\bigcup_{j=1}^{d-\ell} V_j\right) = \varnothing,$$

where $V_j := \{x \in \mathbb{R}^n : \varepsilon\langle w_j, x - p_j\rangle \geq |P_Y(x - p_j)|\}$ for every $j = 1, \ldots, d - \ell$.

In the case that $X = \operatorname{span}(N(E))$, the preceding result takes on a much simpler form.

Corollary 1.17. *Let E be an arbitrary subset of $\mathbb{R}^n$, $N : E \to \mathbb{S}^{n-1}$ a continuous mapping, X a linear subspace of $\mathbb{R}^n$ such that $X = \operatorname{span}(N(E))$, and $P : \mathbb{R}^n \to X$ the orthogonal projection. Then there exists a (possibly unbounded) convex body W of class C^1 such that $E \subset \partial W$, $0 \in \operatorname{int}(W)$, $N(x) = n_W(x)$ for all $x \in E$, and $X = \operatorname{span}(n_W(\partial W))$ if and only if the following conditions are satisfied:*

(1) $\langle N(y), x - y\rangle \leq 0$ *for all* $x, y \in E$.

(2) *For all sequences $(x_k)_k$, $(z_k)_k$ contained in E with $(P(x_k))_k$ bounded, we have*

$$\lim_{k\to\infty} \langle N(z_k), x_k - z_k\rangle = 0 \quad \Longrightarrow \quad \lim_{k\to\infty} |N(z_k) - N(x_k)| = 0.$$

(3) $0 < \inf_{y\in E} \langle N(y), y\rangle$.

2. Proof of Theorem 1.11

Let us first recall some terminology from [Azagra 2013]. We say that a function $C:\mathbb{R}^n\to\mathbb{R}$ is a k-dimensional *corner function* on $\mathbb{R}^n$ if it is of the form

$$C(x)=\max\{\ell_1+b_1,\ \ell_2+b_2,\ \dots,\ \ell_k+b_k\},$$

where the $\ell_j:\mathbb{R}^n\to\mathbb{R}$ are linear functions such that the functions $L_j:\mathbb{R}^{n+1}=\mathbb{R}^n\times\mathbb{R}\to\mathbb{R}$ defined by $L_j(x,x_{n+1})=x_{n+1}-\ell_j(x)$, $1\le j\le k$, are linearly independent in $(\mathbb{R}^{n+1})^*$, and the b_j are in $\mathbb{R}$. This is equivalent to saying that the functions $\{\ell_2-\ell_1,\dots,\ell_k-\ell_1\}$ are linearly independent in $(\mathbb{R}^n)^*$.

We also say that a convex function $f:\mathbb{R}^n\to\mathbb{R}$ is supported by C at a point $x\in\mathbb{R}^n$ provided we have $C\le f$ and $C(x)=f(x)$.

Now let us prove Theorem 1.11.

Case 1. We will first assume that f is differentiable (and therefore of class C^1, since f is convex). If f is affine, say $f(x)=a\langle u,x\rangle+b$, then the result is trivially true with $X=\{0\}$, $c(0)=b$, and $v=au$. On the other hand, if f is essentially coercive then the result also holds obviously with $X=\mathbb{R}^n$, $v=0$, and $c=f$. So we may assume that f is neither affine nor essentially coercive. In particular there exist $x_0,y_0\in\mathbb{R}^n$ with $Df(x_0)\neq Df(y_0)$. It is then clear that $L_1(x,x_{n+1})=x_{n+1}-Df(x_0)(x)$ and $L_2(x,x_{n+1})=x_{n+1}-Df(y_0)(x)$ are two linearly independent linear functions on $(\mathbb{R}^{n+1})^*$; hence f is supported at x_0 by the two-dimensional corner $x\mapsto\max\{f(x_0)+Df(x_0)(x-x_0),\ f(y_0)+Df(y_0)(x-y_0)\}$.

Let us then define k as the greatest integer such that f is supported at x_0 by a $(k+1)$-dimensional corner. By assumption we have $1\le k<n$. Then we also have that there exist $\ell_1,\dots,\ell_{k+1}\in(\mathbb{R}^n)^*$ with $L_j(x,x_{n+1})=x_{n+1}-\ell_j(x)$, $j=1,\dots,k+1$, linearly independent in $(\mathbb{R}^{n+1})^*$, and $b_1,\dots,b_{k+1}\in\mathbb{R}$ so that $C=\max_{1\le j\le k+1}\{\ell_j+b_j\}$ supports f at x_0.

Observe that the $\{L_j-L_1\}_{j=2}^{k+1}$ are linearly independent in $(\mathbb{R}^{n+1})^*$; hence so are the $\{\ell_j-\ell_1\}_{j=2}^{k+1}$ in $(\mathbb{R}^n)^*$, and therefore $\bigcap_{j=2}^{k+1}\operatorname{Ker}(\ell_j-\ell_1)$ has dimension $n-k$. Then we can find linearly independent vectors $w_1,\dots,w_{n-k}$ such that $\bigcap_{j=2}^{k+1}\operatorname{Ker}(\ell_j-\ell_1)=\operatorname{span}\{w_1,\dots,w_{n-k}\}$.

Now, given any $y\in\mathbb{R}^n$, if

$$\frac{d}{dt}(f-\ell_1)(y+tw_q)|_{t=t_0}\neq 0$$

for some t_0 then $Df(y+t_0w_q)-\ell_1$ is linearly independent with $\{\ell_j-\ell_1\}_{j=2}^{k+1}$, which implies that $(x,x_{n+1})\mapsto x_{n+1}-Df(y+t_0w_q)$ is linearly independent with $L_1,\dots,L_{k+1}$, and therefore the function

$$x\mapsto\max\{\ell_1(x)+b_1,\dots,\ell_{k+1}(x)+b_{k+1},\ Df(y+t_0w_q)(x-y-t_0w_q)+f(y+t_0w_q)\}$$

is a $(k+2)$-dimensional corner supporting f at x_0, which contradicts the choice of k. Thus we must have

$$\frac{d}{dt}(f-\ell_1)(y+tw_q)=0\quad\text{for all } y\in\mathbb{R}^n, t\in\mathbb{R},\quad\text{with } y+tw_q\in\mathbb{R}^n,\ q=1,\dots,n-k. \tag{2-1}$$

This implies

$$(f-\ell_1)\left(y+\sum_{j=1}^{n-k}t_jw_j\right)=(f-\ell_1)(y) \tag{2-2}$$

if $y \in \mathbb{R}^n$ and $t_1, \dots, t_{n-k} \in \mathbb{R}$. Let P be the orthogonal projection of $\mathbb{R}^n$ onto the subspace $X := \operatorname{span}\{w_1, \dots, w_{n-k}\}^\perp$. For each $z \in X$ we may define

$$\tilde{c}(z) = (f - \ell_1)\left(z + \sum_{j=1}^{n-k} t_j w_j\right)$$

if $z + \sum_{j=1}^{n-k} t_j w_j \in \mathbb{R}^n$ for some $t_1, \dots, t_{n-k}$. It is clear that $\tilde{c} : X \to \mathbb{R}$ is well-defined and convex, and satisfies

$$f - \ell_1 = \tilde{c} \circ P.$$

Now let us write

$$\ell_1(x) = \langle u, x\rangle + \langle v, x\rangle,$$

where $u \in X$ and $v \in X^\perp$. We then have

$$f(x) = c(P(x)) + \langle v, x\rangle,$$

where $c : X \to \mathbb{R}$ is defined by

$$c(x) = \tilde{c}(x) + \langle u, x\rangle.$$

Moreover, since $\bigcap_{j=2}^{k+1} \operatorname{Ker}(\ell_j - \ell_1) = X^\perp$, it is clear that the restriction of the corner function $C = \max_{1 \le j \le k+1}\{\ell_j + b_j\}$ to X is a $(k+1)$-dimensional corner function on X, which has dimension k, and it is obvious that $(k+1)$-dimensional corner functions on k-dimensional spaces are essentially coercive; therefore, because $c(x) \ge C(x)$ for all $x \in X$, we deduce that c is essentially coercive.

Now let us see that X is the only linear subspace of $\mathbb{R}^n$ for which f admits a decomposition of the form

$$f(x) = c(P_X(x)) + \langle v, x\rangle, \tag{2-3}$$

with c essentially coercive and $v \in X^\perp$. Assume that we have two subspaces Z_1, Z_2 for which (2-3) holds, say

$$f(x) = \varphi_1(P_{Z_1}(x)) + \langle \xi_1, x\rangle, \tag{2-4}$$

and

$$f(x) = \varphi_2(P_{Z_2}(x)) + \langle \xi_2, x\rangle, \tag{2-5}$$

with φ_j essentially coercive and $\xi_j \in X_j^\perp$. In order to show that $Z_1 = Z_2$, it is enough to check that $Z_1^\perp = Z_2^\perp$. Suppose this equality does not hold; then, either $\in Z_1^\perp \setminus Z_2^\perp \neq \varnothing$ or $\in Z_2^\perp \setminus Z_1^\perp \neq \varnothing$. Assume for instance that there exists $\xi_0 \in Z_1^\perp \setminus Z_2^\perp$. Then, on the one hand (2-4) implies that the function $t \mapsto f(t\xi_0) = \varphi_1(0) + t\langle \xi_1, \xi_0\rangle$ is linear, and on the other hand (2-5) implies that the same function $t \mapsto f(t\xi_0) = \varphi_2(P_{Z_2}(t\xi_0)) + t\langle \xi_2, \xi_0\rangle$ is essentially coercive (indeed, we have $\lim_{|t|\to\infty} |P_{Z_2}(t\xi_0)| = \infty$ because $\xi_0 \notin Z_2^\perp$). This is absurd, so we must have $Z_1^\perp \subset Z_2^\perp$. By a similar argument, just switching the roles of Z_1 and Z_2, we also obtain that $Z_2^\perp \subset Z_1^\perp$. Therefore $Z_1^\perp = Z_2^\perp$, as we wanted to check.[3]

[3] It is worth noting that the preceding argument also shows that the dimension of X_f is k, the largest integer such that f is supported at some point by a $(k+1)$-dimensional corner function. In particular, it follows that a function is essentially coercive in $\mathbb{R}^n$ if and only if it is supported by an $(n+1)$-dimensional corner function.

Next, let us see that $\xi_1 = \xi_2$. For every $v \in Z_1^{\perp}$ we have

$$\varphi_1(0) + \langle \xi_1, v\rangle = f(v) = \varphi_2(0) + \langle \xi_2, v\rangle.$$

Since the equality of two affine functions implies the equality of their linear parts, we have

$$\langle \xi_1, v\rangle = \langle \xi_2, v\rangle$$

for all $v \in Z_1^{\perp}$, and because $\xi_1, \xi_2 \in Z_1^{\perp}$ this shows that $\xi_1 = \xi_2$.

Once we know that $X_1 = X_2$ and $\xi_1 = \xi_2$, it immediately follows from (2-4) and (2-5) that $\varphi_1 = \varphi_2$. This shows that the decomposition is unique.

Finally let us prove that if f is essentially coercive in the direction of a subspace Y (say that there exists a linear form ℓ on $\mathbb{R}^n$ such that $|f(x) - \ell(x)| \to \infty$ as $|P_Y(x)| \to \infty$), then $Y \subseteq X_f$. Indeed, otherwise there would exist a vector $\xi \in X^{\perp} \setminus Y^{\perp}$, and the function

$$\mathbb{R} \ni t \mapsto f(t\xi) = c(P_X(t\xi)) + t\langle v, \xi\rangle = c(0) + t\langle v, \xi\rangle$$

would be affine; hence so would be the function

$$\mathbb{R} \ni t \mapsto f(t\xi) - \ell(t\xi).$$

But this function cannot be affine, because $\xi \notin Y^{\perp}$ implies $|P_Y(t\xi)| \to \infty$ as $|t| \to \infty$, and we have $|f(x) - \ell(x)| \to \infty$ as $|P_Y(x)| \to \infty$. This completes the proof of Theorem 1.11 in the case that f is everywhere differentiable.

Case 2. In the case that $f : \mathbb{R}^n \to \mathbb{R}$ is convex but not everywhere differentiable, we can use [Azagra 2013, Theorem 1.1] in order to find a C^1 (or even real-analytic) convex function $g : \mathbb{R}^n \to \mathbb{R}$ such that $f - 1 \le g \le f$. Then we may apply Case 1 in order to find a unique subspace $X \subseteq \mathbb{R}^n$, an essentially coercive convex function $C : X \to \mathbb{R}$ and a vector $v \in X^{\perp}$ such that

$$g(z) = c(P(z)) + \langle v, z\rangle$$

for all $z \in \mathbb{R}^n$. Now take $x \in X$ and $\xi \in X^{\perp}$. The function $\mathbb{R} \ni t \mapsto g(t\xi)$, is affine, and because $f \le g+1$ and f is convex, so must be the function $\mathbb{R} \ni t \mapsto f(t\xi)$, and with the same linear part (this immediately follows from the fact that the only convex functions which are bounded above on $\mathbb{R}$ are constants). This shows that

$$f(x + t\xi) = f(x) + t\langle v, \xi\rangle$$

for every $x \in X$, $\xi \in X^{\perp}$, $t \in \mathbb{R}$. Equivalently, we can write

$$f(z) = \varphi(P(z)) + \langle v, z\rangle \quad \text{for all } z \in \mathbb{R}^n,$$

where $\varphi : X \to \mathbb{R}$ is defined by $\varphi(x) = f(x)$ for all $x \in X$. Moreover, φ is essentially coercive because so is $g_{|X}$ and we have $|f - g| \le 1$. This shows the existence of the decomposition in the statement. The uniqueness of the decomposition, as well as the last part of the statement of Theorem 1.11, follows by the same arguments as in Case 1, because that part of the proof does not use the differentiability of f. The proof of Theorem 1.11 is thus complete. □

3. Necessity of Theorem 1.13

Let F be a convex function of class $C^1(\mathbb{R}^n)$ such that $(F, \nabla F)$ extends (f, G) from E, and $X_F = X$.

***Condition* (i).** The inequality $f(x) - f(y) - \langle G(y), x - y\rangle \geq 0$ for all $x, y \in E$ follows from the fact that F is convex and differentiable with $(F, \nabla F) = (f, G)$ on E.

***Condition* (ii).** Assume that $(|\nabla F(x_k)|)_k$ tends to $+\infty$ for a sequence $(x_k)_k \subset \mathbb{R}^n$ but

$$\frac{\langle \nabla F(x_k), x_k\rangle - F(x_k)}{|\nabla F(x_k)|}$$

does not go to $+\infty$. Then, passing to a subsequence, we may assume that there exists $M > 0$ such that $\langle \nabla F(x_k), x_k\rangle - F(x_k) \leq M|\nabla F(x_k)|$ for all k. We define $z_k = 2M(\nabla F(x_k)/|\nabla F(x_k)|)$. By convexity, we have, for all k, that

$$0 \leq F(z_k) - F(x_k) - \langle \nabla F(x_k), z_k - x_k\rangle \leq F(z_k) - M|\nabla F(x_k)|,$$

which contradicts the assumption that $|\nabla F(x_k)| \to \infty$.

***Condition* (iii).** Making use of Theorem 1.11 and bearing in mind that $X_F = X$, we can write $F = c \circ P_X + \langle v, \cdot\rangle$, where $P_X : \mathbb{R}^n \to X$ is the orthogonal projection onto the subspace X, the function $c : X \to \mathbb{R}$ is convex and essentially coercive on X, and $v \perp X$. It is easy to see that c is differentiable on X and that $\nabla F(x) = \nabla c(P_X(x)) + v$ for all $x \in \mathbb{R}^n$. Since $F = G$ on E, we easily get $G(x) - G(y) \in X$ for all $x, y \in E$.

***Condition* (v).** Let us consider sequences $(x_k)_k$, $(z_k)_k$ on E such that $(P_X(x_k))_k$ and $(\nabla F(z_k))_k$ are bounded and

$$\lim_{k\to\infty} (F(x_k) - F(z_k) - \langle \nabla F(z_k), x_k - z_k\rangle) = 0. \tag{3-1}$$

Suppose that $|\nabla F(x_k) - \nabla F(z_k)|$ does not converge to 0. Then, using that $(P_X(x_k))_k$ is bounded, there exist some $x_0 \in X$ and $\varepsilon > 0$ for which, possibly after passing to a subsequence, $P_X(x_k)$ converges to x_0 and $|\nabla F(x_k) - \nabla F(z_k)| \geq \varepsilon$ for every k. By using the decomposition $F = c \circ P_X + \langle v, \cdot\rangle$ and some elementary properties of orthogonal projections with (3-1) we obtain

$$\lim_{k\to\infty} \big(c(P_X(x_k)) - c(P_X(z_k)) - \langle \nabla c(P_X(z_k)), P_X(x_k) - P_X(z_k)\rangle\big) = 0.$$

Since $\nabla F(y) - v = \nabla c(P_X(y))$ for all $y \in \mathbb{R}^n$ we have that $(\nabla c(P_X(z_k)))_k$ is bounded and

$$|\nabla c(P_X(x_k)) - \nabla c(P_X(z_k))| \geq \varepsilon$$

for every k. Additionally

$$\lim_{k\to\infty} (c(x_0) - c(P_X(z_k)) - \langle \nabla c(P_X(z_k)), x_0 - P_X(z_k)\rangle) = 0.$$

The contradiction follows from the lemma below.

Lemma 3.1. *Let $h : X \to \mathbb{R}$ be a differentiable convex function, $x_0 \in X$, and $(y_k)_k$ be a sequence in X such that $(\nabla h(y_k))_k$ is bounded and*

$$\lim_{k\to\infty} (h(x_0) - h(y_k) - \langle \nabla h(y_k), x_0 - y_k\rangle) = 0.$$

Then $\lim_{k\to\infty} |\nabla h(x_0) - \nabla h(y_k)| = 0$.

Proof. Suppose not. Then, up to extracting a subsequence, we would have $|\nabla h(x_0) - \nabla h(y_k)| \geq \varepsilon$ for some positive ε and for every k. Now, for every k, we set

$$\alpha_k := h(x_0) - h(y_k) - \langle \nabla h(y_k), x_0 - y_k\rangle, \qquad v_k := \frac{\nabla h(y_k) - \nabla h(x_0)}{|\nabla h(y_k) - \nabla h(x_0)|}.$$

In [Azagra and Mudarra 2017, Lemma 2.1] it is proved that $\alpha_k = 0$ implies $|\nabla h(x_0) - \nabla h(y_k)| = 0$, which is absurd. Thus we must have $\alpha_k > 0$ for every k. By convexity we have

$$\begin{aligned}\sqrt{\alpha_k}\langle \nabla h(x_0 + \sqrt{\alpha_k} v_k), v_k\rangle &\geq h(x_0 + \sqrt{\alpha_k} v_k) - h(x_0)\\ &\geq h(y_k) + \langle \nabla h(y_k), x_0 + \sqrt{\alpha_k} v_k - y_k\rangle - h(x_0) = -\alpha_k + \sqrt{\alpha_k}\langle \nabla h(y_k), v_k\rangle\end{aligned}$$

for all k. Hence, we obtain

$$\langle \nabla h(x_0 + \sqrt{\alpha_k} v_k) - \nabla h(x_0), v_k\rangle \geq -\sqrt{\alpha_k} + |\nabla h(y_k) - \nabla h(x_0)| \geq -\sqrt{\alpha_k} + \varepsilon.$$

But the above inequality is impossible, as ∇h is continuous and $\alpha_k \to 0$. □

***Condition* (iv).** By applying Theorem 1.11 we may write

$$F(x) = c(P_X(x)) + \langle v, x\rangle,$$

with $c : X \to \mathbb{R}$ convex and essentially coercive, and $v \perp X$. This implies

$$X = \operatorname{span}\{\nabla c(x) - \nabla c(y) : x, y \in X\},$$

and because $\nabla F = \nabla(c \circ P_X) + v$, also that

$$X = \operatorname{span}\{\nabla F(x) - \nabla F(y) : x, y \in \mathbb{R}^n\}.$$

Let us define $Y := \operatorname{span}\{\nabla F(x) - \nabla F(y) : x, y \in E\} \subset X$ and assume that $Y \neq X$. Let k and d denote the dimensions of Y and X respectively. We can find points $x_0, x_1, \dots, x_k \in E$ such that $Y = \operatorname{span}\{\nabla F(x_j) - \nabla F(x_0) : j = 1, \dots, k\}$. We claim that there exists $p_1 \in \mathbb{R}^n$ such that $\nabla F(p_1) - \nabla F(x_0) \notin Y$. Indeed, otherwise we would have $\nabla F(p) - \nabla F(x_0) \in Y$ for all $p \in \mathbb{R}^n$, which implies

$$\nabla F(p) - \nabla F(q) = (\nabla F(p) - \nabla F(x_0)) - (\nabla F(q) - \nabla F(x_0)) \in Y \quad \text{for all } p, q \in \mathbb{R}^n.$$

This is a contradiction since $X \neq Y$. Then the subspace Y_1 spanned by Y and the vector $\nabla F(p_1) - \nabla F(x_0)$ has dimension $k+1$. If $d = k+1$, we are done. If $d > k+1$, using the same argument as above, we can find a point $p_2 \in \mathbb{R}^n$ such that $\nabla F(p_2) - \nabla F(x_0) \notin Y_1$. By induction, we obtain points $p_1, \dots, p_{d-k} \in \mathbb{R}^n$

such that the set $\{\nabla F(p_j) - \nabla F(x_0)\}_{j=1}^{d-k}$ is linearly independent and $X = Y \oplus \operatorname{span}\{\nabla F(p_j) - \nabla F(x_0) : j = 1, \dots, d-k\}$, which shows that

$$X = \operatorname{span}\{u - w : u, w \in \nabla F(E) \cup \{\nabla F(p_1), \dots, \nabla F(p_{d-k})\}\}.$$

This shows the necessity of (iv)(a). Obviously we have $\nabla F(p_j) - \nabla F(x_0) \in X \setminus Y$ for all $j = 1, \dots, d-k$, and we claim that

$$p_j \in \mathbb{R}^n \setminus \bar{E} \quad \text{for all } j = 1, \dots, d-k.$$

Indeed, if there exists a sequence $(q_\ell)_\ell \subset E$ with $(q_\ell)_\ell \to p_j$ for some $j = 1, \dots, d-k$, then, because Y is closed and ∇F is continuous, $\nabla F(p_j) - \nabla F(x_0) = \lim_\ell (\nabla F(q_\ell) - \nabla F(x_0)) \in Y$, which is a contradiction. By the (already shown) necessity of condition (v), applied with $E^* = E \cup \{p_1, \dots, p_{d-k}\}$ in place of E, we have

$$\lim_{\ell \to \infty} |\nabla F(x_\ell) - \nabla F(z_\ell)| = 0 \tag{3-2}$$

whenever $(x_\ell)_\ell$, $(z_\ell)_\ell$ are sequences in E^* such that $(P_X(x_\ell))_\ell$ and $(\nabla F(z_\ell))_\ell$ are bounded and

$$\lim_{\ell \to \infty} (F(x_\ell) - F(z_\ell) - \langle \nabla F(z_\ell), x_\ell - z_\ell \rangle) = 0.$$

But the fact that $\operatorname{dist}(\nabla F(p_j) - \nabla F(x_0), Y) > 0$ for each $j = 1, \dots, d-k$ prevents the limiting condition (3-2) from holding true with $(z_\ell)_\ell \subset \{p_1, \dots, p_{d-k}\}$ and $(x_\ell)_\ell \subset E$. This implies that the inequalities

$$\begin{aligned}
F(p_j) &\geq F(p_i) + \langle \nabla F(p_i), p_j - p_i \rangle, && 1 \leq i, j \leq d-k,\ i \neq j, \\
F(p_j) &\geq \sup_{z \in E, |\nabla F(z)| \leq N} \{F(z) + \langle \nabla F(z), p_j - z \rangle\}, && 1 \leq j \leq d-k,\ N \in \mathbb{N}, \\
F(x) &\geq F(p_j) + \langle \nabla F(p_j), x - p_j \rangle, && 1 \leq j \leq d-k,\ x \in \mathbb{R}^n,
\end{aligned}$$

which generally hold by convexity of F, must all be strict. Moreover, the last of these inequalities, together with (3-2), also implies

$$\inf_{x \in E,\, |P_X(x)| \leq N} \Big\{ F(x) - \max_{1 \leq j \leq d-k} \{F(p_j) + \langle \nabla F(p_j), x - p_j \rangle\} \Big\} > 0$$

for all $N \in \mathbb{N}$. Setting $w_j = \nabla F(p_j)$ and $\beta_j = F(p_j)$, $j = 1, \dots, d-k$, this shows the necessity of (iv)(b)–(d).

4. Sufficiency of Theorem 1.13

First of all, with the notation of condition (iv), if $Y \neq X$, we define

$$E^* = E \cup \{p_1, \dots, p_{d-k}\}$$

and extend the functions f and G to E^* by setting

$$f(p_j) := \beta_j, \quad G(p_j) := w_j \quad \text{for } j = 1, \dots, d-k. \tag{4-1}$$

If $Y = X$, we just set $E^* = E$ and ignore any reference to the points p_j and their companions w_j and β_j in what follows.

Lemma 4.1. *We have*:

(a) $X = \mathrm{span}(\{G(x) - G(y) : x, y \in E^*\})$.

(b) *There exists* $r > 0$ *such that* $f(p_i) - f(p_j) - \langle G(p_j), p_i - p_j\rangle \geq r$ *for all* $1 \leq i \neq j \leq d-k$.

(c) *For every* $N \in \mathbb{N}$, *there exists* $r_N > 0$ *with* $f(p_i) - f(z) - \langle G(z), p_i - z\rangle \geq r_N$ *for all* $z \in E$ *with* $|G(z)| \leq N$ *and all* $1 \leq i \leq d-k$.

(d) *For every* $N \in \mathbb{N}$, *there exists* $r_N > 0$ *with* $f(x) - f(p_i) - \langle G(p_i), x - p_i\rangle \geq r_N$ *for all* $x \in E$ *with* $|P_X(x)| \leq N$ *and all* $1 \leq i \leq d-k$.

Proof. This follows immediately from condition (iv) and the definitions of (4-1). □

Lemma 4.2. *The jet* (f, G) *defined on* E^* *satisfies the inequalities of the assumption* (i) *on* E^*. *Moreover, if* $(x_k)_k$, $(z_k)_k$ *are sequences in* E^* *such that* $(P_X(x_k))_k$ *and* $(G(z_k))_k$ *are bounded, then*

$$\lim_{k\to\infty} (f(x_k) - f(z_k) - \langle G(z_k), x_k - z_k\rangle) = 0 \quad \Longrightarrow \quad \lim_{k\to\infty} |G(x_k) - G(z_k)| = 0.$$

Proof. Suppose that $(x_k)_k$, $(z_k)_k$ are sequences in E^* such that $(P_X(x_k))_k$ and $(G(z_k))_k$ are bounded and $\lim_{k\to\infty}(f(x_k)-f(z_k)-\langle G(z_k), x_k-z_k\rangle)=0$. In view of Lemma 4.1(b), (c) and (d), it is immediate that there exists k_0 such that either there is some $1\leq i\leq d-k$ with $x_k = z_k = p_i$ for all $k\geq k_0$ or else $x_k, z_k \in E$ for all $k \geq k_0$. In the first case, the conclusion is trivial. In the second case, $\lim_{k\to\infty} |G(x_k) - G(z_k)| = 0$ follows from condition (v) of Theorem 1.13. □

We now consider the minimal convex extension of the jet (f, G) from E^*, defined by

$$m(x) = m(f, G, E^*)(x) := \sup_{y\in E^*} \{f(y) + \langle G(y), x - y\rangle\}, \quad x \in \mathbb{R}^n.$$

It is clear that m, being the supremum of a family of affine functions, is a convex function on $\mathbb{R}^n$. In fact, we have the following.

Lemma 4.3. *The function* $m(x)$ *is finite for every* $x \in \mathbb{R}^n$. *In addition*, $m = f$ *on* E^* *and* $G(x) \in \partial m(x)$ *for all* $x \in E^*$.

Here $\partial m(x) := \{\xi \in \mathbb{R}^n : m(y) \geq m(x) + \langle \xi, y - x\rangle$ for all $y \in \mathbb{R}^n\}$ is the subdifferential of f at x.

Proof. Fix a point $z_0 \in E^*$. For any given point $x \in \mathbb{R}^n$ it is clear that there exists a sequence $(y_k)_k$ (possibly stationary) in E^* such that

$$f(z_0) + \langle G(z_0), x - z_0\rangle \leq f(y_k) + \langle G(y_k), x - y_k\rangle \quad \text{for all } k,$$

and $f(y_k) + \langle G(y_k), x - y_k\rangle \to m(x)$ as $k\to\infty$. On the other hand, by the first statement of Lemma 4.2, we have

$$f(y_k) + \langle G(y_k), x - y_k\rangle \leq f(z_0) + \langle G(y_k), x - z_0\rangle.$$

Then it is clear that $m(x) < +\infty$ when $(G(y_k))_k$ is a bounded sequence. We next show that this sequence can never be unbounded. Indeed, in such case, by the condition (ii) in Theorem 1.13 (which obviously holds with E^* in place of E), we would have a subsequence for which $\lim_{k\to\infty}|G(y_k)| = +\infty$, which in turn implies

$$\lim_{k\to\infty} \frac{\langle G(y_k), y_k\rangle - f(y_k)}{|G(y_k)|} = +\infty.$$

Hence, by the assumption on $(y_k)_k$ we would have

$$\frac{f(y_k) - \langle G(y_k), y_k\rangle}{|G(y_k)|} \geq \frac{f(z_0) + \langle G(z_0), x - z_0\rangle}{|G(y_k)|} - \left\langle \frac{G(y_k)}{|G(y_k)|}, x\right\rangle.$$

Since $\lim_{k\to\infty}|G(y_k)| = +\infty$, the right-hand term is bounded below, and this leads to a contradiction. Therefore $m(x) < +\infty$ for all $x \in \mathbb{R}^n$. In addition, by using the definition of m and the first statement of Lemma 4.2 for the jet (f, G), we easily obtain that $m = f$ on E^* and that $G(x)$ belongs to $\partial m(x)$ for all $x \in E^*$. □

Lemma 4.4. *The function m is essentially coercive in the direction of X, and in fact, with the notation of Theorem 1.11 we have*

$$X_m = X.$$

Proof. By Lemma 4.1(a), we have $X = \operatorname{span}(\{G(x) - G(y) : x, y \in E^*\})$. Let us first see that m is essentially coercive in the direction of X. If $X = \{0\}$ then m is affine and the result is obvious. Therefore we can assume $\dim(X) \geq 1$ and take points $x_0, x_1, \ldots, x_k \in E$ such that $\{v_1, \ldots, v_k\}$ is a basis of X, where

$$v_j = G(x_j) - G(x_0), \qquad j = 1, \ldots, k.$$

Then

$$C(x) = \max\{f(x_0) + \langle G(x_0), x - x_0\rangle,\ f(x_1) + \langle G(x_1), x - x_1\rangle,\ \ldots,\ f(x_k) + \langle G(x_k), x - x_k\rangle\}$$

defines a k-dimensional corner function such that

$$C(x) \leq m(x) \quad \text{for all } x \in \mathbb{R}^n,$$

and it is not difficult to see that C is essentially coercive in the direction of X; hence so is m.

In particular, by Theorem 1.11, it follows that $X \subseteq X_m$.

Now, if $X_m \neq X$, we can take a vector $w \in X_m \setminus \{0\}$ such that $w \perp X$, and then we obtain, for all $t \in \mathbb{R}$, that

$$\begin{aligned} m(x_0 + tw) - f(x_0) - \langle G(x_0), tw\rangle &= \sup_{z\in E}\{f(z) - f(x_0) + \langle G(z) - G(x_0), tw\rangle + \langle G(z), x_0 - z\} \\ &= \sup_{z\in E}\{f(z) - f(x_0) + \langle G(z), x_0 - z\} \leq 0. \end{aligned}$$

By convexity, this implies

$$m(x_0 + tw) = f(x_0) + \langle G(x_0), tw\rangle$$

for all $t \in \mathbb{R}$, and in particular the function $\mathbb{R} \ni t \mapsto m(x_0 + tw)$ cannot be essentially coercive, contradicting the assumption that $w \in X_m$. Therefore we must have $X_m = X$. □

Making use of Theorem 1.11 in combination with Lemma 4.4, we can write

$$m = c \circ P_X + \langle v, \cdot \rangle \quad \text{on } \mathbb{R}^n, \tag{4-2}$$

where $c : X \to \mathbb{R}$ is convex and essentially coercive on X and $v \perp X$. In addition, the subdifferential mappings of m and c satisfy the following.

Claim 4.5. *Given $x \in \mathbb{R}^n$ and $\eta \in \partial m(x)$, we have $\eta - v \in X$ and $\eta - v \in \partial c(P_X(x))$.*

Proof. Suppose that $x \in \mathbb{R}^n$ and $\eta \in \partial m(x)$ but $\eta - v \notin X$. Then we can find $w \in X^\perp$ with $\langle \eta - v, w\rangle = 1$. Using (4-2) we get that

$$\langle \eta, w\rangle \le m(x+w) - m(x) = c(P_X(x+w)) + \langle v, x+w\rangle - c(P_X(x)) - \langle v, x\rangle = \langle v, w\rangle.$$

This implies $\langle \eta - v, w\rangle \le 0$, a contradiction. This shows that $\eta - v \in X$. Now, let $z \in X$ and $x \in \mathbb{R}^n$. We have

$$c(z) - c(P_X(x)) = m(z) - \langle v, z\rangle - m(x) + \langle v, x\rangle \ge \langle \eta - v, z - x\rangle = \langle \eta - v, z - P_X(x)\rangle.$$

Therefore, $\eta - v \in \partial c(P_X(x))$. □

By combining the previous claim with the second part of Lemma 4.3 we obtain that

$$G(x) - v \in \partial c(P_X(x)) \subset X \quad \text{for all } x \in E^*. \tag{4-3}$$

Lemma 4.6. *The function c is differentiable on $\overline{P_X(E^*)}$, and, if $y \in P_X(E^*)$, then $\nabla c(y) = G(x) - v$, where $x \in E^*$ is such that $P_X(x) = y$.*

Proof. Let us suppose that c is not differentiable at some $y_0 \in \overline{P_X(E^*)}$. Then, by the convexity of c on X, we may assume that there exist a sequence $(h_k)_k \subset X$ with $|h_k| \searrow 0$ and a number $\varepsilon > 0$ such that

$$\varepsilon \le \frac{c(y_0 + h_k) + c(y_0 - h_k) - 2c(y_0)}{|h_k|} \quad \text{for all } k.$$

We now consider sequences $(y_k)_k \subset P_X(E^*)$ and $(x_k)_k \subset E^*$ with

$$P_X(x_k) = y_k \quad \text{and} \quad y_k \to y_0.$$

In particular, the sequence $(P_X(x_k))_k$ is bounded. Since each h_k belongs to X, we can use (4-2) to rewrite the last inequality as

$$\varepsilon \le \frac{m(y_0 + h_k) + m(y_0 - h_k) - 2m(y_0)}{|h_k|} \quad \text{for all } k. \tag{4-4}$$

By the definition of m we can pick two sequences $(z_k)_k, (\tilde{z}_k)_k \subset E^*$ with the following properties:

$$m(y_0 + h_k) \ge f(z_k) + \langle G(z_k), y_0 + h_k - z_k\rangle \ge m(y_0 + h_k) - \frac{|h_k|}{2^k},$$
$$m(y_0 - h_k) \ge f(\tilde{z}_k) + \langle G(\tilde{z}_k), y_0 - h_k - \tilde{z}_k\rangle \ge m(y_0 - h_k) - \frac{|h_k|}{2^k}$$

for every k. We claim that $(G(z_k))_k$ must be bounded. Indeed, otherwise, possibly after passing to a subsequence and using condition (ii) of Theorem 1.13, we would obtain that

$$\lim_{k\to\infty} |G(z_k)| = \lim_{k\to\infty} \frac{\langle G(z_k), z_k\rangle - f(z_k)}{|G(z_k)|} = +\infty.$$

Due to the choice of $(z_k)_k$ we must have

$$\begin{aligned} m(y_0) &= \lim_{k\to\infty} (f(z_k) + \langle G(z_k), x_0 + h_k - z_k\rangle) \\ &= \lim_{k\to\infty} |G(z_k)| \left(\frac{f(z_k) - \langle G(z_k), z_k\rangle}{|G(z_k)|} + \left\langle \frac{G(z_k)}{|G(z_k)|}, x_0 + h_k \right\rangle \right) = -\infty, \end{aligned}$$

which is absurd. Similarly one can show that $(G(\tilde{z}_k))_k$ is bounded. Now we write

$$\begin{aligned} f(x_k) - f(z_k) - \langle G(z_k), x_k - z_k\rangle = f(x_k) - \langle v, x_k\rangle - (m(y_0 + k_k) - \langle v, y_0 + h_k\rangle) + m(y_0 + h_k) \\ - f(z_k) - \langle G(z_k), y_0 + h_k - z_k\rangle + \langle G(z_k) - v, y_0 + h_k - x_k\rangle. \end{aligned}$$

By (4-2), the first term in the sum equals $c(P_X(x_k)) - c(y_0 + h_k)$, which converges to 0 because $P_X(x_k) \to y_0$ and c is continuous. Thanks to the choice of the sequence $(z_k)_k$, the second term also converges to 0. From (4-3), we have $G(z_k) - v \in X$ for all k, and then the third term in the sum is actually $\langle G(z_k) - v, y_0 - P_X(x_k) + h_k\rangle$, which converges to 0, as $(G(z_k))_k$ is bounded and $P_X(x_k) \to y_0$. We then have

$$\lim_{k\to\infty} (f(x_k) - f(z_k) - \langle G(z_k), x_k - z_k\rangle) = 0,$$

where $(P_X(x_k))_k$ and $(G(z_k))_k$ are bounded sequences. By Lemma 4.2, $\lim_{k\to\infty} |G(x_k) - G(z_k)| = 0$, and similarly one can show that $\lim_{k\to\infty} |G(x_k) - G(\tilde{z}_k)| = 0$. This obviously implies

$$\lim_{k\to\infty} |G(z_k) - G(\tilde{z}_k)| = 0. \tag{4-5}$$

By the choice of the sequences $(z_k)_k$, $(\tilde{z}_k)_k$ and by inequality (4-4) we have, for every k,

$$\begin{aligned} \varepsilon &\le \frac{f(z_k) + \langle G(z_k), y_0 + h_k - z_k\rangle}{|h_k|} + \frac{f(\tilde{z}_k) + \langle G(\tilde{z}_k), y_0 - h_k - \tilde{z}_k\rangle}{|h_k|} \\ &\qquad - \frac{f(z_k) + \langle G(z_k), y_0 - z_k\rangle + f(\tilde{z}_k) + \langle G(\tilde{z}_k), y_0 - \tilde{z}_k\rangle}{|h_k|} \\ &= \left\langle G(z_k) - G(\tilde{z}_k), \frac{h_k}{|h_k|} \right\rangle + \frac{1}{2^{k-1}} \le |G(z_k) - G(\tilde{z}_k)| + \frac{1}{2^{k-1}}. \end{aligned}$$

Then (4-5) leads us to a contradiction. We conclude that c is differentiable on $\overline{P_X(E^*)}$.

We now prove the second part of the lemma. Consider $y \in P_X(E^*)$ and $x \in E^*$ with $P_X(x) = y$. By (4-3), we have $G(x) - v \in \partial c(y)$. Because c is differentiable at y, we further obtain that $G(x) - v = \nabla c(y)$. $\square$

In order to complete the proof of Theorem 1.13, we will need the following lemma.

Lemma 4.7. *Let $h : X \to \mathbb{R}$ be a convex and coercive function such that h is differentiable on a closed subset A of X. There exists $H \in C^1(X)$ convex and coercive such that $H = h$ and $\nabla H = \nabla h$ on A.*

Proof. Since h is convex, its gradient ∇h is continuous on A (see [Rockafellar 1970, Corollary 24.5.1] for instance). Then, for all $x, y \in A$, we have

$$0 \le \frac{h(x) - h(y) - \langle \nabla h(y), x - y \rangle}{|x - y|} \le \left\langle \nabla h(x) - \nabla h(y), \frac{x - y}{|x - y|} \right\rangle \le |\nabla h(x) - \nabla h(y)|,$$

where the last term tends to 0 as $|x - y| \to 0$ uniformly on $x, y \in K$ for every compact subset K of A. This shows that the pair $(h, \nabla h)$ defined on A satisfies the conditions of the classical Whitney extension theorem for C^1 functions. Therefore, there exists a function $\tilde{h} \in C^1(X)$ such that $\tilde{h} = h$ and $\nabla \tilde{h} = \nabla h$ on A. We now define

$$\phi(x) := |h(x) - \tilde{h}(x)| + 2d(x, A)^2, \quad x \in X. \tag{4-6}$$

Claim 4.8. *The function ϕ is differentiable on A, with $\nabla \phi(x_0) = 0$ for every $x_0 \in A$.*

Proof. The function $d(\,\cdot\,, A)^2$ is obviously differentiable, with a null gradient, at x_0; hence we only have to see that $|h - \tilde{h}|$ is differentiable, with a null gradient, at x_0. Since $\nabla \tilde{h}(x_0) = \nabla h(x_0)$, the claim boils down to the following easy exercise: if two functions h_1, h_2 are differentiable at x_0, with $\nabla h_1(x_0) = \nabla h_2(x_0)$, then $|h_1 - h_2|$ is differentiable, with a null gradient, at x_0. □

Now, because $d(\,\cdot\,, A)^2$ is continuous and positive on $X \setminus A$, according to Whitney's approximation theorem [1934] we can find a function $\varphi \in C^\infty(X \setminus A)$ such that

$$|\varphi(x) - \phi(x)| \le d(x, A)^2 \quad \text{for every } x \in X \setminus A. \tag{4-7}$$

Let us define $\tilde{\varphi} : X \to \mathbb{R}$ by $\tilde{\varphi} = \varphi$ on $X \setminus A$ and $\tilde{\varphi} = 0$ on A.

Claim 4.9. *The function $\tilde{\varphi}$ is differentiable on X and $\nabla \tilde{\varphi} = 0$ on A.*

Proof. It is obvious that $\tilde{\varphi}$ is differentiable on $\operatorname{int}(A) \cup (X \setminus A)$ and $\nabla \tilde{\varphi} = 0$ on $\operatorname{int}(A)$. We only have to check that $\tilde{\varphi}$ is differentiable on ∂A. If $x_0 \in \partial A$ we have

$$\frac{|\tilde{\varphi}(x) - \tilde{\varphi}(x_0)|}{|x - x_0|} = \frac{|\tilde{\varphi}(x)|}{|x - x_0|} \le \frac{|\phi(x)| + d(x, A)^2}{|x - x_0|} \to 0$$

as $|x - x_0| \to 0^+$, because both ϕ and $d(\,\cdot\,, A)^2$ vanish at x_0 and are differentiable, with null gradients, at x_0. Therefore $\tilde{\varphi}$ is differentiable at x_0, with $\nabla \tilde{\varphi}(x_0) = 0$. □

Now we set

$$g := \tilde{h} + \tilde{\varphi}$$

on X. It is clear that $g = h$ on A. Also, by Claim 4.9, g is differentiable on X with $\nabla g = \nabla h$ on A. By combining (4-6) and (4-7) we easily obtain that

$$g(x) \ge \tilde{h}(x) + \phi(x) - d(x, A)^2 \ge h(x), \quad x \in X \setminus A.$$

Therefore $g \ge h$ on X and in particular g is coercive on X, because so is h, by assumption.

We next consider the *convex envelope* of g. Recall that, for a function $\psi : X \to \mathbb{R}$, the convex envelope of ψ is defined by

$$\operatorname{conv}(\psi)(x) = \sup\{\Phi(x) : \Phi \text{ is convex}, \Phi \le \psi\}$$

(another expression for $\operatorname{conv}(\psi)$, which follows from Carathéodory's theorem, is

$$\operatorname{conv}(\psi)(x) = \inf\left\{ \sum_{j=1}^{n+1} \lambda_j \psi(x_j) : \lambda_j \geq 0,\ \sum_{j=1}^{n+1} \lambda_j = 1,\ x = \sum_{j=1}^{n+1} \lambda_j x_j \right\};$$

see [Rockafellar 1970, Corollary 17.1.5] for instance). The following result is a restatement of a particular case of the main theorem in [Kirchheim and Kristensen 2001]; see also [Griewank and Rabier 1990].

Theorem 4.10 (Kirchheim–Kristensen). *If $\psi : X \to \mathbb{R}$ is differentiable and $\lim_{|x|\to\infty} \psi(x) = \infty$, then* $\operatorname{conv}(\psi) \in C^1(X)$.

If we define

$$H = \operatorname{conv}(g),$$

we immediately get that H is convex on X and $H \in C^1(X)$. By the definition of H we have $h \leq H \leq g$ on X, which implies that H is coercive. Also, because $g = h$ on A, we have $H = h$ on A. In order to show that $\nabla H = \nabla h$ on A, we use the following well-known criterion for differentiability of convex functions, whose proof is straightforward.

Lemma 4.11. *If ψ is convex, Φ is differentiable at x, $\psi \leq \Phi$, and $\psi(x) = \Phi(x)$, then ψ is differentiable at x, with $\nabla\psi(x) = \nabla\Phi(x)$.*

(This fact can also be phrased as: a convex function ψ is differentiable at x if and only if ψ is superdifferentiable at x.)

Since h is convex and H is differentiable on X with $h = H$ on A and $h \leq H$ on X, the preceding lemma shows that $\nabla H = \nabla h$ on A.

This completes the proof of Lemma 4.7. □

Now we are able to finish the proof of Theorem 1.13. Setting $A := \overline{P_X(E^*)}$, we see from Lemma 4.6 that c is differentiable on A. Moreover, since $c : X \to \mathbb{R}$ is convex and essentially coercive on X, there exists $\eta \in X$ such that $h := c - \langle \eta, \cdot \rangle$ is convex, differentiable on A and coercive on X. Applying Lemma 4.7 to h, we obtain $H \in C^1(X)$ convex and coercive on X with $(H, \nabla H) = (h, \nabla h)$ on A. Thus, the function $\varphi := H + \langle \eta, \cdot \rangle$ is convex, essentially coercive on X and of class $C^1(X)$ with $(\varphi, \nabla\varphi) = (c, \nabla c)$ on A. We next show that $F := \varphi \circ P_X + \langle v, \cdot \rangle$ is the desired extension of (f, G). Since φ is $C^1(X)$ and convex, it is clear that F is $C^1(\mathbb{R}^n)$ and convex as well. Bearing in mind Theorem 1.11 and the fact that φ is essentially coercive, it follows that $X_F = X$. Also, since $\varphi(y) = c(y)$ for $y \in P_X(E)$, we obtain from (4-2) and Lemma 4.3 that

$$F(x) = \varphi(P_X(x)) + \langle v, x \rangle = c(P_X(x)) + \langle v, x \rangle = m(x) = f(x).$$

Finally, from the second part of Lemma 4.6, we have, for all $x \in E$, that

$$\nabla F(x) = \nabla\varphi(P_X(x)) + v = G(x) - v + v = G(x).$$

The proof of Theorem 1.13 is complete. □

5. Necessity of Theorem 1.14

We already know that conditions (i), (ii) and (iv) are necessary for the existence of a convex function $F \in C^1(\mathbb{R}^n)$ with $(f, G) = (F, \nabla F)$ on E and $X_F = X$. Let us assume that F is also Lipschitz, and let us prove that in this case condition (iii) is satisfied as well. If $\operatorname{Lip}(F) = 0$ then F is constant, so we have $X = X_F = \{0\} = Y$, and condition (iii) is trivially satisfied. Otherwise we have $X = X_F \neq \{0\}$, and assuming that $Y \neq X$ we may find points $x_0, x_1, \ldots, x_k \in E$ and $p_1, \ldots, p_{d-k} \in \mathbb{R}^n \setminus \overline{E}$ such that

$$Y = \operatorname{span}\{G(x_j) - G(x_0) : j = 1, \ldots, k\},$$
$$\nabla F(p_j) - G(x_0) \in X \setminus Y \quad \text{for every } j = 1, \ldots, d-k$$

and the set $\{\nabla F(p_j) - G(x_0) : j = 1, \ldots, d-k\}$ is linearly independent. Now we define, for each $j = 1, \ldots, d-k$, the subspace Y_j spanned by Y and the vector $\nabla F(p_j) - G(x_0)$. Obviously we can find $w_j \in Y_j \cap Y^{\perp}$ with $|w_j| = 1$ and $Y_j = Y \oplus [w_j]$, for every $j = 1, \ldots, d-k$. Moreover, w_j can be taken so that

$$\mu_j := \langle \nabla F(p_j) - G(x_0), w_j \rangle > 0 \quad \text{for all } j = 1, \ldots, d-k.$$

Let us take $\varepsilon > 0$ small enough so that

$$\varepsilon < \frac{\mu_j}{2\operatorname{Lip}(F) + 2\|G\|_\infty} \quad \text{for all } j = 1, \ldots, d-k.$$

Note that, because $\mu_j \leq 2\operatorname{Lip}(F)$ for each j, we have $\varepsilon \leq 1$. Now, assume that there exists some $x \in \overline{E}$ with $x \in V_j := \{x \in \mathbb{R}^n : \varepsilon\langle w_j, x - p_j\rangle \geq |P_Y(x - p_j)|\}$ for some $j = 1, \ldots, d-k$ Using the convexity of F we can easily write

$$\begin{aligned}
&F(x) - F(p_j) - \langle \nabla F(p_j), x - p_j\rangle \\
&\qquad \leq \langle \nabla F(x) - \nabla F(p_j), x - p_j\rangle \\
&\qquad = \langle \nabla F(x) - G(x_0), x - p_j\rangle + \langle G(x_0) - \nabla F(p_j), x - p_j\rangle \\
&\qquad = \langle \nabla F(x) - G(x_0), x - p_j\rangle - \mu_j\langle w_j, x - p_j\rangle + \langle P_Y(G(x_0) - \nabla F(p_j)), x - p_j\rangle.
\end{aligned}$$

Since we are assuming that $x \in \overline{E}$, the continuity of ∇F yields $\nabla F(x) - G(x_0) \in Y$. Then, the last term coincides with

$$\begin{aligned}
&\langle \nabla F(x) - G(x_0), P_Y(x - p_j)\rangle - \mu_j\langle w_j, x - p_j\rangle + \langle P_Y(G(x_0) - \nabla F(p_j)), P_Y(x - p_j)\rangle \\
&\qquad \leq (2\|G\|_\infty + 2\operatorname{Lip}(F))|P_Y(x - p_j)| - \mu_j\langle w_j, x - p_j\rangle \\
&\qquad \leq 0,
\end{aligned}$$

where the last inequality follows from the definition of ε and the fact that $x \in V_j$. We have thus shown that

$$F(x) - F(p_j) - \langle \nabla F(p_j), x - p_j\rangle = 0,$$

which implies, by condition (CW^1), that $\nabla F(p_j) = \nabla F(x)$, where $x \in \overline{E}$. It follows that $\nabla F(p_j) - G(x_0) = \nabla F(x) - G(x_0) \in Y$, which contradicts the choice of p_j. Therefore $\overline{E}$ and $\bigcup_{j=1}^{d-k} V_j$ are disjoint.

6. Sufficiency of Theorem 1.14: keeping control of the Lipschitz constant

If m denotes the *minimal convex extension* of the jet (f, G) from E, we can write

$$m = c \circ P_{X_m} + \langle v, \cdot \rangle,$$

where $v \in \mathbb{R}^n$ and $c : X_m \to \mathbb{R}$ is a coercive convex function. Moreover, we know that

$$X_m = Y = \operatorname{span}\{G(x) - G(y) : x, y \in E\}$$

and therefore

$$m = c \circ P_Y + \langle v, \cdot \rangle. \tag{6-1}$$

Let us prove some properties of m, c and v.

Lemma 6.1. *Let us define $K = \|G\|_\infty = \sup_{y \in E} |G(y)|$. We have*:

(1) *The function m is K-Lipschitz on $\mathbb{R}^n$.*

(2) *The vector v belongs to the subdifferential of m at some point $y_0 \in Y$, and $|v| \leq K$.*

(3) *There exists points $x_1, \dots, x_k \in E$ such that $\{G(x_j) - v\}_{j=1}^k$ is a basis of Y.*

(4) *The function c is $2K$-Lipschitz on Y.*

(5) *There exist numbers $0 < \alpha \leq 2K$ and $\beta \in \mathbb{R}$ such that $c(y) \geq \alpha |y| + \beta$ for every $y \in Y$.*

Proof. (1) The function m is a supremum of K-Lipschitz affine functions on $\mathbb{R}^n$ and therefore m is K-Lipschitz as well.

(2) Since c is coercive on Y, there exists a point $y_0 \in Y$ with $c(y) \geq c(y_0)$ for every $y \in Y$. We then have, for every $x \in \mathbb{R}^n$, that

$$m(x) = c(P_Y(x)) + \langle v, x \rangle \geq c(y_0) + \langle v, x \rangle = c(y_0) + \langle v, y_0 \rangle + \langle v, x - y_0 \rangle = m(y_0) + \langle v, x - y_0 \rangle,$$

which implies $v \in \partial m(y_0)$. Since m is K-Lipschitz, we obtain, for every $x \in \mathbb{R}^n$,

$$K|x - y_0| + m(y_0) \geq m(x) \geq m(y_0) + \langle v, x - y_0 \rangle,$$

which implies $\langle v, (x - y_0)/|x - y_0| \rangle \leq K$ for every $x \in \mathbb{R}^n \setminus \{y_0\}$. This shows that $|v| \leq K$.

(3) Recall that $\eta - v \in Y$ for every $\eta \in \partial m(x)$. In particular we have $G(x) - v \in Y$ for every $x \in E$. Let us take some $x_1 \in Y$ with $G(x_1) - v \neq 0$. If $\dim(Y) = 1$, there is nothing to say. If $\dim(Y) > 1$, we claim that there exists some $x_2 \in E$ such that $G(x_2) - v$ and $G(x_1) - v$ are linearly independent. Indeed, assume that $G(x) - v$ and $G(x_1) - v$ are proportional for every $x \in E$. Then we would have for every $x, y \in E$ that

$$G(x) - G(y) = (G(x) - v) + (v - G(y))$$

is proportional to $G(x_1) - v$; hence $\dim(Y) = 1$, a contradiction. Using an inductive argument we easily obtain (3).

(4) This follows at once from (1), (2), and the fact that $c = m - \langle v, \cdot \rangle$ on Y.

(5) It is well known and easy to show that for every coercive convex function c there exist numbers $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $c(y) \geq \alpha|y| + \beta$ for every $y \in Y$. Now, because c is $2K$-Lipschitz, we have

$$c(0) + 2K|y| \geq c(y) \geq \alpha|y| + \beta, \qquad y \in Y.$$

This clearly implies $\alpha \leq 2K$. □

Defining new data. Let us consider $w_1, \dots, w_{d-k} \in Y^{\perp} \cap X$, $\varepsilon \in (0, 1)$, $p_1, \dots, p_{d-k}$ and $V_1, \dots, V_{d-k}$ as in condition (iii) of Theorem 1.14. Using Lemma 6.1(5), we consider a positive $T > 0$ large enough so that

$$(\varepsilon\alpha)T \geq 2 - \beta - \max_{j=1,\dots,d-k} \{\alpha|P_Y(p_j)| + m(p_j) - \langle v, p_j\rangle\},$$

$$(\varepsilon\alpha)T \min_{1\leq i\neq j\leq d-k} \{1 - \langle w_i, w_j\rangle\} \geq 1 + \max_{1\leq i,j\leq d-k} \{c(P_Y(p_j)) - c(P_Y(p_i)) + \varepsilon\alpha\langle w_j, p_i - p_j\rangle\}.$$

Note that, since the vectors $\{w_i\}_{i=1}^{d-k}$ have norm equal to 1, we have $\langle w_i, w_j\rangle = 1$ if and only if $w_i = w_j$, which is equivalent (as the vectors $\{w_1, \dots, w_{d-k}\}$ are linearly independent) to $i = j$. So it is clear that we can find a positive $T > 0$ satisfying both inequalities. We define the following new data:

$$q_j = p_j + T w_j, \quad f(q_j) = m(q_j) + 1, \quad G(q_j) = v + \varepsilon\alpha w_j, \quad j = 1, \dots, d-k. \tag{6-2}$$

Note that $q_i = q_j$ if and only if $p_i - p_j = T(w_j - w_i)$. Since $w_i \neq w_j$ whenever $i \neq j$, it is clear that we can take T large enough so that the points q_i and q_j are distinct if $i \neq j$. On the other hand, because each w_j is orthogonal to Y, we immediately see that $q_j \in V_j$ and, in particular, $q_j \notin \overline{E}$ for every $j = 1, \dots, d-k$.

Lemma 6.2. *The following inequalities are satisfied*:

(1) $f(q_j) - f(x) - \langle G(x), q_j - x\rangle \geq 1$ *for every* $x \in E$, $j = 1, \dots, d-k$.

(2) $f(x) - f(q_j) - \langle G(q_j), x - q_j\rangle \geq 1$ *for every* $x \in E$, $j = 1, \dots, d-k$.

(3) $f(q_i) - f(q_j) - \langle G(q_j), q_i - q_j\rangle \geq 1$ *for every* $1 \leq i \neq j \leq d-k$.

Proof. (1) Since $f(q_j) = m(q_j) + 1$, the definition of m leads us to

$$f(q_j) - f(x) - \langle G(x), q_j - x\rangle = m(q_j) - f(x) - \langle G(x), q_j - x\rangle + 1 \geq 1$$

for $x \in E$, $j = 1, \dots, d-k$.

(2) We fix $x \in E$ and $j = 1, \dots, d-k$. The decomposition of m yields

$$m(q_j) = c(P_Y(p_j) + P_Y(T w_j)) + \langle v, q_j\rangle = c(P_Y(p_j)) + \langle v, q_j\rangle = m(p_j) + \langle v, q_j - p_j\rangle.$$

We obtain from this

$$\begin{aligned}
f(x) - f(q_j) - \langle G(q_j), x - q_j\rangle &= m(x) - m(p_j) + \langle v, p_j - q_j\rangle - \langle G(q_j), x - q_j\rangle - 1 \\
&= c \circ (P_Y(x)) + \langle v, x\rangle - m(p_j) + \langle v, p_j - q_j\rangle - \langle v + \varepsilon\alpha w_j, x - q_j\rangle - 1 \\
&= c \circ (P_Y(x)) - m(p_j) + \langle v, p_j\rangle - \varepsilon\alpha\langle w_j, x - q_j\rangle - 1 \\
&= c \circ (P_Y(x)) - m(p_j) + \langle v, p_j\rangle - \varepsilon\alpha\langle w_j, x - p_j\rangle - \varepsilon\alpha\langle w_j, p_j - q_j\rangle - 1 \\
&= c \circ (P_Y(x)) - m(p_j) + \langle v, p_j\rangle - \varepsilon\alpha\langle w_j, x - p_j\rangle + \varepsilon\alpha T - 1.
\end{aligned}$$

Now, using Lemma 6.1(5), the last term is greater than or equal to

$$\begin{aligned}\alpha|P_Y(x)|+\beta-m(p_j)+\langle v,p_j\rangle-\varepsilon\alpha\langle w_j,x-p_j\rangle+\varepsilon\alpha T-1 \\ \geq \alpha|P_Y(x-p_j)|-\alpha|P_Y(p_j)|+\beta-m(p_j)+\langle v,p_j\rangle-\varepsilon\alpha\langle w_j,x-p_j\rangle+\varepsilon\alpha T-1 \\ \geq \alpha|P_Y(x-p_j)|-\varepsilon\alpha\langle w_j,x-p_j\rangle+1,\end{aligned}$$

where the last inequality follows from the choice of T. Now, since $x\in E$, condition (iii) tells us that x does not belong to the cone V_j, which implies that the last term is greater than or equal to

$$\varepsilon\alpha\langle w_j,x-p_j\rangle-\varepsilon\alpha\langle w_j,x-p_j\rangle+1=1.$$

This establishes the inequalities of (2).

(3) Consider $1\leq i\neq j\leq d-k$. Notice that

$$f(q_i)-f(q_j)=c(P_Y(p_i+Tw_i))-c(P_Y(p_j+Tw_j))+\langle v,q_i-q_j\rangle=c(P_Y(p_i))-c(P_Y(p_j))+\langle v,q_i-q_j\rangle.$$

This implies

$$\begin{aligned}f(q_i)-f(q_j)-\langle G(q_j),q_i-q_j\rangle &= c(P_Y(p_i))-c(P_Y(p_j))+\langle v,q_i-q_j\rangle-\langle v+\varepsilon\alpha w_j,q_i-q_j\rangle \\ &= c(P_Y(p_i))-c(P_Y(p_j))-\varepsilon\alpha\langle w_j,p_i-p_j+T(w_i-w_j)\rangle \\ &= c(P_Y(p_i))-c(P_Y(p_j))-\varepsilon\alpha\langle w_j,p_i-p_j\rangle+\varepsilon\alpha T(1-\langle w_i,w_j\rangle)\geq 1,\end{aligned}$$

where the last inequality follows from the choice of T. □

Properties of the new jet. We now define the set $E^*=E\cup\{q_1,\dots,q_{d-k}\}$. Note that we have already extended the definition of (f,G) to E^*.

Lemma 6.3. *We have*:

(1) $X=\operatorname{span}\{G(x)-G(y):x,y\in E^*\}$.

(2) *G is continuous on E^* and $f(x)\geq f(y)+\langle G(y),x-y\rangle$ for all $x,y\in E^*$.*

(3) *$|G(x)|\leq 3K$ for every $x\in E^*$.*

(4) *If $(x_\ell)_\ell$, $(z_\ell)_\ell$ are sequences in E^* such that $(P_X(x_\ell))_\ell$ is bounded and*

$$\lim_{\ell\to\infty}(f(x_\ell)-f(z_\ell)-\langle G(z_\ell),x_\ell-z_\ell\rangle)=0,$$

then $\lim_{\ell\to\infty}|G(x_\ell)-G(z_\ell)|=0$.

Proof. (1) By Lemma 6.1, there are points $x_1,\dots,x_k\in E$ with $Y=\operatorname{span}\{G(x_j)-v:j=1,\dots,k\}$, where v is that of (6-1). Since the vectors $w_1,\dots,w_{d-k}$ are linearly independent, the definitions of (6-2) show that

$$\operatorname{span}\{G(q_j)-v:j=1,\dots,d-k\}=\operatorname{span}\{(\varepsilon\alpha)w_j:j=1,\dots,d-k\}=X\cap Y^\perp.$$

We thus have

$$X=\operatorname{span}\{G(x_1)-v,\dots,G(x_k)-v,G(q_1)-v,\dots,G(q_{d-k})-v\}.$$

For every two points $x, y \in E^*$, we can write

$$G(x) - G(y) = (G(x) - v) - (G(y) - v),$$

but notice that $G(z) - v \in Y = \operatorname{span}\{G(x_i) - v\}_{i=1}^k$ for every $z \in E$ and obviously $G(z) - v \in \operatorname{span}\{G(q_j) - v\}_{j=1}^{d-k}$ if $z \in E^* \setminus E$. This implies $G(x) - G(y) \in X$ for every $x, y \in E^*$. Conversely, if $z \in E^*$, we can write

$$G(z) - v = (G(z) - G(x_1)) + (G(x_1) - v),$$

where the first term belongs to $\operatorname{span}\{G(x) - G(y) : x, y \in E^*\}$ and the second one belongs to $Y = \operatorname{span}\{G(x) - G(y) : x, y \in E\}$. We conclude that $X = \operatorname{span}\{G(x) - G(y) : x, y \in E^*\}$.

(2) The points $q_1, \dots, q_{d-k}$ are distinct and none of them belong to $\bar{E}$. Because G is continuous on E, G is in fact continuous on E^*. Condition (i) of Theorem 1.14 together with Lemma 6.2 tell us that

$$f(x) \geq f(y) + \langle G(y), x - y\rangle \quad \text{for all } x, y \in E^*.$$

(3) From (6-2), $G(q_j) = v + (\varepsilon\alpha)w_j$ for $j = 1, \dots, d-k$. Now Lemma 6.1 tells us that $|v| \leq K$ and $\alpha \leq 2K$, where K denotes $\sup_{y\in E} |G(y)|$. Since $\varepsilon \in (0, 1)$ and the vectors w_j have norm equal to 1, we can write $|G(p_j)| \leq |v| + \alpha \leq 3K$.

(4) Suppose that$(x_\ell)_\ell$, $(z_\ell)_\ell$ are sequences in E^* such that $(P_X(x_\ell))_k$ is bounded and

$$\lim_{\ell\to\infty} (f(x_\ell) - f(z_\ell) - \langle G(z_\ell), x_\ell - z_\ell\rangle) = 0.$$

In view of Lemma 6.2, it is immediate that there exists ℓ_0 such that either there is some $1 \leq j \leq d-k$ with $x_\ell = z_\ell = q_j$ for all $\ell \geq \ell_0$ or else $x_\ell, z_\ell \in E$ for all $\ell \geq \ell_0$. In the first case, the conclusion is trivial. In the second case, $\lim_{\ell\to\infty} |G(x_\ell) - G(z_\ell)| = 0$ follows from condition (iv) of Theorem 1.14. □

We now define

$$m^*(x) = \sup_{y\in E^*} \{f(y) + \langle G(y), x - y\rangle\}$$

for every $x \in E^*$. We already know that $X_{m^*} = \operatorname{span}\{G(x) - G(y) : x, y \in E^*\}$. From Lemma 6.3, $X_{m^*} = X$. The function m^* is convex and $m^* = f$ on E^*. Also, for every $x \in E^*$, we have $G(x) \in \partial m^*(x)$ and, by virtue of Lemma 6.3, m^* is $3K$-Lipschitz on $\mathbb{R}^n$. The function m^* has the decomposition

$$m^* = c^* \circ P_X + \langle v^*, \cdot\rangle \quad \text{on } \mathbb{R}^n, \tag{6-3}$$

where $c^* : X \to \mathbb{R}$ is convex and coercive on X, and $v^* \in \mathbb{R}^n$. With the same proof as that of Lemma 6.1(2), we see that $v^* \in \partial m^*(z_0)$ for some $z_0 \in X$, the function c^* is $6K$-Lipschitz and $|v^*| \leq 3K$. We study the differentiability of c^* in the following lemma, which follows from the corresponding result of the general (not necessarily Lipschitz) case.

Lemma 6.4. *The function c^* is differentiable on $\overline{P_X(E^*)}$, and, if $y \in P_X(E^*)$, then $\nabla c^*(y) = G(x) - v^*$, where $x \in E^*$ is such that $P_X(x) = y$.*

Construction of the extension.

Lemma 6.5. *Let $h : X \to \mathbb{R}$ be a convex, Lipschitz and coercive function such that h is differentiable on a closed subset A of X. There exists $H \in C^1(X)$ convex, Lipschitz and coercive such that $H = h$ and $\nabla H = \nabla h$ on A. Moreover, H can be taken so that* $\operatorname{Lip}(H) \leq M \operatorname{Lip}(h)$, *where $M = M(n) > 0$ is a constant only depending on n.*

Proof. Since h is convex, its gradient ∇h is continuous on A. Then, for all $x, y \in A$, we have

$$0 \leq \frac{h(x) - h(y) - \langle \nabla h(y), x - y\rangle}{|x-y|} \leq \left\langle \nabla h(x) - \nabla h(y), \frac{x-y}{|x-y|} \right\rangle \leq |\nabla h(x) - \nabla h(y)|,$$

where the last term tends to 0 as $|x - y| \to 0$ uniformly on $x, y \in K$ for every compact subset K of A. This shows that the pair $(h, \nabla h)$ defined on A satisfies the conditions of the classical Whitney extension theorem for C^1 functions. Therefore, there exists a function $\tilde{h} \in C^1(X)$ such that $\tilde{h} = h$ and $\nabla \tilde{h} = \nabla h$ on A. In fact, we can arrange $\operatorname{Lip}(\tilde{h}) \leq \kappa \operatorname{Lip}(h)$, where $\kappa = \kappa(n) > 0$ is a constant only depending on n; see [Azagra and Mudarra 2017, Claim 2.3]. Let us define $L = \operatorname{Lip}(h)$.

For each $\varepsilon > 0$, let $\theta_\varepsilon : \mathbb{R} \to \mathbb{R}$ be defined by

$$\theta_\varepsilon(t) = \begin{cases} 0 & \text{if } t \leq 0, \\ t^2 & \text{if } t \leq \frac{1}{2}(L+\varepsilon), \\ (L+\varepsilon)\big(t - \frac{1}{2}(L+\varepsilon)\big) + \big(\frac{1}{2}(L+\varepsilon)\big)^2 & \text{if } t > \frac{1}{2}(L+\varepsilon). \end{cases}$$

Observe that $\theta_\varepsilon \in C^1(\mathbb{R})$, $\operatorname{Lip}(\theta_\varepsilon) = L + \varepsilon$. Now set

$$\Phi_\varepsilon(x) = \theta_\varepsilon(d(x, A)),$$

where $d(x, A)$ stands for the distance from x to A, notice that $\Phi_\varepsilon(x) = d(x, A)^2$ on an open neighborhood of A, and define

$$H_\varepsilon(x) = |\tilde{h}(x) - h(x)| + 2\Phi_\varepsilon(x).$$

Note that $\operatorname{Lip}(\Phi_\varepsilon) = \operatorname{Lip}(\theta_\varepsilon)$ because $d(\cdot\,, A)$ is 1-Lipschitz, and therefore

$$\operatorname{Lip}(H_\varepsilon) \leq \operatorname{Lip}(\tilde{h}) + L + 2(L+\varepsilon) \leq (3+\kappa)L + 2\varepsilon. \tag{6-4}$$

Claim 6.6. *H_ε is differentiable on A, with $\nabla H_\varepsilon(x) = 0$ for every $x \in A$.*

Proof. The proof is the same as that of Claim 4.8. □

Now, because Φ_ε is continuous and positive on $X \setminus A$, by using mollifiers and a partition of unity, one can construct a function $\varphi_\varepsilon \in C^\infty(X \setminus A)$ such that

$$|\varphi_\varepsilon(x) - H_\varepsilon(x)| \leq \Phi_\varepsilon(x) \quad \text{for every } x \in X \setminus A, \tag{6-5}$$

and

$$\operatorname{Lip}(\varphi_\varepsilon) \leq \operatorname{Lip}(H_\varepsilon) + \varepsilon \tag{6-6}$$

(see for instance [Greene and Wu 1979, Proposition 2.1] for a proof in the more general setting of Riemannian manifolds, or [Azagra et al. 2007] for possibly infinite-dimensional Riemannian manifolds).

Let us define $\tilde{\varphi} = \tilde{\varphi}_\varepsilon : X \to \mathbb{R}$ by

$$\tilde{\varphi}(x) = \begin{cases} \varphi_\varepsilon(x) & \text{if } x \in X \setminus A, \\ 0 & \text{if } x \in A. \end{cases}$$

Claim 6.7. *The function $\tilde{\varphi}$ is differentiable on X, and it satisfies $\nabla\tilde{\varphi}(x_0) = 0$ for every $x_0 \in A$.*

Proof. The proof is the same as that of Claim 4.9. □

Note also that

$$\mathrm{Lip}(\tilde{\varphi}) = \mathrm{Lip}(\varphi_\varepsilon) \leq \mathrm{Lip}(H_\varepsilon) + \varepsilon \leq (3+\kappa)L + 3\varepsilon. \tag{6-7}$$

Next we define

$$g = g_\varepsilon := \tilde{h} + \tilde{\varphi}. \tag{6-8}$$

The function g is differentiable on X, and coincides with h on A. Moreover, we also have $\nabla g = \nabla h$ on A (because $\nabla\tilde{\varphi} = 0$ on A). And, for $x \in X \setminus A$, we have

$$g(x) \geq \tilde{h}(x) + H_\varepsilon(x) - \Phi_\varepsilon(x) = \tilde{h}(x) + |h(x) - \tilde{h}(x)| + \Phi_\varepsilon(x) \geq h(x) + \Phi_\varepsilon(x).$$

This shows that $g \geq h$, which in turn implies that g is coercive. Also, notice that according to (6-7) and the definition of g, we have

$$\mathrm{Lip}(g) \leq \mathrm{Lip}(\tilde{h}) + \mathrm{Lip}(\tilde{\varphi}) \leq \kappa L + (3+\kappa)L + 3\varepsilon = (3+2\kappa)L + 3\varepsilon. \tag{6-9}$$

If we define $H = \mathrm{conv}(g)$ we thus get that H is convex on X and $F \in C^1(X)$, with

$$\mathrm{Lip}(H) \leq \mathrm{Lip}(g) \leq (3+2\kappa)L + 3\varepsilon. \tag{6-10}$$

Thus, we can take ε small enough so that $\mathrm{Lip}(H) \leq 2(3+2\kappa)L$. Finally, we know (by an already familiar argument) that $H = h$ and $\nabla H = \nabla h$ on A. Also, because h is a coercive convex function, we have that $H \geq h$ is also coercive. This completes the proof of Lemma 6.5. □

Now we are able to finish the proof of Theorem 1.14. Setting $A := \overline{P_X(E^*)}$, we see from Lemma 6.4 that c^* is differentiable on A. Moreover, since $c^* : X \to \mathbb{R}$ is convex and coercive on X, Lemma 6.5 provides us with a Lipschitz, convex and coercive function H of class $C^1(X)$ such that $(H, \nabla H) = (c^*, \nabla c^*)$ on A and

$$\mathrm{Lip}(H) \leq M\,\mathrm{Lip}(c^*) \leq 6MK,$$

where $M > 0$ is a dimensional constant. Recall that K denotes $\sup_{y\in E} |G(y)|$. We next show that $F := H \circ P_X + \langle v^*, \cdot\rangle$ is the desired extension of (f, G). Since H is $C^1(X)$ and convex, it is clear that F is $C^1(\mathbb{R}^n)$ and convex as well. Because H is coercive on X, it follows (using Theorem 1.11) that $X_F = X$. Also, since $H(y) = c^*(y)$ for $y \in P_X(E)$, we obtain from (6-3) that

$$F(x) = H(P_X(x)) + \langle v^*, x\rangle = c^*(P_X(x)) + \langle v^*, x\rangle = m^*(x) = f(x).$$

Additionally, from the second part of Lemma 6.4, we have, for all $x \in E$, that

$$\nabla F(x) = \nabla H(P_X(x)) + v^* = G(x) - v^* + v^* = G(x).$$

Finally, note that

$$\operatorname{Lip}(F) \le \operatorname{Lip}(H) + |v^*| \le 6MK + 3K = (6M+3)K = (6M+3)\sup_{y\in E} |G(y)|.$$

The proof of Theorem 1.14 is complete.

7. Proof of Theorem 1.16

Let us assume first that there exists such a convex body W, and let us check that N and $P = P_X : \mathbb{R}^n \to X$ satisfy conditions (1)–(4). Define $F : \mathbb{R}^n \to \mathbb{R}$ by

$$F(x) = \theta(\mu_W(x)), \quad x \in \mathbb{R}^n,$$

where $\theta : \mathbb{R} \to [0, +\infty)$ is a C^1 Lipschitz convex function with $\theta(t) = t^2$ whenever $|t| \le 2$ and $\theta(t) = at$ whenever $|t| \ge 2$, for a suitable $a > 0$. We have $\partial W = F^{-1}(1)$, and in particular $F = 1$ on E; additionally

$$N(x) = \frac{\nabla F(x)}{|\nabla F(x)|} \quad \text{for all } x \in E.$$

It is clear that F is a Lipschitz convex function of class $C^1(\mathbb{R}^n)$. Moreover, by elementary properties of the Minkowski functional and the fact that $\nabla F(0) = 0$, we have

$$X_F = \operatorname{span}\{\nabla F(x) : x \in \mathbb{R}^n\} = \operatorname{span}\{\nabla \mu_W(x) : x \in \partial W\} = \operatorname{span}\{n_W(x) : x \in \partial W\} = X.$$

Therefore $(F, \nabla F)$ satisfies conditions (i)–(iv) of Theorem 1.14 on the set $E^* := E \cup \{0\}$ with projection $P = P_X : \mathbb{R}^n \to X$. Then condition (1) follows directly from (i) (or from the fact that W is convex and N is normal to ∂W). In order to check (2), take two sequences $(x_k)_k$, $(z_k)_k$ contained in E with $(P(x_k))_k$ bounded. Now suppose that

$$\lim_{k\to\infty} \langle N(z_k), x_k - z_k\rangle = 0.$$

Then we also have, using $F(x_k) = 1 = F(z_k)$, that

$$\lim_{k\to\infty} \left(F(x_k) - F(z_k) - \langle \nabla F(z_k), x_k - z_k\rangle\right) = 0,$$

and according to (i) of Theorem 1.14 we obtain

$$\lim_{k\to\infty} |\nabla F(x_k) - \nabla F(z_k)| = 0. \tag{7-1}$$

Suppose, seeking a contradiction, that we do not have $\lim_{k\to\infty} |N(x_k) - N(z_k)| = 0$. Then, after possibly passing to subsequences, we may assume that there exists some $\varepsilon > 0$ such that

$$|N(x_k) - N(z_k)| \ge \varepsilon \quad \text{for all } k \in \mathbb{N}.$$

Since $F(x_k) = 1$, $F(0) = 0$ and $\nabla F(x_k) \in X$, the convexity of F yields

$$0 \le F(0) - F(x_k) - \langle \nabla F(x_k), -x_k\rangle = -1 + \langle \nabla F(x_k), x_k\rangle = -1 + \langle \nabla F(x_k), P(x_k)\rangle$$

and this shows that $\inf_k |\nabla F(x_k)| > 0$. Thanks to (7-1), we have $\inf_k |\nabla F(z_k)| > 0$ too and both $(\nabla F(x_k))_k$ and $(\nabla F(z_k))_k$ are bounded above because F is Lipschitz. So we may assume, possibly after extracting subsequences again, that $\nabla F(x_k)$ and $\nabla F(z_k)$ converge, respectively, to vectors $\xi, \eta \in \mathbb{R}^n \setminus \{0\}$. By (7-1) we then get $\xi = \eta$; hence also

$$\varepsilon \leq |N(x_k) - N(z_k)| = \left| \frac{\nabla F(x_k)}{|\nabla F(x_k)|} - \frac{\nabla F(z_k)}{|\nabla F(z_k)|} \right| \to \left| \frac{\xi}{|\xi|} - \frac{\eta}{|\eta|} \right| = 0,$$

a contradiction.

Let us now check (3). Since $0 \in \operatorname{int}(W)$, we can find $r > 0$ such that $B(0, r) \subset W$. Let $y \in E$. If y is parallel to $N(y)$, then $\langle N(y), y\rangle = |y| \geq r$. Otherwise, by convexity of W, the triangle of vertices 0, $rN(y)$ and y, with angles α, β, γ at those vertices, is contained in W, so is the triangle of vertices 0, $rN(y)$, p, where p is the intersection of the line segment $[0, y]$ with the line $L = \{rN(y) + tv : t \in \mathbb{R}\}$, where v is perpendicular to $N(y)$ in the plane $\operatorname{span}\{y, N(y)\}$. Then we have $|p| < |y|$, and $|p| \cos\alpha = r$; hence

$$\langle N(y), y\rangle = |y| \cos\alpha > |p| \cos\alpha = r > 0.$$

Finally condition (4) follows immediately from (iii) of Theorem 1.14 applied with $E^* = E \cup \{0\}$ (and from the fact that $\nabla F(0) = 0$).

Conversely, assume that $N : E \to \mathbb{S}^{n-1}$ and $P = P_X : \mathbb{R}^n \to X$ satisfy (1)–(4), and let us construct a suitable W with the help of Theorem 1.14. Choose r such that

$$0 < r < \inf_{y \in E} \langle N(y), y\rangle, \tag{7-2}$$

and define $E^* = E \cup \{0\}$, $f : E^* \to \mathbb{R}$, $G : E^* \to \mathbb{R}^n$ by

$$f(0) = 0, \qquad f(x) = 1 \quad \text{if } x \in E, \qquad G(0) = 0, \qquad G(x) = \frac{2}{r} N(x) \quad \text{if } x \in E.$$

It is clear that condition (3) implies $\operatorname{dist}(0, E) > 0$; hence the continuity of G on E^* is obvious. As for checking that

$$f(x) - f(y) - \langle G(y), x - y\rangle \geq 0 \quad \text{for all } x, y \in E^*,$$

the only interesting case is that of $x = 0$, $y \in E$, for which we have

$$f(0) - f(y) - \langle G(y), x - y\rangle = -1 + \frac{2}{r}\langle N(y), y\rangle \geq -1 + 2 = 1 > 0.$$

Therefore condition (i) of Theorem 1.14 is fulfilled. Conditions (ii) and (iii) follow immediately from (4). It only remains for us to check (iv). As before, an a priori less trivial situation consists in taking $x_k = 0$, $(z_k)_k \subseteq E$. Note that $(G(z_k))_k$ is always bounded. Assuming that

$$\lim_{k \to \infty} (f(x_k) - f(z_k) - \langle G(z_k), x_k - z_k\rangle) = 0,$$

we get $\lim_{k \to \infty} \langle G(z_k), z_k\rangle = 1$, which implies

$$\lim_{k \to \infty} \langle N(z_k), z_k\rangle = \frac{r}{2},$$

contradicting (7-2). Therefore this situation cannot occur. The rest of cases are immediately dealt with.

Thus we may apply Theorem 1.14 in order to find a convex function $F \in C^1(\mathbb{R}^n)$ such that $(F, \nabla F)$ extends the jet (f, G), and $X_F = X$. We then define $W = F^{-1}((-\infty, 1])$. It is easy to check that W is a (possibly unbounded) convex body of class C^1 such that $E \subset \partial W$, $0 \in \mathrm{int}(W)$, $N(x) = n_W(x)$ for all $x \in E$. Moreover, because $F(0) = 0$ and $\nabla F(0) = 0$, one can see from the proof of Theorem 1.14 that

$$X = \mathrm{span}(\nabla F(E) \cup \{\nabla F(q_1), \dots, \nabla F(q_{d-\ell})\}),$$

where the q_j are such that $F(q_j) \geq 1$ (see Lemma 6.2). In particular, the q_j do not belong to $\mathrm{int}(W)$ and then $\mu_W(q_j) > 0$ for every $j = 1, \dots, d - \ell$. This implies

$$\mathrm{span}(n_W(\partial W)) = \mathrm{span}\{\nabla F(x) : x \in \mathbb{R}^n \setminus \mu_W^{-1}(0)\} \supseteq \mathrm{span}(\nabla F(E) \cup \{\nabla F(q_1), \dots, \nabla F(q_{d-\ell})\}) = X.$$

Since $X_F = X$, this argument shows that $\mathrm{span}(n_W(\partial W)) = X$. □

Acknowledgement

The authors wish to thank the referee for many suggestions that improved the exposition, and for the statement of the stronger version of Corollary 1.4 included in the final version of this paper.

References

[Azagra 2013] D. Azagra, "Global and fine approximation of convex functions", *Proc. Lond. Math. Soc.* (3) **107**:4 (2013), 799–824. MR Zbl

[Azagra and Hajłasz 2017] D. Azagra and P. Hajłasz, "Lusin-type properties of convex functions", preprint, 2017.

[Azagra and Mudarra 2017] D. Azagra and C. Mudarra, "Whitney extension theorems for convex functions of the classes C^1 and $C^{1,\omega}$", *Proc. Lond. Math. Soc.* (3) **114**:1 (2017), 133–158. MR Zbl

[Azagra et al. 2007] D. Azagra, J. Ferrera, F. López-Mesas, and Y. Rangel, "Smooth approximation of Lipschitz functions on Riemannian manifolds", *J. Math. Anal. Appl.* **326**:2 (2007), 1370–1378. MR Zbl

[Azagra et al. 2018] D. Azagra, E. Le Gruyer, and C. Mudarra, "Explicit formulas for $C^{1,1}$ and $C^{1,\omega}_{\mathrm{conv}}$ extensions of 1-jets in Hilbert and superreflexive spaces", *J. Funct. Anal.* **274**:10 (2018), 3003–3032. MR Zbl

[Brudnyi and Shvartsman 2001] Y. Brudnyi and P. Shvartsman, "Whitney's extension problem for multivariate $C^{1,\omega}$-functions", *Trans. Amer. Math. Soc.* **353**:6 (2001), 2487–2512. MR Zbl

[Daniilidis et al. 2018] A. Daniilidis, M. Haddou, E. Le Gruyer, and O. Ley, "Explicit formulas for $C^{1,1}$ Glaeser–Whitney extensions of 1-Taylor fields in Hilbert spaces", *Proc. Amer. Math. Soc.* **146**:10 (2018), 4487–4495. MR Zbl

[Fefferman 2005] C. L. Fefferman, "A sharp form of Whitney's extension theorem", *Ann. of Math.* (2) **161**:1 (2005), 509–577. MR Zbl

[Fefferman 2006] C. Fefferman, "Whitney's extension problem for C^m", *Ann. of Math.* (2) **164**:1 (2006), 313–359. MR Zbl

[Fefferman 2009] C. Fefferman, "Whitney's extension problems and interpolation of data", *Bull. Amer. Math. Soc.* (*N.S.*) **46**:2 (2009), 207–220. MR Zbl

[Fefferman et al. 2016] C. Fefferman, A. Israel, and G. K. Luli, "Finiteness principles for smooth selection", *Geom. Funct. Anal.* **26**:2 (2016), 422–477. MR Zbl

[Fefferman et al. 2017] C. Fefferman, A. Israel, and G. K. Luli, "Interpolation of data by smooth nonnegative functions", *Rev. Mat. Iberoam.* **33**:1 (2017), 305–324. MR Zbl

[Ghomi 2001] M. Ghomi, "Strictly convex submanifolds and hypersurfaces of positive curvature", *J. Differential Geom.* **57**:2 (2001), 239–271. MR Zbl

[Glaeser 1958] G. Glaeser, "Étude de quelques algèbres tayloriennes", *J. Analyse Math.* **6** (1958), 1–124. MR Zbl

[Greene and Wu 1979] R. E. Greene and H. Wu, "C^∞ approximations of convex, subharmonic, and plurisubharmonic functions", *Ann. Sci. École Norm. Sup.* (4) **12**:1 (1979), 47–84. MR Zbl

[Griewank and Rabier 1990] A. Griewank and P. J. Rabier, "On the smoothness of convex envelopes", *Trans. Amer. Math. Soc.* **322**:2 (1990), 691–709. MR Zbl

[Jiménez-Sevilla and Sánchez-González 2013] M. Jiménez-Sevilla and L. Sánchez-González, "On smooth extensions of vector-valued functions defined on closed subsets of Banach spaces", *Math. Ann.* **355**:4 (2013), 1201–1219. MR Zbl

[Kirchheim and Kristensen 2001] B. Kirchheim and J. Kristensen, "Differentiability of convex envelopes", *C. R. Acad. Sci. Paris Sér. I Math.* **333**:8 (2001), 725–728. MR Zbl

[Le Gruyer 2009] E. Le Gruyer, "Minimal Lipschitz extensions to differentiable functions defined on a Hilbert space", *Geom. Funct. Anal.* **19**:4 (2009), 1101–1118. MR Zbl

[Rockafellar 1970] R. T. Rockafellar, *Convex analysis*, Princeton Mathematical Series **28**, Princeton University Press, 1970. MR Zbl

[Schulz and Schwartz 1979] K. Schulz and B. Schwartz, "Finite extensions of convex functions", *Math. Operationsforsch. Statist. Ser. Optim.* **10**:4 (1979), 501–509. MR Zbl

[Whitney 1934] H. Whitney, "Analytic extensions of differentiable functions defined in closed sets", *Trans. Amer. Math. Soc.* **36**:1 (1934), 63–89. MR Zbl

[Yan 2014] M. Yan, "Extension of convex function", *J. Convex Anal.* **21**:4 (2014), 965–987. MR Zbl

Received 4 Sep 2017. Revised 13 Mar 2018. Accepted 30 Jul 2018.

DANIEL AZAGRA: azagra@mat.ucm.es
ICMAT (CSIC-UAM-UC3-UCM), Departamento de Análisis Matemático, Facultad Ciencias Matemáticas, Universidad Complutense de Madrid, Madrid, Spain

CARLOS MUDARRA: carlos.mudarra@icmat.es
ICMAT (CSIC-UAM-UC3-UCM), Madrid, Spain

ANALYSIS AND PDE
Vol. 12, No. 4, 2019
dx.doi.org/10.2140/apde.2019.12.1101

CLASSIFICATION OF POSITIVE SINGULAR SOLUTIONS TO A NONLINEAR BIHARMONIC EQUATION WITH CRITICAL EXPONENT

RUPERT L. FRANK AND TOBIAS KÖNIG

For $n \geq 5$, we consider positive solutions u of the biharmonic equation

$$\Delta^2 u = u^{(n+4)/(n-4)} \quad \text{on } \mathbb{R}^n \setminus \{0\},$$

with a nonremovable singularity at the origin. We show that $|x|^{(n-4)/2} u$ is a periodic function of $\ln |x|$ and we classify all periodic functions obtained in this way. This result is relevant for the description of the asymptotic behavior of local solutions near singularities and for the Q-curvature problem in conformal geometry.

1. Introduction and main results

In this paper we are interested in positive solutions u of the equation

$$\Delta^2 u = u^{(n+4)/(n-4)} \quad \text{in } \mathbb{R}^n \setminus \{0\} \tag{1}$$

for $n \geq 5$. As we will explain later in more detail, this equation serves on one hand as a model problem for higher-order equations with critical nonlinearity and on the other hand has a concrete meaning in the Q-curvature problem in conformal geometry. It is well known that the absence of the maximum principle for equations involving the bi-Laplacian poses great challenges both on a conceptual and on a technical level. Nevertheless we succeed here in proving a classification result for positive solutions of (1) which is completely analogous to its second-order counterpart.

We will work throughout with classical solutions of (1), that is, $u \in C^4(\mathbb{R}^n \setminus \{0\})$. Because of the regularity theory in [Uhlenbeck and Viaclovsky 2000] (which extends that in [Chang et al. 1999] to $n \geq 5$) this is not a restriction.

In the fundamental work [Lin 1998] it was shown that all solutions u with a *removable singularity* at the origin (so that (1) holds in all of $\mathbb{R}^n$) are given by

$$u(x) = c_n \left(\frac{\lambda}{1 + \lambda^2 |x - x_0|^2} \right)^{(n-4)/2}, \quad c_n = \big((n-4)(n-2)n(n+2)\big)^{(n-4)/8}, \tag{2}$$

for some $\lambda > 0$ and $x_0 \in \mathbb{R}^n$. Solutions of the closely related equation $\Delta^2 u = |u|^{8/(n-4)} u$ in $\mathbb{R}^n$ are, in particular, given by optimizers of the Sobolev inequality

$$\int_{\mathbb{R}^n} (\Delta u)^2 \, \mathrm{d}x \geq \mathcal{S}_n \left(\int_{\mathbb{R}^n} |u|^{2n/(n-4)} \, \mathrm{d}x \right)^{(n-4)/n}.$$

Frank was partially supported through the National Science Foundation grant DMS-1363432 .
MSC2010: 34C25, 35B09, 35J30, 53A30.
Keywords: fourth-order equation, critical exponent, classification, periodic solutions.

These optimizers were classified in [Lieb 1983] in an equivalent dual formulation and are again given by constant multiples of the functions in (2). For a classification of positive solutions with removable singularities of the four-dimensional analogue of (1) we refer to [Chang and Yang 1997; Lin 1998] and for the higher-order case to [Wei and Xu 1999; Martinazzi 2009].

In this paper we will be concerned with solutions u of (1) with *nonremovable singularities*. It was also shown by Lin [1998] that such solutions are necessarily radial. We pass to logarithmic coordinates (in this context also known as *Emden–Fowler coordinates*) and write

$$u(x) = |x|^{-(n-4)/2} v(\ln|x|).$$

By a short computation we find that (1) for u is equivalent to the following ordinary differential equation for v:

$$v^{(4)} - \frac{n(n-4)+8}{2} v'' + \frac{n^2(n-4)^2}{16} v - |v|^{8/(n-4)} v = 0 \quad \text{in } \mathbb{R}. \tag{3}$$

Note that positive solutions u of (1) correspond to positive solutions v of (3) and so $|v|^{8/(n-4)}v = v^{(n+4)/(n-4)}$. For some of our results, however, we also need to consider not necessarily positive functions v, and for such functions (3) is the relevant extension. We set

$$a_0 = \left(\frac{n(n-4)}{4}\right)^{(n-4)/4}.$$

Our first main result classifies all positive periodic solutions of (3) and describes their shape.

Theorem 1. (i) *Let $v \in C^4(\mathbb{R})$ be a solution of (3). Then $\inf_{\mathbb{R}} |v| \le a_0$, with equality if and only if v is a nonzero constant.*

(ii) *Let $a \in (0, a_0)$. Then there is a unique (up to translations) bounded solution $v \in C^4(\mathbb{R})$ of (3) with minimal value a. This solution is periodic, has a unique local maximum and minimum per period and is symmetric with respect to its local extrema.*

To state our second main result, we denote by v_a the unique solution to (3) obtained from Theorem 1 by requiring that $v_a(0) = \min_{\mathbb{R}} v_a = a$. Also, denote by L_a the minimal period of v_a. For the constant solution $v_{a_0} \equiv a_0$, we set $L_{a_0} = 0$.

The following theorem provides a classification of positive solutions u of (1) with nonremovable singularities in terms of a two-parameter family.

Theorem 2. *Let $u \in C^4(\mathbb{R}^n \setminus \{0\})$ be a positive solution of (1) whose singularity at the origin is nonremovable. Then there are $a \in (0, a_0]$ and $L \in [0, L_a]$ such that*

$$u(x) = |x|^{-(n-4)/2} v_a(\log|x| + L),$$

where v_a is the solution of (3) introduced after Theorem 1. Moreover, $\partial u/\partial|x| < 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$.

This theorem answers an open question raised in [Guo et al. 2017a] and shows, in particular, that the positivity of the scalar curvature in their conjecture is not necessary.

It is easy to see that as $a \to 0$ one has $L_a \to \infty$ and $v_a(t + L_a/2) \to c_n(2\cosh t)^{-(n-4)/2}$. Undoing the logarithmic change of variables we therefore recover the nonsingular solution (2) in the limit $a \to 0$.

We believe that Theorems 1 and 2 will have several applications. Firstly, they should be a key step in describing the asymptotic behavior near the origin of positive solutions u of $\Delta^2 u = u^{(n+4)/(n-4)}$ in a punctured ball $\{0 < |x| < \rho\}$. This would be the fourth-order analogue of a celebrated result of Caffarelli, Gidas and Spruck [Caffarelli et al. 1989]; see also [Korevaar et al. 1999]. Secondly, we believe that our theorems will prove useful in the construction of constant Q-curvature metrics with isolated singularities in the spirit of the classical works [Schoen 1988; Mazzeo and Pacard 1999] for the scalar curvature; see [Baraket and Rebhi 2002; Guo et al. 2017b] for results in this direction in the fourth-order case. For an introduction to the Q-curvature problem see, for instance, [Hang and Yang 2016].

We end this introduction by comparing the statements and proofs of Theorems 1 and 2 with their second-order counterpart, which concerns positive solutions u of

$$-\Delta u = u^{(n+2)/(n-2)} \quad \text{in } \mathbb{R}^n \setminus \{0\} \tag{4}$$

for $n \geq 3$. A famous result of Caffarelli, Gidas and Spruck [Caffarelli et al. 1989] says that if this equation is valid on all of $\mathbb{R}^n$, then

$$u(x) = c_n' \left(\frac{\lambda}{1 + \lambda^2 |x - x_0|^2} \right)^{(n-2)/2}, \qquad c_n' = (n(n-2))^{(n-2)/4},$$

for some $\lambda > 0$ and $x_0 \in \mathbb{R}^n$. Moreover, they show that if u is a positive solution of (4) with a nonremovable singularity, then u is radial. Using this information, Schoen [1989] observed that all solutions can be classified by standard phase-plane analysis. Indeed, setting

$$u(x) = |x|^{-(n-2)/2} v(\ln |x|)$$

one obtains

$$-v'' + \frac{(n-2)^2}{4} v - v^{(n+2)/(n-2)} = 0 \quad \text{in } \mathbb{R}$$

and the positive solutions of this equation are given by the constant $((n-2)/2)^{(n-2)/2}$, by the homoclinic solution $c_n'(2\cosh(t+T))^{(n-2)/2}$ and by periodic solutions uniquely parametrized, up to translations, by their minimal value in $(0, ((n-2)/2)^{(n-2)/2})$. Moreover, these periodic solutions have a unique local maximum and minimum per period and are symmetric with respect to their local extrema.

Thus, our Theorems 1 and 2 provide exactly the same conclusions as in the second-order case. Their proofs, however, are considerably more difficult, because the phase "plane" in the fourth-order case is four-dimensional. Moreover, solutions to fourth-order equations show, in general, a much richer and typically more erratic behavior than solutions to second-order equations; see, e.g., the introduction of the textbook [Peletier and Troy 2001] for examples. To emphasize the structure of our equation we abbreviate

$$A = \frac{n(n-4)+8}{2}, \qquad B = \frac{n^2(n-4)^2}{16}, \qquad p = \frac{n+4}{n-4}, \tag{5}$$

and

$$f(v) = |v|^{p-1}v - Bv \tag{6}$$

and rewrite (3) as

$$v^{(4)} - Av'' - f(v) = 0 \quad \text{in } \mathbb{R}. \tag{7}$$

Of fundamental importance for us is that the coefficients A and B in (3) satisfy the inequalities

$$A > 0 \quad \text{and} \quad 4B < A^2. \tag{8}$$

These inequalities guarantee that the characteristic equation $\xi^4 - A\xi^2 + B = 0$ associated to the linearization of (7) around the zero solution has four distinct, real solutions. The picture that has emerged from the analysis of fourth-order equations is that under this structural assumption the solution set is better behaved than that of general fourth-order equations and resembles in some sense the solution set of second-order equations; see, e.g., [Peletier and Troy 2001; van den Berg 2000; Buffoni et al. 1996]. The reason is that certain techniques are available which are reminiscent of the maximum principle. Technically, this better, second-order-like behavior can be proved for *bounded* solutions of the equation and for such solutions there are certain substitutes for two-dimensional phase-plane arguments (see, in particular, Propositions 4 and 6). Parts of our analysis will rely on results of van den Berg [2000] for bounded solutions, which in turn rely on results of Buffoni, Champneys and Toland [Buffoni et al. 1996]. Our crucial new ingredient, however, which does not appear in these works, is that *global solutions are necessarily bounded* (Lemma 11). We emphasize that boundedness is a nonlocal property and breaks the local character of the ODE analysis.

Most of our results (except for the explicit expression of the homoclinic solution) hold, mutatis mutandis, for any equation of the form (7) with f given by (6), where $p > 1$ is arbitrary and A and B are arbitrary subject to (8).

2. Classification of global ODE solutions

In this section we will classify all solutions v of (7) which are defined on all of $\mathbb{R}$. Positivity will not play a role here.

We begin with some preliminary remarks, which we will use several times below. The function $v \mapsto f(v)$ in (6) has exactly three zeros, namely, at 0 and at $\pm B^{1/(p-1)} = \pm a_0$. These correspond to exactly three constant solutions. Moreover, if $v(t)$ is a solution to (7), then so are the functions

- $v(-t)$ (because (7) contains only even-order derivatives),
- $-v(t)$ (because f is odd) and
- $v(t+T)$ for any $T \in \mathbb{R}$ (because (7) is autonomous).

We now state the main result of this section.

Proposition 3. *Let $v \in C^4(\mathbb{R})$ be a solution of (7). Then one of the following three alternatives holds*:

(a) $v \equiv \pm B^{1/(p-1)}$, *or* $v \equiv 0$.

(b) $v(t) = \pm c_n (2\cosh(t-T))^{-(n-4)/2}$ *for some* $T \in \mathbb{R}$ *with* c_n *from* (2).

(c) v *is periodic, has a unique local maximum and minimum per period and is symmetric with respect to its local extrema.*

For the proof of this proposition we will need two results, taken from [van den Berg 2000], which quantify the intuition that the set of bounded solutions to the fourth-order equation (7) behaves in some respects similarly to the set of solutions of a second-order equation. As we pointed out in the introduction, for this it is crucial that the relation $4B < A^2$ is satisfied. The first result is that every *bounded* entire solution v is uniquely determined by only *two* (instead of four) initial values.

Proposition 4 [van den Berg 2000, Theorem 1]. *Let* $v, w \in C^4(\mathbb{R})$ *be bounded solutions of* (7) *and suppose that* $v(0) = w(0)$ *and* $v'(0) = w'(0)$. *Then* $v \equiv w$.

Since this result is of crucial importance for us, we give a (slightly more direct) proof with our notation in the Appendix. Proposition 4 has the following consequence.

Corollary 5. *Let* $v \in C^4(\mathbb{R})$ *be a bounded solution of* (7):

(i) *Suppose that* $v'(t_0) = 0$ *for some* $t_0 \in \mathbb{R}$. *Then* v *is symmetric with respect to* t_0; *i.e., for all* $t \in \mathbb{R}$, $v(t_0 + t) = v(t_0 - t)$.

(ii) *Suppose that* $v(t_0) = 0$ *for some* $t_0 \in \mathbb{R}$. *Then* v *is antisymmetric with respect to* t_0; *i.e., for all* $t \in \mathbb{R}$, $v(t_0 - t) = -v(t_0 + t)$.

Proof. (i) Since (7) is autonomous, we may assume $t_0 = 0$. Moreover, if v is a solution, then so is $w(t) := v(-t)$. Thus $v(0) = w(0)$ and, by assumption, $v'(0) = w'(0) = 0$. Proposition 4 gives $v \equiv w$.

(ii) Again, we may assume $t_0 = 0$. Moreover, if v solves (7), then so does $w(t) := -v(-t)$. Since $v(0) = w(0)$ and $v'(0) = w'(0)$, we conclude by Proposition 4 that $v \equiv w$. □

In order to state the second result from [van den Berg 2000] that we need, we introduce

$$F(v) = \int_0^v f(s)\,\mathrm{d}s = \frac{|v|^{p+1}}{p+1} - \tfrac{1}{2}Bv^2$$

as well as the following quantity, also referred to as the *energy*:

$$\mathcal{E}_v(t) = -v'''(t)v'(t) + \tfrac{1}{2}(v''(t))^2 + \tfrac{1}{2}A(v'(t))^2 + F(v(t)).$$

Using (7) one easily finds that for every solution v of (7)

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}_v(t) = 0;$$

that is, the *energy is conserved*. We emphasize that this conservation is a local property and valid on the maximal interval of existence and does not require any a priori boundedness assumptions like Proposition 4 and the following Proposition 6 and Lemma 7.

The second result says that, as in the second-order case, the energy is a parameter which orders bounded solutions in the (v, v')-phase plane.

Proposition 6 [van den Berg 2000, Theorem 2]. *Let $v, w \in C^4(\mathbb{R})$ be bounded solutions of (7) with $v(0) = w(0)$ and either $v'(0) > w'(0) \geq 0$ or $v'(0) < w'(0) \leq 0$. Then $\mathcal{E}_v > \mathcal{E}_w$.*

For the proof we refer to [van den Berg 2000]. The assumption there is satisfied since $4B < A^2$. (Note that no a priori bound on the solutions is necessary for our f.)

Next, we state two lemmas concerning the asymptotic behavior of solutions at infinity.

Lemma 7 [van den Berg 2000, Lemma 4]. *Let $v \in C^4(\mathbb{R})$ be a bounded solution of (7). If v is eventually monotone for $t \to \infty$, then*

$$\lim_{t\to\infty} v(t) \in \{0, \pm B^{1/(p-1)}\} \qquad \text{and} \qquad \lim_{t\to\infty} v^{(k)}(t) = 0 \quad \text{for } k = 1, 2, 3.$$

Similarly, if v is eventually monotone for $t \to -\infty$, then

$$\lim_{t\to-\infty} v(t) \in \{0, \pm B^{1/(p-1)}\} \qquad \text{and} \qquad \lim_{t\to-\infty} v^{(k)}(t) = 0 \quad \text{for } k = 1, 2, 3.$$

The following lemma from [Guo et al. 2017a] shows that (7) does not have a solution which tends to either plus or minus infinity at infinity; that is, solutions that blow up do so in finite time.

Lemma 8 [Guo et al. 2017a, Lemma 2.1]. *Let $v \in C^4(\mathbb{R})$ be a solution of (7). If $a_+ := \lim_{t\to\infty} v(t) \in \mathbb{R} \cup \{\pm\infty\}$ exists, then $a_+ \in \mathbb{R}$. Similarly, if $a_- := \lim_{t\to-\infty} v(t) \in \mathbb{R} \cup \{\pm\infty\}$ exists, then $a_- \in \mathbb{R}$.*

This lemma is proved in [Guo et al. 2017a] for *positive* solutions. An inspection of the proof shows, however, that this positivity is not needed.

We now use the above results to show uniqueness, up to translations, of the positive homoclinic solution. A similar result for $p = 2$ appears in [Amick and Toland 1992] with a different proof.

Lemma 9. *Let $v, w \in C^4(\mathbb{R})$ be positive solutions of (7) with $\lim_{|t|\to\infty} v(t) = \lim_{|t|\to\infty} w(t) = 0$ and $v'(0) = w'(0) = 0$. Then $v \equiv w$.*

Proof. Let us first prove that 0 is the only zero of v' and w'. Indeed, if v' had another zero at, say, $t_0 > 0$, then by repeated application of Corollary 5 (note that by assumption, v is bounded) we deduce that v must be periodic of period $2t_0$. In particular $0 < v(0) = v(2kt_0)$ for all $k \in \mathbb{N}$, which contradicts the assumption that $v(t) \to 0$ as $t \to \infty$. The argument for w is analogous. Hence we must have

$$v'(t) < 0 \quad \text{and} \quad w'(t) < 0 \quad \text{for all } t > 0. \tag{9}$$

Next, by Lemma 7 and by energy conservation,

$$\mathcal{E}_v = \lim_{t\to\infty} \mathcal{E}_v(t) = F(0) = 0 \quad \text{and} \quad \mathcal{E}_w = \lim_{t\to\infty} \mathcal{E}_w(t) = F(0) = 0. \tag{10}$$

If $v(0) = w(0)$, we are done by Proposition 4.

To complete the proof, let us suppose for contradiction that $v(0) > w(0)$. We claim that this implies $v > w$ everywhere. Indeed, otherwise there is $t_0 > 0$ such that $v > w$ on $[0, t_0)$ and $v(t_0) = w(t_0)$. Then by (9) we infer that $v'(t_0) \leq w'(t_0) < 0$. If $v'(t_0) = w'(t_0)$, then Proposition 4 implies $v \equiv w$, contradicting $v(0) > w(0)$. If $v'(t_0) < w'(t_0) < 0$, then Proposition 6 implies $\mathcal{E}_v > \mathcal{E}_w$, which contradicts (10). Hence $v > w$ everywhere.

We can now derive the desired contradiction. For every $R > 0$, we have, using integration by parts and the fact that v and w satisfy (7),

$$\begin{aligned} 0 &= \int_{-R}^{R} w(v^{(4)} - Av'' - f(v)) \\ &= b(R) + \int_{-R}^{R} v(w^{(4)} - Aw'' - f(w)) + \int_{-R}^{R} wv(w^{p-1} - v^{p-1}) \\ &= b(R) + \int_{-R}^{R} wv(w^{p-1} - v^{p-1}). \end{aligned}$$

Here, $b(R)$ contains all the boundary terms coming from the integrations by parts. By Lemma 7 we have $b(R) \to 0$ as $R \to \infty$. But since $\int_{-R}^{R} wv(w^{p-1} - v^{p-1})$ is a negative and strictly decreasing function of R, we obtain a contradiction by choosing R large enough. □

For the concrete values of A, B and p in (5) one can compute the homoclinic solution explicitly. We emphasize that this is the only place in the proof of Proposition 3 where the precise form of A, B and p enters.

Corollary 10. *Suppose that v is a positive solution of (3) with $\lim_{|t|\to\infty} v(t) = 0$. Then there is $T \in \mathbb{R}$ such that*

$$v(t) = c_n(2\cosh(t - T))^{-(n-4)/2}, \quad t \in \mathbb{R},$$

with c_n from (2).

Proof. A straightforward calculation shows that $w(t) = c_n(2\cosh(t))^{-(n-4)/2}$ solves (3). From the assumptions on v it follows that v has a global maximum at some $T \in \mathbb{R}$. Since $v'(T) = 0$, we can apply Lemma 9 to deduce that $v(\cdot + T) = w$. □

The following lemma is one of the key new results in this paper.

Lemma 11. *Let $v \in C^4(\mathbb{R})$ be a solution of (7). Then v is bounded.*

Proof. By replacing $v(t)$ by $v(-t)$, we only need to show that v is bounded on $[0, \infty)$. We consider the set $Z_+ = \{t \geq 0 : v'(t) = 0\}$.

If Z_+ is bounded (in particular, if it is empty), then v is monotone for large t and thus admits a limit a_+ as $t \to \infty$. By Lemma 8, a_+ is finite and therefore v is bounded on $[0, \infty)$.

We now assume that Z_+ is unbounded. Since $F(u) \to \infty$ as $|u| \to \infty$, there is an $R > |v(0)|$ such that $F(u) > \mathcal{E}_v$ for all $|u| \geq R$. We claim that $|v| < R$ on $[0, \infty)$ which, in particular, implies that v is bounded on $[0, \infty)$. Indeed, by contradiction assume that $M_R := \{t \geq 0 : |v(t)| \geq R\}$ is nonempty and define $t^* := \inf M_R$. Since $|v(0)| < R$, we must have $t^* > 0$ and $|v(t^*)| = R$. Replacing $v(t)$ by $-v(t)$ if necessary (which does not change the set Z_+), we may assume that $v(t^*) = R$. Then also $v'(t^*) \geq 0$. Since Z_+ is unbounded, the set $Z_+ \cap [t^*, \infty)$ is nonempty and we can set $T := \inf(Z_+ \cap [t^*, \infty))$. Then $v'(T) = 0$ and $v' \geq 0$ on $[t^*, T]$ by continuity of v'. Thus $v(T) \geq v(t^*) = R$, and we deduce that

$$\mathcal{E}_v(T) = \tfrac{1}{2} v''(T)^2 + F(v(T)) \geq F(v(T)) > \mathcal{E}_v,$$

a contradiction to energy conservation. □

We are now ready to prove the main result of this section.

Proof of Proposition 3. Let $v \in C^4(\mathbb{R})$ be a solution to (7) and set

$$Z := \{t \in \mathbb{R} : v'(t) = 0\}.$$

We distinguish several cases:

Suppose first that $Z = \varnothing$, so v is strictly monotone. We will show that this case cannot occur. Up to replacing $v(t)$ by $v(-t)$, we may assume that v is strictly increasing, and so both limits $a_\pm = \lim_{t\to\pm\infty} v(t)$ exist in $\mathbb{R} \cup \{\pm\infty\}$. By Lemma 8 both limits are finite. By Lemma 7, we are reduced to studying three cases, each of which will lead to a contradiction via an energy argument.

If $a_- = 0$ and $a_+ = B^{1/(p-1)}$, then using Lemma 7 we get $\lim_{t\to-\infty} \mathcal{E}_v(t) = F(0) = 0$, while $\lim_{t\to+\infty} \mathcal{E}_v(t) = F(B^{1/(p-1)}) < 0$, a contradiction to energy conservation. Analogously, a contradiction is obtained if $a_- = -B^{1/(p-1)}$ and $a_+ = 0$.

It remains to consider the case $a_- = -B^{1/(p-1)}$ and $a_+ = B^{1/(p-1)}$. Then as above, by Lemma 7,

$$\lim_{|t|\to\infty} \mathcal{E}_v(t) = F(B^{1/(p-1)}) < 0. \tag{11}$$

On the other hand, by [van den Berg 2000, Corollary 6], the inequality

$$\mathcal{E}_v(t) \geq \tfrac{1}{2} v''(t)^2 + F(v(t)) \tag{12}$$

holds for all $t \in \mathbb{R}$. But now evaluating the energy at t_0 such that $v(t_0) = 0$ gives, together with (12), that $\mathcal{E}_v(t_0) \geq \frac{1}{2} v''(t_0)^2 + F(0) \geq 0$, in contradiction to (11) and energy conservation. Altogether, we have shown that the case $Z = \varnothing$ cannot occur.

If $|Z| = 1$, we may assume, up to a translation, that $Z = \{0\}$. Then v is strictly monotone on $(-\infty, 0)$ and $(0, \infty)$, and so both limits $a_\pm = \lim_{t\to\pm\infty} v(t)$ exist in $\mathbb{R} \cup \{\pm\infty\}$. By Lemma 8 these limits are finite, so v is bounded and, by Corollary 5, even. Therefore $a_+ = a_-$. By Lemma 7, only three cases can occur: $a_+ = a_- = 0$ or $a_+ = a_- = \pm B^{1/(p-1)}$. In the first case, monotonicity implies that either $v > 0$ or $v < 0$, and we conclude that $v(t) = \pm c_n (2\cosh(t))^{-(n-4)/2}$ by Corollary 10.

As for the other cases, let us assume without loss of generality that $a_+ = a_- = B^{1/(p-1)}$ (otherwise replace v by $-v$). We derive a contradiction as follows. Since v is strictly monotone on $[0, \infty)$, $v(0) \neq B^{1/(p-1)}$, and from $\mathcal{E}_v(0) = \frac{1}{2} v''(0)^2 + F(v(0)) \geq F(B^{1/(p-1)})$ we infer that $v(0) = -B^{1/(p-1)}$ (since F attains its global minimal value only at $\pm B^{1/(p-1)}$). Hence v changes sign; i.e., there is $t_0 \in \mathbb{R}$ such that $v(t_0) = 0$. By Corollary 5, v is antisymmetric with respect to t_0. But this is a contradiction to the fact that both a_+ and a_- are positive. Altogether we have thus shown that if $|Z| = 1$, then $v(t) = \pm c_n (2\cosh(t))^{-(n-4)/2}$.

Finally, let us consider the case where $|Z| \geq 2$. By continuity of v', we see that unless v is constant (and hence $v \equiv \pm B^{1/(p-1)}$ or $v \equiv 0$), the closed set Z cannot be dense; i.e., there are real numbers $c < d$ such that $v'(c) = v'(d) = 0$ and $v' \neq 0$ on (c, d). By Lemma 11, v is bounded and therefore we can use Corollary 5 as in the first part of the proof of Lemma 9 to conclude that v must be periodic of period $2(d - c)$. Moreover, since v is strictly monotone on (c, d), there is only one maximum and minimum per period interval, and these are strict. The symmetry with respect to the extrema follows at once from Corollary 5. ☐

We end this section with one more result that will be needed in the proof of Theorem 2.

Lemma 12. *Let $v \in C^4(\mathbb{R})$ be a positive solution of (3). Then*

$$v' < \sqrt{\tfrac{1}{2}A - \sqrt{\left(\tfrac{1}{2}A\right)^2 - B}}\; v.$$

For our values of A and B we have

$$\sqrt{\tfrac{1}{2}A - \sqrt{\left(\tfrac{1}{2}A\right)^2 - B}} = \tfrac{1}{2}(n-4),$$

but the lemma is true for general A and B satisfying (8).

Proof. Because of (8) we can introduce the two positive numbers

$$\lambda = \tfrac{1}{2}A - \sqrt{\left(\tfrac{1}{2}A\right)^2 - B} \quad \text{and} \quad \mu = \tfrac{1}{2}A + \sqrt{\left(\tfrac{1}{2}A\right)^2 - B}. \tag{13}$$

Using $\lambda + \mu = A$ and $\lambda\mu = B$ we can write (3) in terms of the auxiliary function

$$\phi(t) := v''(t) - \lambda v(t)$$

as

$$\phi'' - \mu\phi = v^p. \tag{14}$$

According to Proposition 3, ϕ attains its maximum on $\mathbb{R}$. Since $v > 0$, the maximum principle implies that $\phi < 0$.

The function $w := v'/v$ satisfies

$$w' = -w^2 + \lambda + \frac{\phi}{v}. \tag{15}$$

According to Proposition 3 there is a $t_0 \in \mathbb{R}$ with $v'(t_0) = 0$, and therefore also $w(t_0) = 0$. We shall show now that $M := \{t > t_0 : \ w(t) \geq \sqrt{\lambda}\}$ is empty, which yields the claimed inequality.

Suppose by contradiction that $M \neq \varnothing$ and let $t_1 := \inf M$. It is easy to see that $t_1 > t_0$. Then certainly $w'(t_1) \geq 0$. On the other hand, since $w(t_1) = \sqrt{\lambda}$, (15) implies

$$w'(t_1) = \frac{\phi(t_1)}{v(t_1)} < 0,$$

where the inequality comes from $\phi < 0$ and $v > 0$. This is a contradiction. □

3. Proofs of the main results

3.1. ***Proof of Theorem 1.*** We begin with the proof of part (i) of Theorem 1. Let $v \in C^4(\mathbb{R})$ be a solution of (7). By Proposition 3, the only case where

$$\inf_{\mathbb{R}} |v| \leq B^{1/(p-1)} \tag{16}$$

may fail to hold is when v is periodic. In this case, v possesses a local minimum at, say, $t_0 \in \mathbb{R}$. Note that if v has a zero then (16) is automatically fulfilled, so we may assume that v has a fixed sign and, up to replacing v by $-v$, we may assume that $v > 0$. But by [Guo et al. 2017a, Lemma 2.6], either v is constant (and hence $v \equiv B^{1/(p-1)}$) or $v(t_0) < B^{1/(p-1)}$, so that (16) holds with strict inequality.

We turn now to proving part (ii) of Theorem 1. We proceed via a shooting argument. The value $a \in (0, B^{1/(p-1)})$ will be considered to be fixed throughout the following argument.

For $\beta \geq 0$, we denote by v_β the unique solution of (7) with the initial values

$$v(0) = a, \quad v'(0) = 0, \quad v''(0) = \beta, \quad v'''(0) = 0, \tag{17}$$

and by $T_\beta \in (0, \infty]$ its maximal forward time of existence. Also, let $b := -\min_{v \in \mathbb{R}_+} f(v)$.

Suppose that $\beta > b/A =: \beta_0$. Then we see from

$$v_\beta^{(4)} = A v_\beta'' + f(v_\beta) \tag{18}$$

and (17) that $v_\beta^{(4)} > 0$ initially. Thus, v_β'' increases initially, and since the right-hand side of (18) is positive initially, it is easy to see that it will stay positive on $[0, T_\beta)$. Thus, $v_\beta^{(4)} > 0$ on $[0, T_\beta)$, which implies that v_β and its first three derivatives all keep increasing on $[0, T_\beta)$. Thus, if $T_\beta = \infty$, then v_β is unbounded. On the other hand, if $T_\beta < \infty$, then $v_\beta(t) \to \infty$ as $t \to T_\beta$ (since f is locally Lipschitz). To summarize, v_β increases monotonically on $[0, T_\beta)$ and diverges to $+\infty$ as $t \to T_\beta$ for $\beta \geq \beta_0$.

So we can restrict our search to $\beta \in [0, \beta_0]$. However, for all $\beta \leq \beta_0$, we have the uniform energy bound

$$\mathcal{E}_{v_\beta}(0) = \tfrac{1}{2}\beta^2 + F(a) \leq \tfrac{1}{2}\beta_0^2 + F(a).$$

Since $F(v) \to \infty$ as $v \to \infty$, there is an $R > 0$ such that $F(v) > \frac{1}{2}\beta_0^2 + F(a)$ for all $v > R$. This implies that whenever $\beta \leq \beta_0$ and $v_\beta(t_0) > R$, we must have $v_\beta'(t_0) \neq 0$, for otherwise

$$\mathcal{E}_{v_\beta}(t_0) = \tfrac{1}{2} v_\beta''(t_0)^2 + F(v_\beta(t_0)) \geq F(v_\beta(t_0)) > \tfrac{1}{2}\beta_0^2 + F(a),$$

which contradicts the upper bound on $\mathcal{E}_{v_\beta}(0)$ and energy conservation. In particular, a v_β which enters the interval (R, ∞) cannot leave it again, and hence is certainly not the periodic solution we are looking for.

On the other hand, if $\beta = 0$, we see from (18) that $v_0^{(4)}(0) = f(a) < 0$, and hence $v_0(t)$ and $v_0''(t)$ are strictly decreasing on some small interval $t \in (0, \sigma)$. Since $f(v) < 0$ for $v \in (0, a)$, we deduce from (18) that $v_0^{(k)}(t)$, $k = 1, 2, 3$, stay strictly negative until $v_0(t)$ reaches a negative value. Hence, if $\beta = 0$, there must be t_0 such that $v_0(t_0) < 0$.

All of the previous considerations lead us to defining the shooting sets

$$\begin{aligned} S &:= \{\beta \geq 0 : v_\beta(t) < 0 \text{ for some } t \in (0, T_\beta)\}, \\ T &:= \{\beta \geq 0 : v_\beta(t) > R \text{ for some } t \in (0, T_\beta) \text{ and } v_\beta > 0 \text{ on } [0, t]\}. \end{aligned}$$

Clearly, S and T are open in $[0, \infty)$ because of the continuous dependence of the solution on the initial conditions. Moreover, S and T are disjoint because, as we observed above, once a solution v_β enters the interval (R, ∞), it stays there. We also already argued above that $0 \in S$ and $(\beta_0, \infty) \subset T$; i.e., both $S \neq \varnothing$ and $T \neq \varnothing$.

Since our shooting parameter interval $[0, \infty)$ is connected, we deduce that $S \cup T \neq [0, \infty)$. Hence there must be $\beta^* > 0$ and a corresponding solution $v^* := v_{\beta^*}$ such that $0 \leq v^* \leq R$. In particular, v^* is bounded. This and the fact that f is locally Lipschitz imply that $T_{\beta^*} = \infty$. By even reflection, we obtain a solution defined on all of $\mathbb{R}$, which we still refer to as v^*. Since $\beta^* > 0$, we know v^* has a strict local minimum in 0. By the classification of solutions from Proposition 3, v^* must be periodic. Moreover, it has a unique local maximum and minimum per period and is symmetric with respect to its extrema.

The uniqueness of v^* up to translations follows from Proposition 4. □

3.2. ***Proof of Theorem 2.*** By [Lin 1998, Theorem 4.2], the positivity of u and the nonremovability of the singularity in 0 imply that u is radially symmetric. Since the function v defined by $u(x) = |x|^{-(n-4)/2} v(\ln|x|)$ satisfies (3), we are in a position to apply the classification result from Proposition 3 and we claim that v is either the constant $B^{1/(p-1)} = a_0$ or periodic. Indeed, the only case that remains to be excluded is that $v(t) = c_n (2\cosh(t-T))^{-(n-4)/2}$. But in this case, it is clear that $v(t) \sim c_n e^{t(n-4)/2}$ as $t \to -\infty$ and hence the singularity of u would be removable, contradicting the assumptions. Thus, either v is constant or periodic.

Let $a := \inf v$. Then, by the first part of Theorem 1, $a \in (0, a_0]$, and $a = a_0$ if and only if $v \equiv a_0$. Moreover, for $a < a_0$ the function v is periodic with minimal value a. Therefore, by the second part of Theorem 1, $v(t) = v_a(t+L)$ for some $L \in \mathbb{R}$.

Finally, a simple computation shows that the inequality $\partial u / \partial |x| < 0$ is equivalent to $v' < \frac{1}{2}(n-4)v$, which follows from Lemma 12. □

Appendix: Proof of Proposition 4

In this appendix, we give the proof of Proposition 4, following and simplifying [van den Berg 2000].

Let v and w be bounded solutions of (7) which satisfy $v(0) = w(0)$ and $v'(0) = w'(0)$. We can assume without loss that $v''(0) \geq w''(0)$ (otherwise exchange v and w). We may assume furthermore (up to replacing $v(t)$ and $w(t)$ by $v(-t)$ and $w(-t)$) that $v'''(0) \geq w'''(0)$.

Suppose, by contradiction, that $v \not\equiv w$. Then by uniqueness of ODE solutions, $v^{(k)}(0) \neq w^{(k)}(0)$ for $k = 2$ or $k = 3$. In both cases, we deduce from our hypotheses on the initial conditions that

$$v(t) > w(t) \quad \text{on } (0, \sigma)$$

for some sufficiently small $\sigma > 0$.

With the positive numbers λ and μ from (13) we define the auxiliary functions

$$\phi(t) := v''(t) - \lambda v(t) \quad \text{and} \quad \psi(t) := w''(t) - \lambda w(t).$$

Then by the hypotheses, we have

$$(\phi - \psi)(0) \geq 0 \quad \text{and} \quad (\phi - \psi)'(0) \geq 0. \tag{19}$$

As in (14), equation (7) for v and w implies

$$(\phi - \psi)''(t) - \mu(\phi - \psi)(t) = |v(t)|^{p-1} v(t) - |w(t)|^{p-1} w(t) \quad \text{for all } t \in \mathbb{R}.$$

Since $v(t) > w(t)$ on $(0,\sigma)$ and since the function $u \mapsto |u|^{p-1}u$ is strictly increasing on $\mathbb{R}$, this implies

$$(\phi-\psi)''(t) - \mu(\phi-\psi)(t) > 0 \quad \text{for all } t \in (0,\sigma). \tag{20}$$

The inequalities (19) and (20) and the fact that $\mu > 0$ easily imply $(\phi-\psi)(t) \geq 0$ for $t \in (0,\sigma)$, or equivalently,

$$(v-w)''(t) \geq \lambda(v-w)(t) > 0 \quad \text{for all } t \in (0,\sigma). \tag{21}$$

Since $(v-w)'(0) \geq 0$ by the hypotheses of the lemma and since $\lambda > 0$, we see from (21) that $(v-w)'(t) > 0$ for all $t \in (0,\sigma)$. Hence $v-w$ is strictly increasing on $(0,\sigma)$ and since $\sigma > 0$ was arbitrary with the property that $v-w > 0$ on $(0,\sigma)$, we infer that $v-w$ remains strictly positive for all times.

Repeating the above arguments for the interval $(0,\infty)$ instead of $(0,\sigma)$, we see from (21) that $(v-w)'$ is positive and strictly increasing on $(0,\infty)$. This of course contradicts the boundedness of $v-w$. This proves that in fact we must have $v \equiv w$, concluding the proof of Proposition 4.

References

[Amick and Toland 1992] C. J. Amick and J. F. Toland, "Homoclinic orbits in the dynamic phase-space analogy of an elastic strut", *European J. Appl. Math.* **3**:2 (1992), 97–114. MR Zbl

[Baraket and Rebhi 2002] S. Baraket and S. Rebhi, "Construction of dipole type singular solutions for a biharmonic equation with critical Sobolev exponent", *Adv. Nonlinear Stud.* **2**:4 (2002), 459–476. MR Zbl

[van den Berg 2000] J. B. van den Berg, "The phase-plane picture for a class of fourth-order conservative differential equations", *J. Differential Equations* **161**:1 (2000), 110–153. MR Zbl

[Buffoni et al. 1996] B. Buffoni, A. R. Champneys, and J. F. Toland, "Bifurcation and coalescence of a plethora of homoclinic orbits for a Hamiltonian system", *J. Dynam. Differential Equations* **8**:2 (1996), 221–279. MR Zbl

[Caffarelli et al. 1989] L. A. Caffarelli, B. Gidas, and J. Spruck, "Asymptotic symmetry and local behavior of semilinear elliptic equations with critical Sobolev growth", *Comm. Pure Appl. Math.* **42**:3 (1989), 271–297. MR Zbl

[Chang and Yang 1997] S.-Y. A. Chang and P. C. Yang, "On uniqueness of solutions of nth order differential equations in conformal geometry", *Math. Res. Lett.* **4**:1 (1997), 91–102. MR Zbl

[Chang et al. 1999] S.-Y. A. Chang, M. J. Gursky, and P. C. Yang, "Regularity of a fourth order nonlinear PDE with critical exponent", *Amer. J. Math.* **121**:2 (1999), 215–257. MR Zbl

[Guo et al. 2017a] Z. Guo, X. Huang, L. Wang, and J. Wei, "On Delaunay solutions of a biharmonic elliptic equation with critical exponent", preprint, 2017. arXiv

[Guo et al. 2017b] Z. Guo, J. Wei, and F. Zhou, "Singular radial entire solutions and weak solutions with prescribed singular set for a biharmonic equation", *J. Differential Equations* **263**:2 (2017), 1188–1224. MR Zbl

[Hang and Yang 2016] F. Hang and P. C. Yang, "Lectures on the fourth-order Q curvature equation", pp. 1–33 in *Geometric analysis around scalar curvatures*, edited by F. Han et al., Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap. **31**, World Scientific, Hackensack, NJ, 2016. MR Zbl

[Korevaar et al. 1999] N. Korevaar, R. Mazzeo, F. Pacard, and R. Schoen, "Refined asymptotics for constant scalar curvature metrics with isolated singularities", *Invent. Math.* **135**:2 (1999), 233–272. MR Zbl

[Lieb 1983] E. H. Lieb, "Sharp constants in the Hardy–Littlewood–Sobolev and related inequalities", *Ann. of Math.* (2) **118**:2 (1983), 349–374. MR Zbl

[Lin 1998] C.-S. Lin, "A classification of solutions of a conformally invariant fourth order equation in $\mathbb{R}^n$", *Comment. Math. Helv.* **73**:2 (1998), 206–231. MR Zbl

[Martinazzi 2009] L. Martinazzi, "Classification of solutions to the higher order Liouville's equation on $\mathbb{R}^{2m}$", *Math. Z.* **263**:2 (2009), 307–329. MR Zbl

[Mazzeo and Pacard 1999] R. Mazzeo and F. Pacard, "Constant scalar curvature metrics with isolated singularities", *Duke Math. J.* **99**:3 (1999), 353–418. MR Zbl

[Peletier and Troy 2001] L. A. Peletier and W. C. Troy, *Spatial patterns: higher order models in physics and mechanics*, Progress in Nonlinear Differential Equations and their Applications **45**, Birkhäuser, Boston, 2001. MR Zbl

[Schoen 1988] R. M. Schoen, "The existence of weak solutions with prescribed singular behavior for a conformally invariant scalar equation", *Comm. Pure Appl. Math.* **41**:3 (1988), 317–392. MR Zbl

[Schoen 1989] R. M. Schoen, "Variational theory for the total scalar curvature functional for Riemannian metrics and related topics", pp. 120–154 in *Topics in calculus of variations* (Montecatini Terme, 1987), edited by M. Giaquinta, Lecture Notes in Math. **1365**, Springer, 1989. MR Zbl

[Uhlenbeck and Viaclovsky 2000] K. K. Uhlenbeck and J. A. Viaclovsky, "Regularity of weak solutions to critical exponent variational equations", *Math. Res. Lett.* **7**:5-6 (2000), 651–656. MR Zbl

[Wei and Xu 1999] J. Wei and X. Xu, "Classification of solutions of higher order conformally invariant equations", *Math. Ann.* **313**:2 (1999), 207–228. MR Zbl

Received 2 Nov 2017. Revised 28 May 2018. Accepted 30 Jul 2018.

RUPERT L. FRANK: r.frank@lmu.de, rlfrank@caltech.edu
Mathematisches Institut, Ludwig-Maximilians-Universität München, München, Germany

and

Department of Mathematics, Caltech, Pasadena, CA, United States

TOBIAS KÖNIG: tkoenig@math.lmu.de
Mathematisches Institut, Ludwig-Maximilians-Universität München, München, Germany

ANALYSIS AND PDE
Vol. 12, No. 4, 2019
dx.doi.org/10.2140/apde.2019.12.1115

OPTIMAL MULTILINEAR RESTRICTION ESTIMATES FOR A CLASS OF HYPERSURFACES WITH CURVATURE

IOAN BEJENARU

Bennett, Carbery and Tao (2006) considered the k-linear restriction estimate in $\mathbb{R}^{n+1}$ and established the near optimal $L^{2/(k-1)}$ estimate under transversality assumptions only. In 2017, we showed that the trilinear restriction estimate improves its range of exponents under some curvature assumptions. In this paper we establish almost sharp multilinear estimates for a class of hypersurfaces with curvature for $4 \le k \le n$. Together with previous results in the literature, this shows that curvature improves the range of exponents in the multilinear restriction estimate at all levels of lower multilinearity, that is, when $k \le n$.

1. Introduction

For $n \ge 1$, let $U \subset \mathbb{R}^n$ be an open, bounded and connected neighborhood of the origin and let $\Sigma : U \to \mathbb{R}^{n+1}$ be a smooth parametrization of an n-dimensional submanifold of $\mathbb{R}^{n+1}$ (hypersurface), which we denote by $S = \Sigma(U)$. To this parametrization of S we associate the operator $\mathcal{E}$ defined by

$$\mathcal{E} f(x) = \int_U e^{ix\cdot\Sigma(\xi)} f(\xi)\, d\xi.$$

Given k smooth, compact hypersurfaces $S_i \subset \mathbb{R}^{n+1}$, $i = 1, \dots, k$, where $1 \le k \le n+1$, the k-linear restriction estimate is the inequality

$$\left\| \prod_{i=1}^{k} \mathcal{E}_i f_i \right\|_{L^p(\mathbb{R}^{n+1})} \lesssim \prod_{i=1}^{k} \| f_i \|_{L^2(U_i)}. \tag{1-1}$$

In a more compact format this estimate is abbreviated as

$$\mathcal{R}^*(2 \times \cdots \times 2 \to p).$$

The fundamental question regarding the above estimate is the value of the optimal p for which it holds true. Given that the estimate $\mathcal{R}^*(2 \times \cdots \times 2 \to \infty)$ is trivial, the optimality is translated into the smallest p for which the estimate holds true. Bennett, Carbery and Tao [Bennett et al. 2006] clarified the role of transversality between the surfaces involved and established that, under a transversality condition between $S_1, \dots, S_k$, the optimal exponent is $p = 2/(k-1)$; the actual result in that paper is near-optimal, and the optimal problem is currently open. The optimality can be easily revealed by taking the S_i to be transversal hyperplanes, in which case the estimate becomes the classical Loomis–Whitney inequality.

MSC2010: primary 42B15; secondary 42B25.
Keywords: multilinear restriction estimates, shape operator, wave packets.

It is also known, in some cases (precisely when $k \le 2$), or expected, in most of the others, that curvature assumptions on the surfaces involved improve the range of exponents in (1-1), except for the case $k = n+1$. In [Bejenaru 2017c] we formalized the following conjecture.

Conjecture 1.1. *Under appropriate transversality and curvature conditions on the surfaces S_i, the estimate $\mathcal{R}^*(2 \times \cdots \times 2 \to p)$ holds true for any $p \ge p(k) = 2(n+1+k)/(k(n+k-1))$.*

The case $k = 1$ has been understood for a very long time. Without any curvature assumptions, the optimal exponent is $p = \infty$; once the surface has some nonvanishing principal curvatures, the exponent improves to $p = 2(l+2)/l$, where l is the number of nonvanishing principal curvatures. The case of nonzero Gaussian curvature, corresponding to $l = n$, is the classical result due to Tomas and Stein; see [Stein 1993].

The case $k = 2$ without any curvature assumptions corresponds to the classical L^2 bilinear estimate, where the optimal estimate has been established. Once curvature assumptions are allowed, the best possible exponent in $\mathcal{R}^*(2 \times 2 \to p)$ is $p = (n+3)/(n+1)$ and it was conjectured in [Foschi and Klainerman 2000]. The problem was intensely studied; see [Bourgain 1995; Wolff 2001; Tao 2001; 2003; Tao and Vargas 2000a; Lee 2006; Lee and Vargas 2010; Bejenaru 2017b]. The problem is solved in the regime $p > (n+3)/(n+1)$ for general hypersurfaces with curvature; the end-point $p = (n+3)/(n+1)$ is solved only for cones; see [Tao 2001].

The case $k = n+1$ is fairly well understood. We note that in this case, additional curvature assumptions have no effect on the optimality of p. It is conjectured that if the hypersurfaces $S_i \subset \mathbb{R}^{n+1}$ are transversal, then (1-1) holds true for $p \ge p_0 = 2/n$. If the S_i are transversal hyperplanes, (1-1) is the classical Loomis–Whitney inequality and its proof is elementary. Once the surfaces are allowed to have nonzero principal curvatures, things become far more complicated and the problem has been the subject of extensive research; see, e.g., [Bennett et al. 2006; Guth 2010]. Bennett, Carbery and Tao [Bennett et al. 2006] established a near-optimal version of (1-1), which is (1-1) with an additional R^ϵ factor when the estimate is made over balls of radius R in $\mathbb{R}^{n+1}$. The optimal result for (1-1), that is, without the ϵ-loss, is an open problem. In some cases one can use ϵ-removal techniques to derive the result without the ϵ-loss for $p > 2/n$; see [Bourgain and Guth 2011] for the case of surfaces with nonvanishing Gaussian curvature. The end-point for the multilinear Kakeya version of (1-1) (a slightly weaker statement than (1-1)) was established by Guth [2010] using tools from algebraic topology.

In the remaining cases, $3 \le k \le n$, the k-linear restriction theory has been addressed in [Bennett et al. 2006] only under transversality assumptions and the authors established the near-optimal result for $p \ge 2/(k-1)$. The exponent $2/(k-1)$ is sharp for generic surfaces, but it is not the optimal exponent once curvature assumptions are brought into the problem; indeed note that $p(k) < 2/(k-1)$.

In [Bejenaru 2017c] we looked at the trilinear estimate (corresponding to $k = 3$) and proved the Conjecture 1.1 in the regime $p > p(3)$ for a particular class of surfaces: the double-conic ones. These surfaces have the nice property that they have the exact "amount" of curvature to obtain the estimate with the optimal exponent $p(3)$, and no more, in the sense that they are "flat" in the unnecessary directions.

In this paper we provide the equivalent result for $4 \le k \le n$ for $(k-1)$-conical surfaces. We note that passing from the case $k = 3$ to $k \ge 4$ requires not only additional technical ideas, but also conceptual ones.

We describe bellow the class of hypersurfaces for which we prove the Conjecture 1.1. We start with the definition of a foliation. A $(k-1)$-dimensional foliation of the (n-dimensional) hypersurface S is a decomposition of S into a union of connected disjoint sets $\{S_\alpha\}_{\alpha\in A}$, called the leaves of the foliation, with the following property: every point in S has a neighborhood V and a local system of coordinates $x : V \subset S \to \mathbb{R}^n$ such that for each leaf S_α the coordinates of $V \cap S_\alpha$ are $\xi_k = \text{constant}, \dots, \xi_n = \text{constant}$.

We now formalize the conditions that we impose on our surfaces. As before, $S_i,\ i \in \{1, \dots, k\}$, are hypersurfaces with smooth parametrizations $\Sigma_i : U_i \subset \mathbb{R}^n \to \mathbb{R}^{n+1}$, where each U_i is an open, bounded and connected neighborhood of the origin (note that different U_i may belong to different hyperplanes identified with the same $\mathbb{R}^n$). In addition, we assume the following three hypotheses:

(i) (foliation) For each $i \in \{1, \dots, k\}$, the hypersurface S_i admits the foliation

$$S_i = \bigcup_\alpha S_{i,\alpha},$$

where, for each α, the leaf $S_{i,\alpha}$ is a flat submanifold of dimension $k-1$.

(ii) (the leaves are completely flat) If $S_{N_i(\zeta_i)}$ is the shape operator of S_i at $\zeta_i \in S_i$ with choice of normal $N_i(\zeta_i)$, we assume that for every $v \in T_{\zeta_i} S_{i,\alpha}$ (the tangent plane at $S_{i,\alpha}$ at the point $\zeta_i \in S_{i,\alpha}$) the following holds true:

$$S_{N_i(\zeta_i)} v = 0.$$

(iii) (transversality and curvature) There exists $\nu > 0$ such that for any $\zeta_i \in S_i,\ i \in \{1, \dots, k\}$, for any $l \in \{1, \dots, k\}$ and for any orthonormal basis $v_{k+1}, \dots, v_{n+1} \in (T_{\zeta_l} S_{l,\alpha})^\perp \subset T_{\zeta_l} S_l$ the following holds true:

$$\operatorname{vol}(N_1(\zeta_1), \dots, N_k(\zeta_k), S_{N_l(\zeta_l)} v_{k+1}, \dots, S_{N_l(\zeta_l)} v_{n+1}) \geq \nu. \tag{1-2}$$

In (1-2) vol is the standard volume form of $n+1$ vectors in $\mathbb{R}^{n+1}$; thus the condition quantifies the linear independence of the vectors $N_1(\zeta_1), \dots, N_k(\zeta_k), S_{N_l(\zeta_l)} v_{k+1}, \dots, S_{N_l(\zeta_l)} v_{n+1}$.

The condition (ii) says that the $S_{i,\alpha}$ are, in some sense, completely flat components of the S_i since, besides being subsets of affine planes of dimension $k-1$, the normal $N(\zeta)$ to S_i is constant as we vary ζ along $S_{i,\alpha}$ for fixed α.

The first thing to read in condition (iii) is the transversality condition between $S_1, \dots, S_k$ due to the transversality between any choice on normals. The condition (iii) also says that the submanifolds transversal to the leaves carry the curvature assumptions, in the sense that their tangent space does not contain any eigenvectors of the shape operator. In addition, for each $i \in \{1, \dots, k\}$, we are guaranteed to have transversality between $N_1(\zeta_1), \dots, N_k(\zeta_k)$ and $S_{N_i}(T_{\zeta_i}(S_{l,\alpha})^\perp)$.

In fact (iii) is equivalent to the apparently weaker condition:

(iii′) There exists $\nu > 0$ such that for any $\zeta_i \in S_i,\ i \in \{1, \dots, k\}$, for any $l \in \{1, \dots, k\}$ and for any unit vector $v \in (T_{\zeta_l} S_{l,\alpha})^\perp \subset T_{\zeta_l} S_l$ the following holds true:

$$\operatorname{vol}(N_1(\zeta_1), \dots, N_k(\zeta_k), S_{N_l(\zeta_l)} v) \geq \nu. \tag{1-3}$$

Obviously here vol stands for the $(k+1)$-dimensional volume of the parallelepiped determined by the vectors $N_1(\zeta_1), \dots, N_k(\zeta_k), S_{N_l(\zeta_l)} v$.

At this point we can state the main result of this paper.

Theorem 1.2. *Assume that $S_1, \dots, S_k$ satisfy the conditions (i)–(iii) above. Given any p with*

$$p(k) = \frac{2(n+k+1)}{k(n+k-1)} < p \le \infty,$$

the following holds true:

$$\Big\| \prod_{i=1}^{k} \mathcal{E}_i f_i \Big\|_{L^p(\mathbb{R}^{n+1})} \le C(p) \prod_{i=1}^{k} \|f_i\|_{L^2(U_i)} \quad \textit{for all } f_i \in L^2(U_i). \tag{1-4}$$

To the best of our knowledge this result is the first instance when the k-linear restriction estimate, with $4 \le k \le n$, is proved for the almost optimal exponent, that is, $p > p(k)$. However, very recently Guth [2016a] formulated a weaker version of Conjecture 1.1, which he proved in the case when the S_i are subsets of the paraboloid, and for the same range of parameters $p(k) < p \le \infty$. The formulation of this weaker version is technical and we skip it here. Guth [2016a] used this weaker version to improve the range of the linear restriction theory. It is important to note that Guth employed polynomial partition methods to prove his result. The arguments we use in this paper are very different; see the details below.

The result in Theorem 1.2 and the corresponding one in [Bejenaru 2017c] show that the Conjecture 1.1 holds true at least in some model cases. We hope that this result will lead the way towards a complete resolution of the conjecture, which, in turn, should have important consequences. The multilinear theory discussed above has had major impact in other problems. We mention a few such examples: In harmonic analysis, the bilinear and $(n+1)$-linear restriction theory were used to improve results in the context of the Schrödinger maximal function, see [Bourgain 2013; Lee 2003; Tao and Vargas 2000b; Du et al. 2017], the restriction conjecture, see [Tao 2003; Bourgain and Guth 2011; Guth 2016a; 2016b], and the decoupling conjecture, see [Bourgain and Demeter 2015; Bourgain et al. 2016]. In partial differential equations, the linear theory inspired the Strichartz estimates, see [Tao 2006], while the bilinear restriction theory is used in the context of more sophisticated techniques, such as the profile decomposition, see [Merle and Vega 1998], and concentration compactness methods, see [Kenig and Merle 2006].

Theorem 1.2 reveals the following geometric feature: the optimal k-linear restriction estimate discards the effect of $k-1$ curvatures; indeed, each S_i has precisely $k-1$ vanishing principal curvatures, and thus it relies only on $n+1-k$ principal curvatures being nonzero, although the actual statement has to be more rigorous. This geometric feature of the problem was conjectured by Bennett, Carbery and Tao [Bennett et al. 2006].

We continue with an overview of the paper and highlight some of the elements used in the proof of Theorem 1.2. The reader may look at the paper as split into two parts: Sections 2 through 4 and Sections 6 and 7, with Section 5 marking the transition between the two. In Sections 2 through 4 we adapt to our current setup the standard arguments that are similar to our previous works in the bilinear and trilinear setup, see [Bejenaru 2017b; 2017c]: overview of the geometry of the problem, wave packet theory, table construction and the induction-on-scales argument. All these ideas originate from [Tao 2001].

The second part of the paper, Sections 6 and 7, contains the novel ideas in this paper and they play a key role in establishing the improved estimate (4-4) in Section 4. We note that the equivalent results (to those in Sections 6 and 7 here) in the bilinear and trilinear theory are much simpler, given the structure of the problem, and can be easily derived inside the body of the main argument. The results in Sections 6 and 7 hold in the context of general hypersurfaces; in particular they do not assume the foliation structure or curvature properties used in Theorem 1.2. We also think that these results are new in the literature and may be of independent interest.

The starting ideas originate in the prior work of the author on the multilinear restriction estimate in [Bejenaru 2017a]. In that paper we proved that the k-linear restriction estimate

$$\Big\| \prod_{i=1}^{k} \mathcal{E}_i f_i \Big\|_{L^{2/(k-1)}(B(0,r))} \le C(\epsilon) r^{\epsilon} \prod_{i=1}^{k} \|f_i\|_{L^2(U_i)} \tag{1-5}$$

improves under appropriate localizations of one of the factors f_i. These localizations are precisely the ones carried by the wave packets appearing in the decomposition of one of the factors $\mathcal{E}_i f_i$, and one needs to obtain an appropriate estimate for such superpositions of wave packets. This was an easy task in the case of the trilinear estimate because the estimate is made in L^1 and the triangle inequality holds true. The triangle inequality fails to hold true in the spaces $L^{2/(k-1)}$ with $k \ge 4$; the way to deal with this aspect is to further refine the techniques developed in [Bejenaru 2017a] and derive good "off-diagonal"-type estimates, which in turn give the desired estimate with the correct localization gain; see Theorem 6.1. A further localization to cubes is needed for technical reasons; see Corollary 6.2. This analysis is carried out in Section 6.

In Section 7 we prove the estimate

$$\Big\| \|\mathcal{E}_1 f_1\|_{L^2(S(q))} \prod_{i=2}^{k} \|\mathcal{E}_i f_i\|_{L^2(q)} \Big\|_{l_q^{2/(k-1)}} \le C(\epsilon) r^{\frac{k}{2}} r^{\epsilon} \prod_{i=1}^{k} \|f_i\|_{L^2(U_i)}. \tag{1-6}$$

Here q are cubes of size r and the $l_q^{2/(k-1)}$ is taken over such cubes contained in a larger cube of size r^2; $S(q) = S + q$, where S is a surface with some "good" properties. This estimate has the character of a k-linear restriction estimate, although it is more complex due to the factor $\|\mathcal{E}_1 f_1\|_{L^2(S(q))}$. If S were a point (that is, of dimension zero), then the above estimate is similar to the k-linear restriction estimate; however, the surface we encounter has the maximal dimension that allows (1-6) to hold true. Another interesting aspect is that the maximal dimension of S saturates the estimate (1-6) in the following sense: while for $k < n+1$ (1-5) improves under appropriate localizations of some f_i, (1-6) does not, just as the $(n+1)$-linear restriction estimate does not improve under localizations.

We identified (1-6) as the necessary ingredient to closing the improved estimate (4-4) in Section 4. We note that in the bilinear theory the result used is the one above with $k = 1$ and that means that the term $\prod_{i=2}^{k} \|\mathcal{E}_i f_i\|_{L^2(q)}$ does not appear. In this case the estimate (1-6) corresponds to an energy estimate for a free wave across hypersurfaces that are transversal to its directions of propagation; this is a classical tool in PDE. In the trilinear theory the estimate (1-6) is used for $k = 2$; thus it is an l^2-type estimate that can be dealt with in a direct manner, by using wave packet decompositions for both free waves and some analysis on their interaction. It is in the quadrilinear (or higher) case that the true character of (1-6) comes to light. The analysis of the estimate (1-6) is carried out in Section 7.

1A. *Notation.* We start by clarifying the role of various constants that appear in the argument. N is a large integer that depends only on the dimension. C is a large constant that may change from line to line, and may depend on N, but not on c and C_0 introduced below. C is used in the definition of the following relations: $A \lesssim B$, meaning $A \le CB$, $A \ll B$, meaning $A \le C^{-1}B$, and $A \approx B$, meaning $A \lesssim B \wedge B \lesssim A$. For a given number $r \ge 0$, by $A = O(r)$ we mean that $A \approx r$. C_0 is a constant that is independent of any other constant and its role is to reduce the size of cubes in the inductive argument. We could set $C_0 = 4$ throughout the argument, but we keep it this way so that its role in the argument is not lost. Finally, $c \ll 1$ is a very small variable meant to make expressions $\ll 1$ and most estimates will be stated to hold in a range of c.

We use the standard notation $(\xi_1, \dots, \bar{\xi}_i, \dots, \xi_l) := (\xi_1, \dots, \xi_{i-1}, \xi_{i+1}, \dots, \xi_l)$.

By powers of type $R^{\alpha+}$ we mean $R^{\alpha+\epsilon}$ for arbitrary $\epsilon > 0$. Practically they should be seen as $R^{\alpha+\epsilon}$ for arbitrary $0 < \epsilon \lesssim 1$. The estimates where such powers occur will obviously depend on ϵ.

By $B(x, R)$ we denote the ball centered at x with radius R in the underlying space (most of the time it will be $\mathbb{R}^n$ or $\mathbb{R}^{n+1}$).

Let $\eta_0 : \mathbb{R}^n \to [0, +\infty)$ be a Schwartz function, normalized in L^1, that is, $\|\eta_0\|_{L^1} = 1$, and with Fourier transform supported in the unit ball. Given some $r > 0$ we define $\eta_r(x) = r^{-n}\eta_0(r^{-1}x)$ and note that $\hat{\eta}_r$ is supported in $B(0, r)$. We will abuse notation and use the same η_0 for functions with the same properties, but with a different base space, such as $\eta_0 : \mathbb{R}^{n+1} \to [0, +\infty)$.

A disk $D \subset \mathbb{R}^{n+1}$ has the form

$$D = D(x_D, t_D; r_D) = \{(x, t_D) \in \mathbb{R}^{n+1} : |x - x_D| \le r_D\}$$

for some $(x_D, t_D) \in \mathbb{R}^{n+1}$ and $r_D > 0$. We define the associated smooth cut-off

$$\tilde{\chi}_D(x, t) = \left(1 + \frac{|x - x_D|}{r_D}\right)^{-N}.$$

A cube $Q \subset \mathbb{R}^{n+1}$ of size R has the standard definition

$$Q = \left\{(x, t) \in \mathbb{R}^{n+1} : \|(x - x_Q, t - t_Q)\|_{l^\infty} \le \tfrac{1}{2}R\right\},$$

where $c_Q = (x_Q, t_Q)$ is the center of the cube. Given a constant $\alpha > 0$ we define αQ to be the dilation by α of Q around its center; that is, $\alpha Q = \left\{(x, t) \in \mathbb{R}^{n+1} : \|(x - x_Q, t - t_Q)\|_{l^\infty} \le \alpha \cdot \tfrac{1}{2}R\right\}$.

Given a cube $q \subset \mathbb{R}^{n+1}$ of size r we will use two functions that are highly concentrated in q. One is built with the help of η_0 (as mentioned earlier, we abuse notation here as we should be using the corresponding $\eta_0 : \mathbb{R}^{n+1} \to [0, +\infty)$ with similar properties):

$$\chi_q(x) = \eta_0\left(\frac{x - c(q)}{r}\right).$$

This localization function has nice properties on the Fourier side. The other localization function is

$$\tilde{\chi}_q(x) = \left(1 + \left|\frac{x - c(q)}{r}\right|\right)^{-N}$$

for some large N. This localization has better properties on the physical side.

We recall the standard estimate for superpositions of functions in L^p for $p \le 1$

$$\Big\| \sum_\alpha f_\alpha \Big\|_{L^p}^p \le \sum_\alpha \| f_\alpha \|_{L^p}^p, \tag{1-7}$$

as well as the estimate for sequences

$$\|a_i \cdot b_i\|_{l_i^{2/(k-1)}} \lesssim \|a_i\|_{l_i^2} \|b_i\|_{l_i^{2/(k-2)}}. \tag{1-8}$$

2. Geometry of the surfaces and consequences

We start this section by simplifying the setup. The surfaces are bounded, and therefore we can always break them into smaller (and similar) pieces, as we do to accommodate the additional hypotheses described below.

First note that we can assume each S_i to be of graph type: there is a smooth map $\varphi_i : U_i \subset \mathbb{R}^n \to \mathbb{R}$ such that $S = \{\Sigma_i(\xi) = (\xi, \varphi_i(\xi)) : \xi \in U_i\}$. Here the U_i are open and connected with compact closure. It is less important that the graphs are of type $\zeta_{n+1} = \varphi_i(\zeta_1, \dots, \zeta_n)$ (we can also have $\zeta_k = \varphi_i(\zeta_1, \dots, \bar{\zeta}_k, \dots, \zeta_{n+1})$), although we can accommodate this by a rotation of coordinates. Then each flat leaf $S_{i,\alpha}$ corresponds to a flat leaf $U_{i,\alpha}$, in the sense that $\Sigma_i(U_{i,\alpha}) = S_{i,\alpha}$; this is indeed the case since projections onto hyperplanes along a vector transversal to S_i take $(k-1)$-dimensional affine planes to $(k-1)$-dimensional affine planes.

We can find a system of coordinates $\boldsymbol{x}_i : \mathbb{R}^n \to \mathbb{R}^n$ that parametrizes each leaf $U_{i,\alpha}$ into a new flat leaf $\widetilde{U}_{i,\alpha}$ characterized by $\xi_k =$ constant, $\dots$, $\xi_n =$ constant. Finally, we assume that each U_i has small enough diameter.

Next, we derive a key geometric consequence of our setup. Given a surface S_i we define $\mathcal{N}_i := \{N_i(\zeta_i) : \zeta_i \in S_i\}$ to be the set of normals at S_i. By $\operatorname{dspan} \mathcal{N}_i$ we denote the following subset of the classical span of $\mathcal{N}_i$:

$$\operatorname{dspan} \mathcal{N}_i := \{\alpha N_\alpha + \beta N_\beta : N_\alpha, N_\beta \in \mathcal{N}_i,\ \alpha, \beta \in \mathbb{R}\}.$$

Note that $\operatorname{dspan} \mathcal{N}_i$ is the set of linear combinations of two vectors in $\mathcal{N}_i$; it is not a linear subspace.

Given a set of indexes $I \subset \{1, 2, \dots, k\}$ we also define

$$d\mathcal{N}_I := \{\alpha N_\alpha + \beta N_\beta : N_\alpha \in \mathcal{N}_i,\ N_\beta \in \mathcal{N}_j,\ i, j \in I,\ i \neq j,\ \alpha, \beta \in \mathbb{R}\}.$$

With this notation in place, we claim the following result.

Lemma 2.1. *Assume S_i, $i = 1, \dots, k$, satisfy the conditions (i)–(iii). Let $I = \{3, \dots, k\}$. Then for any $N \in \operatorname{dspan} \mathcal{N}_1$, $N_2 \in \mathcal{N}_2$ and $\widetilde{N} \in d\mathcal{N}_I$, the following holds true*:

$$\operatorname{vol}(N, N_2, \widetilde{N}) \gtrsim |N| \cdot |N_2| \cdot |\widetilde{N}|. \tag{2-1}$$

The above statement is symmetric as we can switch the particular role each S_i, $i = 1, \dots, k$, plays in the above estimate.

Proof. The proof is similar to the one provided in [Bejenaru 2017c]. We write $N = \alpha N_\alpha + \beta N_\beta$ for some $N_\alpha \neq N_\beta$ and consider $\gamma : [0, t_0] \to S_1$, a smooth curve with the property that $N_1(\gamma(0)) = N_\alpha$ and $N_1(\gamma(t_0)) = N_\beta$. We also assume that $|\gamma'(t)| = 1$ on $[0, t_0]$ and that $0 \le t_0 \ll 1$; this is possible because we assumed U_1 to be of small diameter. In addition, if α_0 is such that $\gamma(0) \in S_{1,\alpha_0}$, we can assume that $\gamma'(0) \in (T_{\gamma(0)} S_{1,\alpha_0})^\perp$. Then we have

$$\begin{aligned} N_1(\gamma(t_0)) &= N_1(\gamma(0)) + \int_0^{t_0} S_{N_1(\gamma(s))} \gamma'(s)\, ds \\ &= N_1(\gamma(0)) + t_0 S_{N_1(\gamma(0))} \gamma'(0) + O(t_0^2). \end{aligned}$$

We then continue with

$$\begin{aligned} N &= \alpha N_1(\gamma(0)) + \beta\big(N_1(\gamma(0)) + t_0 S_{N_1(\gamma(0))} \gamma'(0) + O(t_0^2)\big) \\ &= (\alpha+\beta) N_1(\gamma(0)) + \beta t_0 S_{N_1(\gamma(0))} \gamma'(0) + \beta O(t_0^2). \end{aligned}$$

The two vectors $N_1(\gamma(0))$ and $S_{N_1(\gamma(0))}\gamma'(0)$ are transversal; thus $|N| \approx |\alpha+\beta| + t_0|\beta|\,|S_{N_1(\gamma(0))}\gamma'(0)|$ (here we use that $t_0 \ll 1$), and also

$$\begin{aligned} \operatorname{vol}(N, N_2, \tilde{N}) &\approx \operatorname{vol}\big((\alpha+\beta) N_1(\gamma(0)) + \beta t_0 S_{N_1(\gamma(0))}\gamma'(0), N_2, \tilde{N}\big) \\ &\gtrsim |(\alpha+\beta) N_1(\gamma(0)) + \beta t_0 S_{N_1(\gamma(0))}\gamma'(0)| \cdot |\tilde{N}| \approx |N| \cdot |\tilde{N}|, \end{aligned}$$

where we have used the following consequence of (1-2):

$$\operatorname{vol}\big(N_1(\gamma(0)), S_{N_1(\gamma(0))} v, N_2, \tilde{N}\big) \gtrsim |\tilde{N}|,$$

which holds true for any unit vector $v \in (T_{\gamma(0)} S_{1,\alpha_0})^\perp \subset T_{\gamma(0)} S_1$ and any vector $\tilde{N} \in d\mathcal{N}_I$. $\square$

Using an argument similar to that above, one can easily establish the dispersive estimate

$$|N_i(\zeta_1) - N_i(\zeta_2)| \approx d(S_{i,\alpha_1}, S_{i,\alpha_2}), \tag{2-2}$$

where $S_{i,\alpha_1}, S_{i,\alpha_2}$ are the leaves to which ζ_1, ζ_2 belong, respectively. Here the distance between S_{i,α_1} and S_{i,α_2} can be defined either by using geodesics inside the hypersurface S_i (using the induced metric from the ambient space $\mathbb{R}^{n+1}$) or, equivalently, by using the classical distance between sets in $\mathbb{R}^{n+1}$.

3. Free waves, wave packets and tables on cubes

In this section we collect some of the preparatory ingredients that are needed in the proof of our main result. The setup described here originated in the work of Tao [2001] on the bilinear restriction estimate. All of the results here have been discussed in our previous works; see [Bejenaru 2017b; 2017c]. We do not repeat some of the proofs as they are similar to those found in these three mentioned papers.

3A. ***Rephrasing the problem in terms of free waves.*** We reformulate our problem in terms of free waves, this being motivated by the use of wave packets in the proof of Theorem 1.2. Once the wave packet decomposition is made and its properties are clear, the formalization of the problem as an evolution equation can be forgotten.

Assume we are given a surface S with a graph-type parametrization $\zeta_{n+1} = \varphi(\xi)$, where $\xi = (\zeta_1, \dots, \zeta_n)$. We rename the variable ζ_{n+1} by τ; thus the equation of S becomes $\tau = \varphi(\xi)$. We parametrize the physical space by $(x, t) \in \mathbb{R}^n \times \mathbb{R}$. We make the choice that τ is the Fourier variable corresponding to t, while ξ is the Fourier variable corresponding to x. In what follows we use the convention that $\hat{f}$ denotes the Fourier transform of f with respect to the x-variable.

We define the free wave $\phi = \mathcal{E} f$ as

$$\phi(x,t) = \mathcal{E} f(x,t) = \int_{\mathbb{R}^n} e^{i(x\cdot\xi + t\varphi(\xi))} f(\xi)\, d\xi.$$

Note that $\phi(0) = \check{f}$ and $\hat{\phi}(\xi, t) = e^{it\varphi(\xi)}\hat{\phi}(\xi, 0)$. We define the mass of a free wave by $M(\phi(t)) := \|\phi(t)\|_{L^2}^2$ and note that it is time-independent:

$$M(\phi(t)) := \|\phi(t)\|_{L^2}^2 = \|\hat{\phi}(t)\|_{L^2}^2 = \|\hat{\phi}(0)\|_{L^2}^2 = \|\phi(0)\|_{L^2}^2 = M(\phi(0)).$$

The proof of (1-4) relies on estimating $\prod_{i=1}^k \mathcal{E}_i f_i$ on cubes on the physical side and seeing how this behaves as the size of the cubes goes to infinity by using an inductive-type argument with respect to the size of the cubes. Before we formalize this strategy, we note that at every stage of the inductive argument we relocalize functions both on the physical and frequency spaces, and, as a consequence, we need to quantify the new support on the frequency side. This will be done by using the margin of a function.

We assume we are given a reference set V inside which we want to keep all functions supported. If f is supported in $U \subset V$ we define the margin of f relative to V by

$$\operatorname{margin}(f) := \operatorname{dist}(\operatorname{supp}(f), V^c).$$

In terms of free waves $\phi = \mathcal{E} f$, the margin is defined by

$$\operatorname{margin}(\phi(t)) := \operatorname{dist}(\operatorname{supp}_\xi(\hat{\phi}(t)), V^c) = \operatorname{dist}(\operatorname{supp}(f), V^c),$$

where we have used that the Fourier support of $\hat{\phi}(t)$ is time-independent and that $\hat{\phi}(0) = f$. In other words, the margin of a free wave is time-independent.

In practice, we work with k different types of free waves, $\phi_i = \mathcal{E}_i f_i$, $i = 1, \dots, k$. They are assumed to be graphs with different phase functions φ_i and with potentially different ambient domains, that is, the U_i are subsets of different subspaces isomorphic to $\mathbb{R}^n$ (for instance the U_i can be subsets of the hyperplanes $\xi_i = 0$). The above construction changes only by choosing τ to be the coordinate in the direction normal to the ambient hyperplane to which U_i belongs, while ξ are the coordinates in the ambient hyperplane. Obviously, the margin of each ϕ_i is then defined with respect to some V_i in the same ambient hyperplane. When choosing the reference sets V_i we need to impose that the conditions (i)–(iii) hold true on $\Sigma_i(V_i)$ as well.

Next, we prepare the elements that are needed for the induction-on-scale argument. Given that the estimate is trivial for $p = \infty$, it suffices to focus on the result above in the cases $p(k) < p \le 2/(k-1)$ and this is what we will do. Note that the exponent $2/(k-1)$ is precisely the one for which the k-linear restriction theory is expected to hold true without any curvature assumptions.

Definition 3.1. Let $p(k) \leq p \leq 2/(k-1)$. Given $R \geq C_0$ we define $A_p(R)$ to be the best constant for which the estimate

$$\left\| \prod_{i=1}^{k} \phi_i \right\|_{L^p(Q_R)} \leq A_p(R) \prod_{i=1}^{k} M(\phi_i)^{\frac{1}{2}} \tag{3-1}$$

holds true for all cubes Q_R of size R and $\phi_i = \mathcal{E}_i f_i$ obeying the margin requirement

$$\text{margin}^i(\phi_i) \geq M - R^{-\frac{1}{4}}, \quad i = 1, \dots, k. \tag{3-2}$$

The goal is to obtain a uniform estimate on $A_p(R)$ with respect to R. In the absence of the margin requirement above, $A_p(R)$ would be an increasing function. However, since the argument needs to tolerate the margin relaxation, we also define

$$\bar{A}_p(R) := \sup_{1 \leq r \leq R} A_p(r)$$

and the new $\bar{A}_p(R)$ is obviously increasing with respect to R.

Then (1-4), and, as a consequence, the main result of this paper, Theorem 1.2, follow from the next result.

Proposition 3.2. *Assume* $0 < \epsilon < 1$. *If* $R \gg 2^{2C_0}$ *and* $R^{-(1/4)+} \ll c \ll 1$, *there exists* $C(\epsilon)$ *such that the following holds true*:

$$A_p(R) \leq (1+cC)\Big((1+cC)^p \big(\bar{A}_p\big(\tfrac{1}{2}R\big)\big)^p + (C(\epsilon)c^{-C} R^{\frac{n+k+1}{2}(\frac{1}{p} - \frac{k}{2} \cdot \frac{n+k-1}{n+k+1}) + \epsilon})^p\Big)^{\frac{1}{p}}. \tag{3-3}$$

Deriving (1-4) from (3-3) is standard; see the corresponding argument in the trilinear case in [Bejenaru 2017c]. Thus we reduce the proof of Theorem 1.2 to proving (3-3).

3B. ***Tables on cubes.*** Let $Q \subset \mathbb{R}^{n+1}$ be a cube of radius R. Given $j \in \mathbb{N}$ we split Q into $2^{(n+1)j}$ cubes of size $2^{-j}R$ and denote this family by $\mathcal{Q}_j(Q)$; thus we have $Q = \bigcup_{q \in \mathcal{Q}_j(Q)} q$. If $j \in \mathbb{N}$ and $0 \leq c \ll 1$, we define the (c, j)-interior $I^{c,j}(Q)$ of Q by

$$I^{c,j}(Q) := \bigcup_{q \in \mathcal{Q}_j(Q)} (1-c)q. \tag{3-4}$$

Given $j \in \mathbb{N}$ we define a table Φ on Q to be a vector $\Phi = (\Phi^{(q)})_{q \in \mathcal{Q}_j(Q)}$ and define its mass by

$$M(\Phi) = \sum_{q \in \mathcal{Q}_j(Q)} M(\Phi^{(q)}).$$

We define the margin of a table as the minimum margin of its components:

$$\text{margin}(\Phi) = \min_{q \in \mathcal{Q}_j(Q)} \text{margin}(\Phi^{(q)}).$$

We recall from [Bejenaru 2017c] the following result, which originated in [Tao 2001]:

Lemma 3.3. *Assume $0 < p < \infty$, $R \gg 1$, $0 < c \ll 1$ and f is smooth. Given a cube $Q_R \subset \mathbb{R}^{n+1}$ of size R, there exists a cube Q of size $2R$ contained in $4Q_R$ such that*

$$\|f\|_{L^p(Q_R)} \le (1 + cC)\|f\|_{L^p(I^{c,j}(Q))}. \tag{3-5}$$

3C. *Wave packets.* In this section we formalize the wave packet construction for $(k-1)$-conical surfaces. We assume that S is of $(k-1)$-conic type and has the graph-type parametrization $\Sigma : U \to S$, where $\Sigma(\xi) = (\xi, \varphi(\xi))$, with foliations $U = \bigcup_\alpha U_\alpha$, $S = \bigcup_\alpha S_\alpha$, $\Sigma(U_\alpha) = S_\alpha$.

For the foliation $U = \bigcup_\alpha U_\alpha$, we choose a system of coordinates $\boldsymbol{x} : U \to \widetilde{U}$ such that for each leaf U_α the coordinates of U_α are $\xi_k = \text{constant}, \dots, \xi_n = \text{constant}$. Let $\widetilde{U}' = \pi(\widetilde{U})$, where $\pi : \mathbb{R}^n \to \mathbb{R}^{n-k+1}$ is the projection $\pi(\xi_1, \dots, \xi_n) = (\xi_k, \dots, \xi_n)$. Let $\widetilde{\mathcal{L}}$ be a maximal r^{-1}-separated subset of $\widetilde{U}' \subset \mathbb{R}^{n-k+1}$. For each $\tilde{\xi} \in \widetilde{\mathcal{L}}$, we have $\boldsymbol{x}^{-1}(\cdot, \tilde{\xi})$ is a leaf; that is, $\boldsymbol{x}^{-1}(\cdot, \tilde{\xi}) = U_\alpha$ for some α. In each such leaf we pick ξ_T and define $\mathcal{L}$ to be the set obtained this way. It is not important which $\xi_T \in \boldsymbol{x}^{-1}(\cdot, \tilde{\xi})$ is chosen, since from condition (ii) it follows that, for $\xi \in U_\alpha$, the normal $N(\Sigma(\xi))$ to S is constant as ξ varies inside the leaf U_α. We denote by $U(\xi_T)$ the leaf U_α to which ξ_T belongs and by $S(\xi_T) = \Sigma(U(\xi_T))$, the corresponding leaf on S. We note that $d(U(\xi_{T_1}), U(\xi_{T_2})) \approx d(\tilde{\xi}_1, \tilde{\xi}_2)$, which combined with (2-2) gives

$$|N(\Sigma(\xi_{T_1})) - N(\Sigma(\xi_{T_2}))| \approx d(U(\xi_{T_1}), U(\xi_{T_2})) \approx d(\tilde{\xi}_1, \tilde{\xi}_2). \tag{3-6}$$

Let L be the lattice $L = c^{-2} r \mathbb{Z}^n$. With $x_T \in L$, $\xi_T \in \mathcal{L}$, we define the tube

$$T = T(x_T, \xi_T) := \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : |x - x_T + t\nabla\varphi(\xi_T)| \le c^{-2} r\}$$

and denote by $\mathcal{T}$ the set of such tubes. One notices that T is the $c^{-2}r$ neighborhood of the line passing through $(x_T, 0)$ in the direction $N(\Sigma(\xi_T))$.

Associated to a tube $T \in \mathcal{T}$, we define the cut-off $\tilde{\chi}_T$ on $\mathbb{R}^{n+1}$ by

$$\tilde{\chi}_T(x,t) = \tilde{\chi}_{D(x_T - t\nabla\varphi(\xi_T), t; c^{-2}r)}(x).$$

We are ready to state the main result of this section.

Lemma 3.4. *Let Q be a cube of radius $R \gg 1$, let c be such that $R^{-(1/4)+} \ll c \lesssim 1$ and let $J \in \mathbb{N}$ be such that $r = 2^{-J} R \approx R^{1/2}$. Let $\phi = \mathcal{E} f$ be a free wave with* $\operatorname{margin}(\phi) > 0$. *For each $T \in \mathcal{T}$ there is a free wave ϕ_T that is localized in a neighborhood of size $CR^{-1/2}$ of the leaf $S(\xi_T)$ and obeys* $\operatorname{margin}(\phi_T) \ge \operatorname{margin}(f) - CR^{-1/2}$. *The map $f \to \phi_T$ is linear and*

$$\phi = \sum_{T \in \mathcal{T}} \phi_T. \tag{3-7}$$

If $\operatorname{dist}(T, Q) \ge 4R$ *then*

$$\|\phi_T\|_{L^\infty(Q)} \lesssim c^{-C} \operatorname{dist}(T, Q)^{-N} M(\phi)^{\frac{1}{2}}. \tag{3-8}$$

The estimates

$$\sum_T \sup_{q \in Q_J(Q)} \tilde{\chi}_T(x_q, t_q)^{-N} \|\phi_T\|^2_{L^2(q)} \lesssim c^{-C} r M(\phi) \tag{3-9}$$

and

$$\sum_{q_0} M\Big(\sum_T m_{q_0,T}\phi_T\Big) \le (1+cC)M(\phi), \tag{3-10}$$

hold true provided that the coefficients $m_{q_0,T} \ge 0$ *satisfy*

$$\sum_{q_0} m_{q_0,T} = 1 \quad \textit{for all } T \in \mathcal{T}. \tag{3-11}$$

This type of wave packet decomposition was introduced in [Tao 2001] in the context of a bilinear restriction estimate for conical hypersurfaces (1-conical in our language). The strength of this result lies in the use of the small parameter c and the tight mass estimate (3-10). In the case $c \approx 1$, the above decomposition is the standard wave packet decomposition.

In the case of double-conical surfaces the analogous result was proved in [Bejenaru 2017c]. The argument for Lemma 3.4 is entirely similar to the results just mentioned and we will not duplicate it here.

In the case $c \approx 1$, we will use the following variation of (3-9). Fix $N \in \mathbb{N}$; then for each tube $T \in \mathcal{T}$, there are coefficients $c_{N,T}$ such that

$$\sup_{q \in \mathcal{Q}_J(Q)} \tilde{\chi}_T(x_q,t_q)^{-\frac{N}{2}} \|\phi_T\|_{L^2(q)} \lesssim r^{\frac{1}{2}} \cdot c_N(T). \tag{3-12}$$

with the property that

$$\sum_{T \in \mathcal{T}} c_N(T)^2 \lesssim M(\phi). \tag{3-13}$$

4. Table construction and the induction argument

This section contains the main argument for the proof of Theorem 1.2. In Proposition 4.1 we construct tables on cubes, which is a way of reorganizing the information on one term, say ϕ_1, at smaller scales based on information from one of the other interacting terms, $\phi_i, i = 2, \dots, k$. This type of argument is inspired by the work on the conic surfaces of Tao [2001]. Based on this table construction, we will prove the inductive bound claimed in Proposition 3.2.

Proposition 4.1. *Let Q be a cube of size $R \gg 2^{2C_0}$. Assume $\phi_i = \mathcal{E}_i f_i$, $i = 1, \dots, k$, have positive margin. Then there is a table $\Phi_1 = \Phi_c(\phi_1, \phi_2, Q)$ with depth C_0 such that the following properties hold true*:

$$\phi_1 = \sum_{q \in \mathcal{Q}_{C_0}(Q)} \Phi_1^{(q)}, \tag{4-1}$$

$$\operatorname{margin}(\Phi) \ge \operatorname{margin}(\phi) - CR^{-\frac{1}{2}}, \tag{4-2}$$

$$M(\Phi) \le (1+cC)M(\phi), \tag{4-3}$$

and for any $q', q'' \in \mathcal{Q}_{C_0}(Q)$, $q' \ne q''$,

$$\Big\| \Phi_1^{(q')} \prod_{i=2}^k \phi_i \Big\|_{L^{2/(k-1)}((1-c)q'')} \lesssim c^{-C} R^{-\frac{n-k+1}{4}} \prod_{i=1}^k M^{\frac{1}{2}}(\phi_i). \tag{4-4}$$

Remark 1. The above result is stated for scalars $\phi_1, \dots, \phi_k$, but it holds for vector versions as well. Most important is that we can construct $\Phi_1 = \Phi_c(\phi_1, \Phi_2, Q)$, where Φ_2 is a vector-free wave and all its scalar components satisfy similar properties to the ϕ_2 above.

Remark 2. We note that $\Phi_1 = \Phi_c(\phi_1, \phi_2, Q)$ means that the table Φ_1 is constructed from ϕ_1, which is natural in light of (4-1), and ϕ_2. But it does not depend on $\phi_3, \dots, \phi_k$. Obviously, we could have constructed it from ϕ_1 and ϕ_3 (or any other ϕ_k), ending with a different object.

In the proof below we use the results in Sections 6 and 7 in a crucial way. The reason we provide those results in later sections is that, at first reading, it is instructive to get the main points and the motivation for the results in Sections 6 and 7 before the argument becomes too technical.

Proof. There are several scales involved in this argument. The large scale is the size R of the cube Q. The coarse scale is $2^{-C_0} R \gg R^{1/2}$, where this is the size of the smaller cubes in $\mathcal{Q}_{C_0}(Q)$ and the subject of the claims in the proposition. Then there is the fine scale $r = 2^{-j} R$ chosen such that $r \approx R^{1/2}$. Notice that r is the proper scale for wave packets corresponding to time scales R and also that their scale is $c^{-2} r \ll 2^{-C_0} R$, the last one being the scale of cubes in $\mathcal{Q}_{C_0}(Q)$.

We use Lemma 3.4 with $J = j$ to construct the wave packet decomposition for ϕ_1:

$$\phi_1 = \sum_{T_1 \in \mathcal{T}_1} \phi_{1,T_1}.$$

For any $q_0 \in \mathcal{Q}_{C_0}(Q)$ and $T_1 \in \mathcal{T}_1$ we define

$$m_{q_0,T_1} := \|\tilde{\chi}_{T_1} \phi_2\|^2_{L^2(q_0)},$$
$$m_{T_1} := \sum_{q_0 \in \mathcal{Q}_{C_0}(Q)} m_{q_0,T_1}.$$

Based on this we define

$$\Phi_1^{(q_0)} := \sum_{T_1} \frac{m_{q_0,T_1}}{m_{T_1}} \phi_{1,T_1}. \tag{4-5}$$

By combining the definitions above with the decomposition property (3-7), we obtain

$$\phi_1 = \sum_{q_0 \in \mathcal{Q}_{C_0}(Q)} \Phi_1^{(q_0)},$$

thus justifying (4-1).

The margin estimate (4-2) follows from the margin estimate on tubes provided by Lemma 3.4. The coefficients m_{q_0,T_1} satisfy (3-11); thus the estimate (4-3) follows from (3-10).

All that is left to prove is (4-4), which is equivalent to

$$\Bigg(\sum_{\substack{q \in \mathcal{Q}_j(Q) \\ d(q,q_0) \gtrsim cR}} \Big\| \Phi_1^{(q_0)} \prod_{i=2}^{k} \phi_i \Big\|^{\frac{2}{k-1}}_{L^{2/(k-1)}(q)} \Bigg)^{\frac{k-1}{2}} \lesssim c^{-C} r^{-\frac{n-k+1}{2}} \prod_{i=1}^{k} M(\phi_i). \tag{4-6}$$

Note that the cubes q are selected at the finer scale dictated by the size of cubes in $\mathcal{Q}_j(Q)$. In the definition of $\Phi_1^{(q_0)}$, see (4-5), we have the full family $\mathcal{T}_1$. In the above estimate, we estimate the output inside q; thus, in light of (3-8), the terms ϕ_{T_1} with $T_1 \cap q \neq \varnothing$ are the ones that really matter. Indeed, if we split $\Phi_1^{(q_0)}$ as

$$\Phi_1^{(q_0)} := \sum_{T_1 \cap q \neq \varnothing} \frac{m_{q_0,T_1}}{m_{T_1}} \phi_{1,T_1} + \sum_{k \in \mathbb{N}} \sum_{d(T_1,q) \approx 2^k c^{-2} r} \frac{m_{q_0,T_1}}{m_{T_1}} \phi_{1,T_1},$$

we can use (1-7) to reduce the problem to estimating each term in the first sum above. Indeed, in light of (3-9), the contributions of terms from the second sum come with additional decay 2^{-kN}, which, for N large enough, can be easily estimated. Thus it suffices to prove the estimate (4-6) with $\Phi_1^{(q_0)}$ replaced by the first sum above.

For fixed q, it is a straightforward exercise to check that the Setup in Section 6 is satisfied: Simply let $J = \{T_1 \in \mathcal{T}_1 : T_1 \cap q \neq \varnothing\}$ and let $\phi_{T_1} = \mathcal{E}_{1,T_1} f_{1,T_1}$. Thus we can invoke (6-4) to obtain

$$\left\| \left(\sum_{T_1 \cap q \neq \varnothing} \frac{m_{q_0,T_1}}{m_{T_1}} \phi_{T_1} \right) \prod_{i=2}^{k} \phi_i \right\|_{L^{2/(k-1)}(q)} \lesssim C(\epsilon) r^{-\frac{n+1}{2}} r^{\epsilon} \sum_{T_1 \cap q \neq \varnothing} \frac{m_{q_0,T_1}}{m_{T_1}} \|\tilde{\chi}_q \phi_{1,T_1}\|_{L^2} \prod_{i=2}^{k} \|\tilde{\chi}_q \phi_i\|_{L^2}.$$

Since $m_{q_0,T_1} \leq m_{T_1}$,

$$\frac{m_{q_0,T_1}}{m_{T_1}} \leq \frac{m_{q_0,T_1}^{1/2}}{m_{T_1}^{1/2}};$$

from this we obtain

$$\sum_{T_1 \cap q \neq \varnothing} \frac{m_{q_0,T_1}}{m_{T_1}} \|\phi_{1,T_1} \tilde{\chi}_q\|_{L^2} \lesssim \left(\sum_{T_1 \cap q \neq \varnothing} \frac{\|\phi_{1,T_1} \tilde{\chi}_q\|_{L^2}^2}{m_{T_1} \tilde{\chi}_{T_1}(x_q, t_q)} \right)^{\frac{1}{2}} \left(\sum_{T_1 \cap q \neq \varnothing} m_{q_0,T} \tilde{\chi}_{T_1}(x_q, t_q) \right)^{\frac{1}{2}}.$$

Next we claim the estimate

$$\sum_{T_1 \in \mathcal{T}_1} m_{q_0,T_1} \tilde{\chi}_{T_1}(x_q, t_q) \lesssim \|\tilde{\chi}_{S(q)} \phi_2\|_{L^2}^2. \tag{4-7}$$

Using the definition of m_{q_0,T_1} we identify the function

$$\tilde{\chi}_{S(q)} = \left(\sum_{T_1 \in \mathcal{T}_1} \tilde{\chi}(x_q, t_q) \tilde{\chi}_{T_1} \right) \chi_{q_0}$$

which makes (4-7) hold true. Here the surface $S(q)$ is the translation by $c(q)$ of the neighborhood of size r of the cone of normals at S_1, which we denote by $\mathcal{CN}_1 := \{\alpha N_1(\zeta) : \zeta \in S_1,\ \alpha \in \mathbb{R}\}$. It is important to note that we do not consider the whole cone but only the part with $cR \leq \alpha \lesssim R$. Note that $\tilde{\chi}_{S(q)}$ has the following decay property:

$$\tilde{\chi}_{S(q)}(x,t) \lesssim c^{-4} \left(1 + \frac{d((x,t), S(q))}{c^{-2} r} \right)^{-N}.$$

This is a consequence of the fact that the tubes T_1 passing thorough q separate inside q_0 and of the separation between q and q_0, which is quantified by $d(q, q_0) \gtrsim cR$. Quantitatively speaking, given a point in q_0 close to $S(q)$, there are $\lesssim c^{-4}$ tubes T_1 passing through the point and q; this follows from the dispersion estimate (3-6) and the geometry of the family of tubes $\mathcal{T}_1$.

We define

$$A(q)=\Big(\sum_{T_1\cap q\neq\varnothing}\frac{\|\phi_{1,T_1}\tilde{\chi}_q\|_{L^2}^2}{m_{T_1}\tilde{\chi}_{T_1}(x_q,t_q)}\Big)^{\frac{1}{2}},\qquad B(q)=\|\tilde{\chi}_q\phi_2\|_{L^2},$$
$$C(q)=\|\tilde{\chi}_{S(q)}\phi_2\|,\qquad D(q)=\prod_{i=3}^{k}\|\tilde{\chi}_q\phi_i\|_{L^2}.$$

To conclude the proof of (4-6), it suffices to show

$$\Bigg(\sum_{\substack{q\in\mathcal{Q}_j(Q)\\ d(q,q_0)\gtrsim cR}} A(q)^{\frac{2}{k-1}}B(q)^{\frac{2}{k-1}}C(q)^{\frac{2}{k-1}}D(q)^{\frac{2}{k-1}}\Bigg)^{\frac{k-1}{2}}\lesssim r^{\frac{k}{2}}r^{\epsilon}\prod_{i=1}^{k}M(\phi_i)^{\frac{1}{2}}.$$

This will be a consequence of the two inequalities

$$\Bigg(\sum_{\substack{q\in\mathcal{Q}_j(Q)\\ d(q,q_0)\gtrsim cR}} A(q)^2B(q)^2\Bigg)^{\frac{1}{2}}\lesssim r^{\frac{1}{2}}M(\phi_1)^{\frac{1}{2}},\tag{4-8}$$

$$\Bigg(\sum_{\substack{q\in\mathcal{Q}_j(Q)\\ d(q,q_0)\gtrsim cR}} C(q)^{\frac{2}{k-2}}D(q)^{\frac{2}{k-2}}\Bigg)^{\frac{k-2}{2}}\lesssim r^{\frac{k-1}{2}}r^{\epsilon}\prod_{i=2}^{k}M(\phi_i)^{\frac{1}{2}}.\tag{4-9}$$

The proof of (4-8) is similar to the one we used in the bilinear and trilinear theory; see [Bejenaru 2017b; 2017c]. By rearranging the sum, it suffices to show

$$\sum_{T_1}\sum_{q\cap T_1\neq\varnothing}\frac{\|\phi_{1,T_1}\tilde{\chi}_q\|_{L^2}^2\|\phi_2\tilde{\chi}_q\|_{L^2}^2}{m_{T_1}\tilde{\chi}_{T_1}(x_q,t_q)}\lesssim rM(\phi_1).$$

The inner sum is estimated as

$$\sum_{q\cap T_1\neq\varnothing}\frac{\|\phi_2\tilde{\chi}_q\|_{L^2}^2}{m_{T_1}\tilde{\chi}_{T_1}(x_q,t_q)}\lesssim\frac{\|\phi_2\tilde{\chi}_{T_1}\|_{L^2}^2}{m_{T_1}}\lesssim 1,$$

and the outer one is estimated by

$$\sum_{T_1}\sup_q\|\phi_{1,T_1}\tilde{\chi}_q\|_{L^2}^2\lesssim r\sum_{T_1}M(\phi_{1,T_1})\lesssim rM(\phi_1),$$

which is obvious given the size of q in the x_1-direction is $\approx r$ and the mass of ϕ_{1,T_1} is constant across slices in space with $x_1=\text{constant}$.

In proving (4-9), we can take advantage of the fast decay of $\tilde{\chi}_q$ away from q and of $\tilde{\chi}_{S(q)}$ away from $S(q)$, and at the cost of picking factors of type c^{-C}, it suffices to show

$$\Big\|\|\phi_2\|_{L^2(S(q))}\prod_{i=3}^{k}\|\phi_i\|_{L^2(q)}\Big\|_{l_q^{2/(k-2)}}\lesssim r^{\frac{k-1}{2}}r^{\epsilon}\prod_{i=2}^{k}M(\phi_i).\tag{4-10}$$

The $l_q^{2/(k-2)}$ norm is computed over the set of $q \in \mathcal{Q}_j(Q)$, the set of cubes of size r contained in the larger cube of size r^2. This estimate is the subject of Theorem 7.1 in Section 7. The statement of Theorem 7.1 requires S to have certain properties in relation to the other surfaces $S_2, \dots, S_k$; see (P1), (P2) at the beginning of Section 7. The fact that S satisfies these properties follows from Lemma 2.1. □

Proof of Proposition 3.2. This is entirely similar to the argument used in [Bejenaru 2017b; 2017c], see the corresponding proofs there. □

We have finished the proof of our main result Theorem 1.2. Obviously we owe a justification for some estimates used in the body of the proof of Proposition 4.1 and this what will be covered in the next two sections of the paper.

5. The second part: the multilinear estimate revised

We have arrived at the middle point in this article. In the first half, Sections 1 through 4 we proved the main result, Theorem 1.2. In the second part, and Sections 6 and 7, we provide some of the supporting details used in the proof of Theorem 1.2. However we think that these are not just technical results, and they may be of independent interest.

We point out a major difference between the hypotheses used in the two parts. For Theorem 1.2 we assume the particular foliation structure and curvature condition described by conditions (i)–(iii). In the second part, Sections 6 and 7, we provide results in a general setup which we describe below.

We are given k smooth hypersurfaces $S_i = \Sigma_i(U_i)$ with smooth parametrizations Σ_i. These should be seen as new surfaces, different than the ones for which Theorem 1.2 states a result. The most important difference is that the S_i, $i = 1, \dots, k$, used here are generic; in other words they are not assumed to have a foliation structure, nor curvature properties as the surfaces in our main result, Theorem 1.2.

We assume the transversality condition: there exists $\nu > 0$ such that

$$\mathrm{vol}(N_1(\zeta_1), \dots, N_k(\zeta_k)) \geq \nu \tag{5-1}$$

for all choices $\zeta_i \in \Sigma_i(U_i)$. Here by $\mathrm{vol}(N_1(\zeta_1), \dots, N_k(\zeta_k))$ we mean the volume of the k-dimensional parallelepiped spanned by the vectors $N_1(\zeta_1), \dots, N_k(\zeta_k)$.

Each of these (parametrizations of) hypersurfaces generates the corresponding $\mathcal{E}_i$ operator

$$\mathcal{E}_i f(x) = \int_{U_i} e^{ix\cdot\Sigma_i(\xi)} f(\xi)\, d\xi.$$

6. The multilinear estimate: localization and superposition

In this section we provide the proof of a localized version of the multilinear estimate. The motivation comes from the argument in the previous section. The proofs build on the ideas introduced in [Bejenaru 2017a] and later refined in [Bejenaru 2017b].

We work under the setup described in Section 5. Given unit vectors $N_{k+1},\dots,N_{n+1}$, we introduce the following transversality condition: there exists $\nu > 0$ such that

$$|\det(N_1(\zeta_1),\dots,N_k(\zeta_k),N_{k+1},\dots,N_{n+1})| \geq \nu \tag{6-1}$$

for all choices $\zeta_i \in \Sigma_i(U_i)$.

Assume $\Sigma_1(\operatorname{supp} f_1) \subset B(\mathcal{H}_1,\mu)$, where $B(\mathcal{H}_1,\mu)$ is the neighborhood of size μ of the k-dimensional affine subspace $\mathcal{H}_1$. Assume that $|N_1(\zeta_1) - \pi_{\mathcal{H}_1} N_1(\zeta_1)| \lesssim \mu$ for all $\zeta_1 \in \Sigma_1(\operatorname{supp} f_1)$, where $\pi_{\mathcal{H}_1} : \mathbb{R}^{n+1} \to \mathcal{H}_1$ is the projection onto $\mathcal{H}_1$. In addition assume that if N_i, $i = k+1,\dots,n+1$, is a basis of the normal space $\mathcal{H}_1^\perp$ to $\mathcal{H}_1$, then $N_1(\zeta_1),\dots,N_k(\zeta_k),N_{k+1},\dots,N_{n+1}$ are transversal in the sense of (6-1). Under these hypotheses we proved in [Bejenaru 2017b, Theorem 1.3] that

$$\Big\| \prod_{i=1}^k \mathcal{E}_i f_i \Big\|_{L^{2/(k-1)}(B(0,r))} \leq C(\epsilon)\mu^{\frac{n-k+1}{2}} r^\epsilon \prod_{i=1}^k \|f_i\|_{L^2(U_i)}. \tag{6-2}$$

The multilinear estimate (6-2) is a statement about the product of some functions in $L^{2/(k-1)}$. It is very natural to ask how this estimate behaves with respect to superpositions of one factor, that is, replacing f_1 by $\sum_\alpha f_{1,\alpha}$. If $2/(k-1) \geq 1$, then the triangle inequality holds true in $L^{2/(k-1)}$ and the answer is simple: in a sublinear fashion. If $2/(k-1) < 1$, the triangle inequality fails in $L^{2/(k-1)}$ and the sublinearity cannot be argued in the same way. However,

$$\begin{aligned}\Big\| \mathcal{E}_1\Big(\sum_\alpha f_{1,\alpha}\Big) \prod_{i=2}^k \mathcal{E}_i f_i \Big\|_{L^{2/(k-1)}(B(0,R))} &\leq C(\epsilon) R^\epsilon \Big\| \sum_\alpha f_{1,\alpha} \Big\|_{L^2} \prod_{i=2}^k \|f_i\|_{L^2(U_i)} \\ &\leq C(\epsilon) R^\epsilon \sum_\alpha \|f_{1,\alpha}\|_{L^2} \prod_{i=2}^k \|f_i\|_{L^2(U_i)}\end{aligned}$$

and this indicates again sublinear behavior with respect to superpositions of one input. In the above the set of indexes α is taken to be of finite cardinality (to avoid unnecessary distractions) and the key point is that the estimate is independent of the cardinality of this set.

The main question is whether the sublinearity aspect of the estimate holds true for the refinement (6-2) of the multilinear estimate. An a posteriori argument as above fails to give the optimal result when each term $f_{1,\alpha}$ has good localization properties, but $\sum_\alpha f_{1,\alpha}$ does not have such localization properties.

Setup. We are given J, a finite set, and open, bounded and connected sets $U_{1,\alpha} \subset \mathcal{H}_{1,\alpha}$ for all $\alpha \in J$, where $\mathcal{H}_{1,\alpha}$ are affine hyperplanes. For each $\alpha \in J$ we assume the following: there are k-dimensional hyperplanes $\mathcal{H}'_{1,\alpha}$ with the property that $S_{1,\alpha} = \Sigma_{1,\alpha}(U_{1,\alpha}) \subset B(\mathcal{H}'_{1,\alpha},\mu)$, where $B(\mathcal{H}'_{1,\alpha},\mu)$ is the neighborhood of size μ of $\mathcal{H}'_{1,\alpha}$. The following property holds: $|N_1(\zeta_1) - \pi_{\mathcal{H}'_{1,\alpha}} N_1(\zeta_1)| \lesssim \mu$ for all $\zeta_1 \in S_{1,\alpha}$, where $\pi_{\mathcal{H}'_{1,\alpha}} : \mathbb{R}^{n+1} \to \mathcal{H}'_{1,\alpha}$ is the projection onto $\mathcal{H}'_{1,\alpha}$. Let $\widetilde{\mathcal{H}}_{1,\alpha} = \mathcal{H}_{1,\alpha} \cap \mathcal{H}'_{1,\alpha}$ be the $(k-1)$-dimensional affine subspace $\widetilde{\mathcal{H}}_{1,\alpha} \subset \mathcal{H}_{1,\alpha}$; we also assume that $U_{1,\alpha} \subset B(\widetilde{\mathcal{H}}_{1,\alpha},\mu)$.

We assume that $S_{1,\alpha} \subset S_1 = \Sigma_1(U_1)$ for all $\alpha \in J$, and S_1 satisfies the following global property: there is an orthonormal set of vectors N_i, $i = k,\dots,n+1$, such that (6-1) is satisfied.

For each $\alpha \in J$, we assume that if N_i, $i = k+1,\dots,n+1$, is a basis of the normal space $\widetilde{\mathcal{H}}_{1,\alpha}^\perp \subset \mathcal{H}_{1,\alpha}$, then $N_1(\zeta_1),\dots,N_k(\zeta_k),N_{k+1},\dots,N_{n+1}$ are transversal in the sense of (6-1).

For each $\alpha \in J$ we define

$$\mathcal{E}_{1,\alpha} f(x) = \int_{U_{1,\alpha}} e^{ix\cdot\Sigma_{1,\alpha}(\xi)} f(\xi)\, d\xi.$$

Without restricting the generality of the problem, we can assume that $S_{1,\alpha}$ are of graph type; that is, $\Sigma_{1,\alpha}(\xi^\alpha) = (\xi^\alpha, \varphi_{1,\alpha}(\xi^\alpha))$, where ξ^α is the coordinate in $\mathcal{H}_{1,\alpha}$. In addition, for each α, we pick and fix some $\eta_{1,\alpha} \in U_{1,\alpha}$.

The next result states how the multilinear estimate behaves with respect to superposition of localized functions.

Theorem 6.1. *We assume the Setup above. Let $\mu, R > 0$ be such that $R \le \mu^{-1}$. Then for any $\epsilon > 0$, there is $C(\epsilon)$ such that the following holds true*:

$$\left\| \left(\sum_\alpha \mathcal{E}_{1,\alpha} f_{1,\alpha} \right) \prod_{i=2}^{k} \mathcal{E}_i f_i \right\|_{L^{2/(k-1)}(B(0,R))} \le C(\epsilon) \mu^{\frac{n+1-k}{2}} R^\epsilon \left(\sum_\alpha \| f_{1,\alpha} \|_{L^2(U_{1,\alpha})} \right) \prod_{i=2}^{k} \| f_i \|_{L^2(U_i)}. \tag{6-3}$$

In Section 4 we used the following consequence of the above theorem.

Corollary 6.2. *We assume the Setup above. Assume that $\mu \approx r^{-1}$ and q is a cube of size $\approx r$. Then for any $\epsilon > 0$, there is $C(\epsilon)$ such that the following holds true*:

$$\left\| \left(\sum_\alpha \mathcal{E}_{1,\alpha} f_{1,\alpha} \right) \prod_{i=2}^{k} \mathcal{E}_i f_i \right\|_{L^{2/(k-1)}(q)} \le C(\epsilon) r^{-\frac{n+1}{2}} r^\epsilon \left(\sum_\alpha \| \tilde{\chi}_q \mathcal{E}_{1,\alpha} f_{1,\alpha} \|_{L^2} \right) \prod_{i=2}^{k} \| \tilde{\chi}_q \mathcal{E}_i f_i \|_{L^2}. \tag{6-4}$$

We note that the apparent gain of a factor of $r^{-k/2}$ in this corollary over the result in Theorem 6.1 has to do with replacing $\| f_i \|_{L^2(U_i)}$ by $\| \tilde{\chi}_q \mathcal{E}_i f_i \|_{L^2}$.

The result of the corollary is not an immediate consequence of Theorem 6.1; but it follows easily from the arguments used in the proof of Theorem 6.1.

The plan is the following: we introduce some notation specific to this section and then we proceed with the proof of the above two results.

6A. ***Notation.*** Assume $\mathcal{H}_1 \subset \mathbb{R}^{n+1}$ is a hyperplane passing through the origin. Let N_1 be its normal and let $\pi_{N_1} : \mathbb{R}^{n+1} \to \mathcal{H}_1$ be the associated projection along the normal N_1. We denote by $\mathcal{F}_1 : \mathcal{H}_1 \to \mathcal{H}_1$ the Fourier transform and by $\mathcal{F}_1^{-1}$ the inverse Fourier transform. We denote the variables in $\mathbb{R}^{n+1}$ by $x = (x_1, x')$, where x_1 is the coordinate along N_1 and x' is the coordinate along $\mathcal{H}_1$. We denote by ξ' the Fourier variable corresponding to x'. For $f : U_1 \subset \mathcal{H}_1 \to \mathbb{C}$, $f \in L^2(U_1)$, the operator $\mathcal{E}_1$ takes the form

$$\mathcal{E}_1 f(x) = \int_{U_1} e^{i(x'\xi' + x_1 \varphi_1(\xi'))} f(\xi')\, d\xi'. \tag{6-5}$$

We define the differential operator $\nabla\varphi_1(D'/i)$ to be the operator with symbol $\nabla\varphi_1(\xi')$. The following commutator estimate holds true:

$$\left(x' - x_0' - x_1 \nabla\varphi_1\left(\frac{D'}{i}\right)\right)^N \mathcal{E}_1 f = \mathcal{E}_1(\mathcal{F}_1((x' - x_0')^N \mathcal{F}_1^{-1} f)) \quad \text{for all } N \in \mathbb{N}. \tag{6-6}$$

This is a direct computation using (6-5) and it suffices to check it for $N = 1$. The role of (6-6) will be to quantify localization properties of $\mathcal{F}_1^{-1} f$ on hyperplanes with $x_1 =$ constant.

We take $\mathcal{H}_i,\ i = 1, \dots, k$, to be reference hyperplanes that are used in defining $\mathcal{E}_i f_i,\ i = 1, \dots, k$. Their normals are denoted by $N_i,\ i = 1, \dots, k$, respectively. Note that since $S_{1,\alpha} \subset S_1$ for all $\alpha \in J$, it follows that N_1 is transversal to all $\mathcal{H}_{1,\alpha}$. We then pick unit vectors $N_{k+1}, \dots, N_{n+1}$ such that (6-1) is satisfied.

We construct $\mathcal{L} := \{z_1 N_1 + \dots + z_{n+1} N_{n+1} : (z_1, \dots, z_{n+1}) \in \mathbb{Z}^{n+1}\}$ to be the oblique lattice in $\mathbb{R}^{n+1}$ generated by the unit vectors $N_1, \dots, N_{n+1}$. In each $\mathcal{H}_i,\ i = 2, \dots, k$, we construct the induced lattice $\mathcal{L}(\mathcal{H}_i) = \pi_{N_i}(\mathcal{L})$; this is a lattice since the projection is taken along a direction of the original lattice $\mathcal{L}$.

Given $r > 0$ we define $\mathcal{C}(r)$ to be the set of parallelepipeds of size r in $\mathbb{R}^{n+1}$ relative to the lattice $\mathcal{L}$; a parallelepiped in $\mathcal{C}(r)$ has the form

$$q(\boldsymbol{j}) := \left[r\left(j_1 - \tfrac{1}{2}\right), r\left(j_1 + \tfrac{1}{2}\right)\right] \times \dots \times \left[r\left(j_{n+1} - \tfrac{1}{2}\right), r\left(j_{n+1} + \tfrac{1}{2}\right)\right],$$

where $\boldsymbol{j} = (j_1, \dots, j_{n+1}) \in \mathbb{Z}^{n+1}$. For such a parallelepiped we define $c(q) = r\,\boldsymbol{j} = (r j_1, \dots, r j_{n+1}) \in r\mathcal{L}$ to be its center. For each $i = 2, \dots, k$, we let $\mathcal{CH}_i(r) = \pi_{N_i}\mathcal{C}(r)$ be the set of parallelepipeds of size r in the hyperplane $\mathcal{H}_i$. Given two parallelepipeds $q, q' \in \mathcal{C}(r)$ or $\mathcal{CH}_i(r)$ we define $d(q, q')$ to be the distance between them when considered as subsets of the underlying space, which we take to be $\mathbb{R}^{n+1}$ or $\mathcal{H}_i$.

For each $i \in \{2, \dots, k\},\ r > 0$, we define the linear operator $\mathcal{T}_i : \mathcal{H}_i \to \mathcal{H}_i$ to be the operator that takes $\mathcal{L}(\mathcal{H}_i)$ to the standard lattice $\mathbb{Z}^n$ in $\mathcal{H}_i$. Then for each $q \in \mathcal{CH}_i(r)$, define $\chi_q : \mathcal{H}_i \to \mathbb{R}$ by

$$\chi_q(x) = \eta_0\left(\mathcal{T}_i\left(\frac{x - c(q)}{r}\right)\right).$$

Notice that $\mathcal{F}_i \chi_q$ has Fourier support in the ball of radius $\lesssim r^{-1}$. By the Poisson summation formula and properties of η_0,

$$\sum_{q \in \mathcal{CH}_i(r)} \chi_q = 1. \tag{6-7}$$

Using the properties of χ_q, a direct exercise shows that for each $N \in \mathbb{N}$, the following holds true:

$$\sum_{q \in \mathcal{CH}_i(r)} \left\| \left\langle \frac{x - c(q)}{r} \right\rangle^N \chi_q g \right\|_{L^2}^2 \lesssim_N \|g\|_{L^2}^2 \tag{6-8}$$

for any $g \in L^2(\mathcal{H}_i)$. Here, the variable x is the argument of g and belongs to $\mathcal{H}_i$.

Next we turn our attention to similar objects corresponding to the more complex family indexed by $\alpha \in J$. Given $\widetilde{\mathcal{H}}_{1,\alpha} \subset \mathcal{H}_{1,\alpha}$ a subspace of dimension $k - 1$, we let $\tilde{\pi}_\alpha : \mathcal{H}_{1,\alpha} \to \widetilde{\mathcal{H}}_{1,\alpha}$ be the orthogonal projection onto $\widetilde{\mathcal{H}}_{1,\alpha}$. We denote by $(\widetilde{\mathcal{H}}_{1,\alpha})^\perp$ the normal subspace to $\widetilde{\mathcal{H}}_{1,\alpha}$ in $\mathcal{H}_{1,\alpha}$.

We let $\pi_{N_{1,\alpha}} : \mathbb{R}^{n+1} \to \mathcal{H}_{1,\alpha}$ be the projection onto $\mathcal{H}_{1,\alpha}$ and $\tilde{\pi}_{1,\alpha} := \tilde{\pi}_\alpha \circ \pi_{N_{1,\alpha}} : \mathbb{R}^{n+1} \to \widetilde{\mathcal{H}}_{1,\alpha}$ be the projection onto $\widetilde{\mathcal{H}}_{1,\alpha}$. We define the lattices $\mathcal{L}(\mathcal{H}_{1,\alpha}) = \mathbb{Z}^n$ inside $\mathcal{H}_{1,\alpha}$ and $\mathcal{L}(\widetilde{\mathcal{H}}_{1,\alpha}) = \mathbb{Z}^{k-1}$ inside $\widetilde{\mathcal{H}}_{1,\alpha}$ with respect to orthonormal basis in each case. They are constructed such that $\tilde{\pi}_\alpha(\mathcal{L}(\mathcal{H}_{1,\alpha})) = \mathcal{L}(\widetilde{\mathcal{H}}_{1,\alpha})$; this holds true if the orthonormal basis in $\widetilde{\mathcal{H}}_{1,\alpha}$ is a subset of the orthonormal basis in $\mathcal{H}_{1,\alpha}$.

Inside the subspace $\mathcal{H}_{1,\alpha}$ we construct $\mathcal{C}_{1,\alpha}(r)$ to be the set of cubes of size r centered at points from the lattice $r\mathcal{L}(\mathcal{H}_{1,\alpha})$ and sides parallel to the directions of the lattice. Inside the subspace $\widetilde{\mathcal{H}}_{1,\alpha}$ we construct $\widetilde{\mathcal{C}}_{1,\alpha}(r)$ be the set of cubes of size r centered at points from the lattice $r\mathcal{L}(\widetilde{\mathcal{H}}_{1,\alpha})$ with sides parallel to the directions of the lattice. Therefore $\widetilde{\mathcal{C}}_{1,\alpha}(r) = \tilde{\pi}^{\alpha}\mathcal{C}_{1,\alpha}(r)$. Then we define $\mathfrak{S}_{1,\alpha}(r)$ to be the set of infinite cubical strips $\mathfrak{s} = q \times (\widetilde{\mathcal{H}}_{1,\alpha})^{\perp} \subset \mathcal{H}_{1,\alpha}$, where $q \in \widetilde{\mathcal{C}}_{1,\alpha}(r)$. We denote by $c(\mathfrak{s}) := c(q) \subset r\mathcal{L}(\widetilde{\mathcal{H}}_{1,\alpha})$ the center of the strip. We note that given $q_1, q_2 \in \mathcal{C}_{1,\alpha}(r)$, they belong to the same cubical strip in $\mathfrak{S}_{1,\alpha}(r)$ if and only if $\tilde{\pi}_{\alpha} q_1 = \tilde{\pi}_{\alpha} q_2$. For $q \in \mathcal{C}_{1,\alpha}(r)$, we let $\mathfrak{s}(\tilde{\pi}_{\alpha} q)$ be the infinite cubical strip it belongs to as a subset in $\mathfrak{S}_{1,\alpha}(r)$. Given a strip $\mathfrak{s} \in \mathfrak{S}_{1,\alpha}(r)$ we define $\chi_{\mathfrak{s}} : \mathcal{H}_{1,\alpha} \to \mathbb{R}$

$$\chi_{\mathfrak{s}}(x) = \eta_0\left(\frac{\tilde{\pi}_{1,\alpha}(x) - c(\mathfrak{s})}{r}\right),$$

where, by abusing notation, $\eta_0 : \mathbb{R}^{k-1} \to \mathbb{R}$ is entirely similar to the η_0 introduced in Section 6A, except that it acts on $\mathbb{R}^{k-1}$ instead of $\mathbb{R}^n$. A key property of $\chi_{\mathfrak{s}}$ is that it is constant in directions from the subspace $(\widetilde{\mathcal{H}}_{1,\alpha})^{\perp}$.

One unpleasant feature of the above construction is that the lattice $\mathcal{L}$ does not project exactly into the lattices $\mathcal{L}(\widetilde{\mathcal{H}}_{1,\alpha})$ via $\tilde{\pi}_{1,\alpha}$; similarly $\mathcal{C}_{\alpha}(r)$ does not project well into $\widetilde{\mathcal{C}}_{1,\alpha}(r)$ via $\tilde{\pi}_{1,\alpha}$. This is an inherent feature of the fact that there are too many subspaces $\widetilde{\mathcal{H}}_{1,\alpha}$. As a consequence, given $q \in \mathcal{C}(r)$, it is not necessarily true that $\tilde{\pi}_{1,\alpha}(q) \in \widetilde{\mathcal{C}}_{1,\alpha}(r)$; however $\tilde{\pi}_{1,\alpha}(q)$ intersects a finite number of $q' \in \widetilde{\mathcal{C}}_{1,\alpha}(r)$. Abusing notation, we define

$$\mathfrak{s}^{\alpha}(\tilde{\pi}_{1,\alpha}(q)) = \bigcup_{\substack{q' \in \widetilde{\mathcal{C}}_{1,\alpha}(r) \\ q' \cap \tilde{\pi}_{1,\alpha}(q) \neq \varnothing}} \mathfrak{s}^{\alpha}(q'),$$

the strip generated by the projection of q onto $\widetilde{\mathcal{H}}_{1,\alpha}$.

Recalling that $\mathcal{L} := \{z_1 N_1 + \cdots + z_{n+1} N_{n+1} : (z_1, \ldots, z_{n+1}) \in \mathbb{Z}^{n+1}\}$, we denote the coordinates of a point in the lattice by $(z_1, \ldots, z_{n+1})$ and define

$$\|g\|_{l^{\infty}_{z_1,z_{k+1},\ldots,z_{n+1}} l^2_{z_2,\ldots,z_k}(\mathcal{L})} = \sup_{z_1,z_{k+1},\ldots,z_{n+1}} \|g(z_1, \cdot, z_{k+1}, \ldots, z_{n+1})\|_{l^2_{z_2,\ldots,z_k}},$$

where $\cdot$ stands for the variables $z_2, \ldots, z_k$ with respect to which l^2 is computed.

With this notation in place we have the following result:

Lemma 6.3. *Assume* $g_1 \in l^{\infty}_{z_1,z_{k+1},\ldots,z_{n+1}} l^2_{z_2,\ldots,z_k}(\mathcal{L})$ *and* $g_i \in l^2(\mathcal{L}(\mathcal{H}_i))$, $i = 2, \ldots, k$. *Then the following holds true*:

$$\left\| g_1(z) \prod_{i=2}^{k} g_i(\pi_{N_i}(z)) \right\|_{l^{2/(k-1)}(\mathcal{L})} \lesssim \|g_1\|_{l^{\infty}_{z_1,z_{k+1},\ldots,z_{n+1}} l^2_{z_2,\ldots,z_k}(\mathcal{L})} \prod_{i=2}^{k} \|g_i\|_{l^2(\mathcal{L}(\mathcal{H}_i))}. \tag{6-9}$$

Proof. The function $g_i \circ \pi_{N_i}$ is independent of the z_i-variable; therefore it holds true that $g_i \circ \pi_{N_i} \in l^2_{z_1,z_{k+1},\ldots,z_{n+1}} l^2_{z_2,\ldots,z_{i-1}} l^{\infty}_{z_i} l^2_{z_{i+1},\ldots,z_k}$ and

$$\|g_i \circ \pi_{N_i}\|_{l^2_{z_1,z_{k+1},\ldots,z_{n+1}} l^2_{z_2,\ldots,z_{i-1}} l^{\infty}_{z_i} l^2_{z_{i+1},\ldots,z_k}} \leq \|g_i\|_{l^2(\mathcal{L}(\mathcal{H}_i))},$$

where then norms $l^2_{z_1,z_{k+1},\dots,z_{n+1}} l^2_{z_2,\dots,z_{i-1}} l^\infty_{z_i} l^2_{z_{i+1},\dots,z_k}$ are defined in the standard fashion. Then the result is a direct consequence of the Hölder inequality in its discrete version. □

6B. *Proofs of the main results.*

Proof of Theorem 6.1. The argument is based on an induction on scales. Given a $0 < \delta \ll 1$, we break the surfaces into smaller pieces of diameter $\lesssim \delta$. A result on the smaller scales is converted to a result at the original scale at the cost of a large power of δ^{-1}, which is absorbed into $C(\epsilon)$. Thus, the focus will be on providing a result in the context of surfaces with diameter less than δ.

We run an induction with respect to the size of the cube where estimates are made. We show that passing from an estimate on cubes of size R to an estimate on cubes of size $\delta^{-1}R$ can be done by accumulating constants that are independent of δ and R. In implementing this approach, we use a phase-space approach that alters the support of $f_{1,\alpha}, f_2, \dots, f_k$ by a factor $\approx R^{-1/2}$ where $R \ge \delta^{-2}$. This is fine with $f_2, \dots, f_k$ but not with $f_{1,\alpha}$, $\alpha \in J$, given that their support in some directions is $\mu \ll \delta$. This will require extra care.

We work under the hypothesis that $U_i \subset B_i(0;\delta)$, $i = 2, \dots, k$, where $B_i(0;\delta)$ is the ball in the hyperplane $\mathcal{H}_i$. For a function $f_i : \mathcal{H}_i \to \mathbb{C}$, its margin is defined as

$$\operatorname{margin}^i(f_i) := \operatorname{dist}(\operatorname{supp}(f), B_i(0;2\delta)^c), \quad i = 2, \dots, k, \tag{6-10}$$

where supp is the support of f_i.

We work under the hypothesis that $U_{1,\alpha} \subset B'(0;\delta) \times B''(0;\mu)$, where $B'(0;\delta)$ is the ball in the hyperplane $\widetilde{\mathcal{H}}_{1,\alpha}$ centered at the origin and of diameter δ and $B'(0;\mu)$ is the ball in the hyperplane $(\widetilde{\mathcal{H}}_{1,\alpha})^\perp$ centered at the origin and of diameter μ. Accordingly, we split the coordinates in $\mathcal{H}_{1,\alpha}$ as $\xi^\alpha = (\xi^{\prime,\alpha}, \xi^{\prime\prime,\alpha})$, where $\xi^{\prime,\alpha}$ is the coordinate in $\widetilde{\mathcal{H}}_{1,\alpha}$ and $\xi^{\prime\prime,\alpha}$ is the coordinate in $(\widetilde{\mathcal{H}}_{1,\alpha})^\perp$. Given a function $f : \mathcal{H}_{1,\alpha} \to \mathbb{R}$, its margin is defined by

$$\operatorname{margin}^{1,\alpha}(f) := \inf_{\xi^{\prime\prime,\alpha}} \operatorname{dist}\big(\operatorname{supp}_{\xi^{\prime,\alpha}}(f(\cdot,\xi^{\prime\prime,\alpha})), B'(0;2\delta)^c\big), \tag{6-11}$$

where $\operatorname{supp}_{\xi^{\prime,\alpha}}$ is the support of f in the $\xi^{\prime,\alpha}$-variable. On the physical side we denote by $x^{\prime,\alpha}, x^{\prime\prime,\alpha}$ the dual variables to $\xi^{\prime,\alpha}, \xi^{\prime\prime,\alpha}$, respectively. We complete the system of coordinates to $(\xi^\alpha_1, \xi^{\prime,\alpha}, \xi^{\prime\prime,\alpha})$ and $(x^\alpha_1, x^{\prime,\alpha}, x^{\prime\prime,\alpha})$, where ξ^α_1 is the coordinate in the direction of $N_{1,\alpha}$, the normal to $\mathcal{H}_{1,\alpha}$, and x^α_1 is the dual coordinate.

Our induction aims at quantifying the behavior of $A(R)$ defined below.

Definition 6.4. Given $R \ge \delta^{-2}$ we define $A(R)$ to be the best constant for which the estimate

$$\Big\| \Big(\sum_\alpha |\mathcal{E}_{1,\alpha} f_{1,\alpha}| \Big) \prod_{i=2}^k \mathcal{E}_i f_i \Big\|_{L^{2/(k-1)}(Q)} \le A(R) \prod_{i=1}^k \|f_i\|_{L^2} \tag{6-12}$$

holds true for all parallelepipeds $Q \in \mathcal{C}(R)$, with f_i obeying the margin requirement

$$\operatorname{margin}^i(f_i) \ge \delta - R^{-\frac12}, \quad i = 2, \dots, k, \qquad \operatorname{margin}^{1,\alpha}(f_{1,\alpha}) \ge \delta - R^{-\frac12} \quad \text{for all } \alpha \in J, \tag{6-13}$$

and $f_{1,\alpha}$ is supported in $B(\widetilde{\mathcal{H}}_{1,\alpha};\mu) \subset \mathcal{H}_{1,\alpha}$ for all $\alpha \in J$.

Note that in (6-12) we use absolute values. This indicates that we do not use any cancellation properties between the components $\mathcal{E}_{1,\alpha} f_{1,\alpha}$. However, using the stronger statement with the absolute values plays a crucial role in carrying out the induction argument.

We start with the parallelepiped Q of size $\delta^{-1} R$ centered at the origin. To keep notation compact we define

$$H = \prod_{i=2}^{k} \mathcal{E}_i f_i, \quad G = \prod_{i=2}^{k} \| f_i \|_{L^2}.$$

For each $q \in \mathcal{C}(R) \cap Q$, the induction hypothesis is

$$\left\| \left(\sum_{\alpha} |\mathcal{E}_{1,\alpha} f_{1,\alpha}| \right) \cdot H \right\|_{L^{2/(k-1)}(q)} \leq A(R) \left(\sum_{\alpha} \| f_{1,\alpha} \|_{L^2(U_{1,\alpha})} \right) G. \tag{6-14}$$

We claim the following strengthening of (6-14):

$$\begin{aligned}
&\left\| \left(\sum_{\alpha} |\mathcal{E}_{1,\alpha} f_{1,\alpha}| \right) \cdot H \right\|_{L^{2/(k-1)}(q)} \\
&\quad \lesssim_N A(R) \prod_{i=2}^{k} \left(\sum_{q' \in \mathcal{CH}_i(R)} \left\langle \frac{d(\pi_{N_i} q, q')}{R} \right\rangle^{-(2N-n^2)} \left\| \left\langle \frac{x - c(q')}{R} \right\rangle^{N} \chi_{q'} \mathcal{F}_i^{-1} f_i \right\|_{L^2}^2 \right)^{\frac{1}{2}} \\
&\qquad \cdot \sum_{\alpha} \left(\sum_{\mathfrak{s}^\alpha \in \mathfrak{S}_{1,\alpha}(R)} \left\langle \frac{d(\tilde{\pi}_{1,\alpha}(q), \mathfrak{s}^\alpha)}{R} \right\rangle^{-(2N-2k)} \left\| \left\langle \frac{x'^{,\alpha} - c(\mathfrak{s}^\alpha)}{R} \right\rangle^{N} \chi_{\mathfrak{s}^\alpha} \mathcal{F}_{1,\alpha}^{-1} f_{1,\alpha} \right\|_{L^2}^2 \right)^{\frac{1}{2}}.
\end{aligned} \tag{6-15}$$

Similar improvements were provided in [Bejenaru 2017a]; in particular the improvement for the terms f_i, $i = 2, \dots, k$, was established, as claimed above (it can also be derived along similar, but simpler, lines as those in the arguments we provide below for the $f_{1,\alpha}$ terms). The improvement for $f_{1,\alpha}$ was also provided in that paper in the case when there is only one function $f_{1,\alpha}$, that is, J contains one element only. Here we provide an argument for general finite sets J and note that the cardinality of J does not impact $A(R)$.

Therefore, in justifying (6-15) we focus on the improvement for the $f_{1,\alpha}$ terms only. Given $q \in \mathcal{C}(R) \cap Q$ and $d \in \mathbb{N}$, let

$$A^\alpha(q, d) = \left\{ \mathfrak{s} \in \mathfrak{S}_{1,\alpha}(r) : \left\langle \frac{d(\tilde{\pi}_{1,\alpha}(q), \mathfrak{s})}{R} \right\rangle \approx d \right\}.$$

We can modify the sets such that each strip $\mathfrak{s}$ belongs to only one $A^\alpha(q, d)$.

From (6-6) we obtain the identity

$$\begin{aligned}
&\sum_{\alpha} \sum_{\mathfrak{s}^\alpha \in A^\alpha(q,d)} \left| \left(x'^{,\alpha} - c(\mathfrak{s}^\alpha) - x_1^\alpha \nabla_{\xi'^\alpha} \varphi_1 \left(\frac{D^\alpha}{i} \right) \right) \mathcal{E}_{1,\alpha} \mathcal{F}_{1,\alpha} \chi_{\mathfrak{s}^\alpha} \mathcal{F}_{1,\alpha}^{-1} f_{1,\alpha} \right| \\
&\qquad = \sum_{\alpha} \sum_{\mathfrak{s}^\alpha \in A^\alpha(q,d)} |\mathcal{E}_1^\alpha \mathcal{F}_{1,\alpha} (x'^{,\alpha} - c(\mathfrak{s}^\alpha)) \chi_{\mathfrak{s}^\alpha} \mathcal{F}_{1,\alpha}^{-1} f_{1,\alpha}|,
\end{aligned} \tag{6-16}$$

where the differential operator $\nabla_{\xi'^\alpha}\varphi_1(D^\alpha/i)$ has symbol $\nabla_{\xi'^\alpha}\varphi_1(\xi^\alpha)$. We have the following sequence of estimates:

$$\Big\| \sum_\alpha \sum_{\mathfrak{s}^\alpha \in A^\alpha(q,d)} |(x'^{,\alpha} - c(\mathfrak{s}^\alpha) - x_1^\alpha \nabla_{\xi'^\alpha}\varphi_1(\eta_{1,\alpha}))\mathcal{E}_{1,\alpha}\mathcal{F}_{1,\alpha}\chi_{\mathfrak{s}^\alpha}\mathcal{F}_{1,\alpha}^{-1} f_{1,\alpha}| \cdot H \Big\|_{L^{2/(k-1)}(q)}$$

$$\leq \Big\| \sum_\alpha \sum_{\mathfrak{s}^\alpha \in A^\alpha(q,d)} |(x'^{,\alpha} - c(\mathfrak{s}^\alpha) - x_1^\alpha \nabla_{\xi'^\alpha}\varphi_1(\xi^\alpha))\mathcal{E}_{1,\alpha}\mathcal{F}_{1,\alpha}\chi_{\mathfrak{s}^\alpha}\mathcal{F}_{1,\alpha}^{-1} f_{1,\alpha}| \cdot H \Big\|_{L^{2/(k-1)}(q)} + \big\| x_1^\alpha (\nabla_{\xi'^\alpha}\varphi_1(\eta_{1,\alpha}) - \nabla_{\xi'^\alpha}\varphi_1(\xi^\alpha))\mathcal{E}_{1,\alpha}\mathcal{F}_{1,\alpha}\chi_{\mathfrak{s}^\alpha}\mathcal{F}_{1,\alpha}^{-1} f_{1,\alpha} \cdot H \big\|_{L^{2/(k-1)}(q)}.$$

We invoke (6-16) and continue with

$$= \Big\| \sum_\alpha \sum_{\mathfrak{s}^\alpha \in A^\alpha(q,d)} |\mathcal{E}_{1,\alpha}\mathcal{F}_{1,\alpha}(x'^{,\alpha} - c(\mathfrak{s}^\alpha))\chi_{\mathfrak{s}^\alpha}\mathcal{F}_{1,\alpha}^{-1} f_{1,\alpha}| \cdot H \Big\|_{L^{2/(k-1)}(q)} + \Big\| \sum_\alpha \sum_{\mathfrak{s}^\alpha \in A^\alpha(q,d)} x_1^\alpha \mathcal{E}_{1,\alpha}\mathcal{F}_{1,\alpha}(\nabla_{\xi'^\alpha}\varphi_1(\eta_{1,\alpha}) - \nabla_{\xi'^\alpha}\varphi_1(\xi^\alpha))\chi_{\mathfrak{s}^\alpha}\mathcal{F}_{1,\alpha}^{-1} f_{1,\alpha}) \Big\|_{L^{2/(k-1)}}.$$

We apply the induction hypothesis, and use that inside Q we have $|x_1^\alpha| \lesssim \delta^{-1}R$ for all $\alpha \in J$, to further continue with

$$\leq A(R)\Big(\sum_\alpha \sum_{\mathfrak{s}^\alpha \in A^\alpha(q,d)} \|(x'^{,\alpha} - c(\mathfrak{s}^\alpha))\chi_{\mathfrak{s}^\alpha}\mathcal{F}_{1,\alpha}^{-1} f_{1,\alpha}\|_{L^2} \Big) G + A(R)\delta^{-1}R \Big(\sum_\alpha \sum_{\mathfrak{s}^\alpha \in A^\alpha(q,d)} \|(\nabla_{\xi'^\alpha}\varphi_1(\eta_{1,\alpha}) - \nabla_{\xi'^\alpha}\varphi_1(\xi^\alpha))\chi_{\mathfrak{s}^\alpha}\mathcal{F}_{1,\alpha}^{-1} f_{1,\alpha}\|_{L^2} \Big) G$$

$$\lesssim A(R)\Big(\sum_\alpha \sum_{\mathfrak{s}^\alpha \in A^\alpha(q,d)} \|(x'^{,\alpha} - c(\mathfrak{s}^\alpha))\chi_{\mathfrak{s}^\alpha}\mathcal{F}_{1,\alpha}^{-1} f_{1,\alpha}\|_{L^2} + R\|\chi_{\mathfrak{s}^\alpha}\mathcal{F}_{1,\alpha}^{-1} f_{1,\alpha}\|_{L^2} \Big) G$$

$$\lesssim RA(R)\Big(\sum_\alpha \sum_{\mathfrak{s}^\alpha \in A^\alpha(q,d)} \Big\| \Big\langle \frac{x'^{,\alpha} - c(\mathfrak{s}^\alpha)}{R} \Big\rangle \chi_{\mathfrak{s}^\alpha}\mathcal{F}_{1,\alpha}^{-1} f_{1,\alpha} \Big\|_{L^2} \Big) G.$$

Note that it is in the above use of the induction estimate for $\mathcal{E}_{1,\alpha}\mathcal{F}_{1,\alpha}(x'^{,\alpha} - c(\mathfrak{s}^\alpha))\chi_{\mathfrak{s}^\alpha}\mathcal{F}_{1,\alpha}^{-1} f_{1,\alpha}$ that we need to tolerate the relaxed support of $f_{1,\alpha}$. The margin of $f_{1,\alpha}$ is $\geq \delta - (\delta^{-1}R)^{-1/2} = \delta - \delta^{1/2}R^{-1/2}$ and it is affected by the convolution $\mathcal{F}_{1,\alpha}((x'^{,\alpha} - c(\mathfrak{s}^\alpha))\chi_{\mathfrak{s}^\alpha})$ by a factor of at most CR^{-1}, which is smaller than $\frac{1}{2}\delta^{1/2}R^{-1/2}$ provided that δ is small relative to C^{-1}. Hence the new margin is $\geq \delta - \frac{1}{2}\delta^{1/2}R^{-1/2} \geq \delta - R^{-1/2}$, which is the required margin for using the induction hypothesis on cubes of size R.

We claim that for any $\mathfrak{s}^\alpha \in A^\alpha(q,d)$

$$\Big\| \Big\langle \frac{x'^{,\alpha} - c(\mathfrak{s}^\alpha) - x_1^\alpha \nabla_{\xi'^{,\alpha}}\varphi_1(\eta_{1,\alpha})}{R} \Big\rangle \Big\|_{L^\infty(q)} \approx \Big\langle \frac{d(\tilde{\pi}_{1,\alpha}(q), \mathfrak{s}^\alpha)}{R} \Big\rangle \approx d$$

uniformly in α. This statement is invariant to rotations of coordinates, therefore we can assume that $\nabla_{\xi^\alpha}\varphi_1(\eta_{1,\alpha}) = 0$ and moreover that $x'^{,\alpha} = (x_2, \dots, x_k)$ and $x''^{,\alpha} = (x_{k+1}, \dots, x_{n+1})$. This way, $\tilde{\pi}_{1,\alpha}(x) = (0, x_2, \dots, x_k, 0, \dots, 0)$ and the statement is obvious.

From the above we obtain that, for d large,

$$dR\sum_{\alpha}\sum_{\mathfrak{s}^\alpha\in A^\alpha(q,d)}|\mathcal{E}_{1,\alpha}\mathcal{F}_{1,\alpha}\chi_{\mathfrak{s}^\alpha}\mathcal{F}^{-1}_{1,\alpha}f_{1,\alpha}| \lesssim \sum_{\alpha}\sum_{\mathfrak{s}^\alpha\in A^\alpha(q,d)}|(x^{\prime,\alpha}-c(\mathfrak{s}^\alpha)-x_1^\alpha\nabla_{\xi^{\prime}\alpha}\varphi_1(\eta_{1,\alpha}))\mathcal{E}_{1,\alpha}\mathcal{F}_{1,\alpha}\chi_{\mathfrak{s}^\alpha}\mathcal{F}^{-1}_{1,\alpha}f_{1,\alpha}|.$$

Combining all the above estimates gives

$$\begin{aligned} dR\Big\|\sum_{\alpha}\sum_{\mathfrak{s}^\alpha\in A^\alpha(q,d)}|\mathcal{E}_{1,\alpha}\mathcal{F}_{1,\alpha}\chi_{\mathfrak{s}^\alpha}\mathcal{F}^{-1}_{1,\alpha}f_{1,\alpha}|\cdot H\Big\|_{L^{2/(k-1)}(q)} &\lesssim \Big\|\sum_{\alpha}\sum_{\mathfrak{s}^\alpha\in A^\alpha(q,d)}|((x^{\prime,\alpha})-c(\mathfrak{s}^\alpha)-x_1^\alpha\nabla_{\xi^{\prime,\alpha}}\varphi_1(\eta_{1,\alpha}))\mathcal{E}_{1,\alpha}\mathcal{F}_{1,\alpha}\chi_{\mathfrak{s}^\alpha}\mathcal{F}^{-1}_{1,\alpha}f_{1,\alpha}|\cdot H\Big\|_{L^{2/(k-1)}(q)} \\ &\lesssim RA(R)\bigg(\sum_{\alpha}\sum_{\mathfrak{s}^\alpha\in A^\alpha(q,d)}\Big\|\Big\langle\frac{x^{\prime,\alpha}-c(\mathfrak{s}^\alpha)}{R}\Big\rangle\chi_{\mathfrak{s}^\alpha}\mathcal{F}^{-1}_{1,\alpha}f_{1,\alpha}\Big\|_{L^2}\bigg)G. \end{aligned}$$

From this we conclude with (after more iterations of the same argument)

$$\Big\|\sum_{\alpha}\sum_{\mathfrak{s}^\alpha\in A^\alpha(q,d)}|\mathcal{E}_{1,\alpha}\mathcal{F}_{1,\alpha}\chi_{\mathfrak{s}^\alpha}\mathcal{F}^{-1}_{1,\alpha}f_{1,\alpha}|\cdot H\Big\|_{L^{2/(k-1)}(q)} \lesssim d^{-N}A(R)\sum_{\alpha}\sum_{\mathfrak{s}^\alpha\in A^\alpha(q,d)}\Big\|\Big\langle\frac{x^{\prime,\alpha}-c(\mathfrak{s}^\alpha)}{R}\Big\rangle^N\chi_{\mathfrak{s}^\alpha}\mathcal{F}^{-1}_{1,\alpha}f_{1,\alpha}\Big\|_{L^2}\cdot G.$$

Note, that while the argument above assumed d is large, this last inequality holds for all d, since it is trivial for d small. The summation over d is done in the usual manner:

$$\begin{aligned} \Big\|\sum_{\alpha}|\mathcal{E}_{1,\alpha}f_{1,\alpha}|H\Big\|^{\frac{2}{k-1}}_{L^{2/(k-1)}(q)} &= \Big\|\sum_{d}\sum_{\alpha}\sum_{\mathfrak{s}^\alpha\in A^\alpha(q,d)}|\mathcal{E}_{1,\alpha}\mathcal{F}_{1}\chi_{\mathfrak{s}^\alpha}\mathcal{F}^{-1}_{1}f_{1,\alpha}|H\Big\|^{\frac{2}{k-1}}_{L^{2/(k-1)}(q)} \\ &\lesssim \sum_{d}\Big\|\sum_{\alpha}\sum_{\mathfrak{s}^\alpha\in A^\alpha(q,d)}|\mathcal{E}_{1,\alpha}\mathcal{F}_{1,\alpha}\chi_{\mathfrak{s}^\alpha}\mathcal{F}^{-1}_{1,\alpha}f_{1,\alpha}|H\Big\|^{\frac{2}{k-1}}_{L^{2/(k-1)}(q)} \\ &\lesssim (A(R))^{\frac{2}{k-1}}\sum_{d}d^{-N\cdot\frac{2}{k-1}}\bigg(\sum_{\alpha}\sum_{\mathfrak{s}^\alpha\in A^\alpha(q,d)}\Big\|\Big\langle\frac{x^{\prime,\alpha}-c(\mathfrak{s}^\alpha)}{R}\Big\rangle^N\chi_{\mathfrak{s}^\alpha}\mathcal{F}^{-1}_{1}f_{1,\alpha}\Big\|_{L^2}\bigg)^{\frac{2}{k-1}}G^{\frac{2}{k-1}}. \end{aligned}$$

Using (1-8) together with the straightforward estimate

$$\|d^{-\frac{k}{2}}\|_{l^{2/(k-1)}_{\mathbb{N}}} \lesssim 1,$$

and we can continue the sequence of inequalities we started above

$$\begin{aligned}
&\Big\| \sum_\alpha |\mathcal{E}_{1,\alpha} f_{1,\alpha}| H \Big\|_{L^{2/(k-1)}(q)}^{\frac{2}{k-1}} \\
&\lesssim (A(R))^{\frac{2}{k-1}} \Big\| d^{-(N-\frac{k}{2})} \sum_\alpha \sum_{\mathfrak{s}^\alpha \in A^\alpha(q,d)} \Big\| \Big\langle \frac{x^{\prime,\alpha} - c(\mathfrak{s}^\alpha)}{R} \Big\rangle^N \chi_{\mathfrak{s}^\alpha} \mathcal{F}_1^{-1} f_{1,\alpha} \Big\|_{L^2} \Big\|_{l^2_d}^{\frac{2}{k-1}} G^{\frac{2}{k-1}} \\
&\lesssim (A(R))^{\frac{2}{k-1}} \Big(\sum_\alpha \Big\| d^{-(N-\frac{k}{2})} \sum_{\mathfrak{s}^\alpha \in A^\alpha(q,d)} \Big\| \Big\langle \frac{x^{\prime,\alpha} - c(\mathfrak{s}^\alpha)}{R} \Big\rangle^N \chi_{\mathfrak{s}^\alpha} \mathcal{F}_1^{-1} f_{1,\alpha} \Big\|_{L^2} \Big\|_{l^2_d} \Big)^{\frac{2}{k-1}} G^{\frac{2}{k-1}} \\
&\lesssim (A(R))^{\frac{2}{k-1}} \bigg(\sum_\alpha \bigg(\sum_{\mathfrak{s}^\alpha} \Big\langle \frac{d(\tilde{\pi}_{1,\alpha}(q), \mathfrak{s}^\alpha)}{R} \Big\rangle^{-(2N-2k)} \Big\| \Big\langle \frac{x^{\prime,\alpha} - c(\mathfrak{s}^\alpha)}{R} \Big\rangle^N \chi_{\mathfrak{s}^\alpha} \mathcal{F}_1^{-1} f_{1,\alpha} \Big\|_{L^2}^2 \bigg)^{\frac{1}{2}} \bigg)^{\frac{2}{k-1}} G^{\frac{2}{k-1}}.
\end{aligned}$$

In passing to the last line we used that the cardinality of $A^\alpha(q,d)$ is $\approx \langle d \rangle^{k-1}$ in order to bound the $l^1_{\mathfrak{s}^\alpha \in A^\alpha(q,d)}$-norm of the summand by the $l^2_{\mathfrak{s}^\alpha \in A^\alpha(q,d)}$ of the same quantity.

We are done with the justification of (6-15) and continue with the final step in the induction on scales. We define the functions $g_i : \mathcal{L}(\mathcal{H}_i) \to \mathbb{R}$ for $i = 2, \dots, k$ by

$$g_i(\boldsymbol{j}) = \bigg(\sum_{q' \in \mathcal{CH}_i(R)} \Big\langle \frac{d(q(\boldsymbol{j}), q')}{R} \Big\rangle^{-(2N-n^2)} \Big\| \Big\langle \frac{x' - c(q')}{R} \Big\rangle^N \chi_{q'} \mathcal{F}_i^{-1} f_i \Big\|_{L^2}^2 \bigg)^{\frac{1}{2}}$$

for $\boldsymbol{j} \in \mathcal{L}(\mathcal{H}_i)$, and $g_1 : \mathcal{L} \to \mathbb{R}$ by

$$g_1(\boldsymbol{j}) = \sum_\alpha \bigg(\sum_{\mathfrak{s}^\alpha} \Big\langle \frac{d(\tilde{\pi}_{1,\alpha}(q(\boldsymbol{j})), \mathfrak{s}^\alpha)}{R} \Big\rangle^{-(2N-2k)} \Big\| \Big\langle \frac{x^{\prime,\alpha} - c(\mathfrak{s}^\alpha)}{R} \Big\rangle^N \chi_{\mathfrak{s}^\alpha} \mathcal{F}_1^{-1} f_{1,\alpha} \Big\|_{L^2}^2 \bigg)^{\frac{1}{2}}$$

for $\boldsymbol{j} \in \mathcal{L}$. Using (6-8), it is obvious that, provided N is large enough (in terms of n only), the following holds true:

$$\|g_i\|_{l^2(\mathcal{L}(\mathcal{H}_i))} \lesssim \|f_i\|_{L^2}, \quad i = 2, \dots, k.$$

We also claim that

$$\|g_1\|_{l^\infty_{z_1, z_{k+1}, \dots, z_{n+1}} l^2_{z_2, \dots, z_k}(\mathcal{L})} \lesssim \sum_\alpha \|f_{1,\alpha}\|_{L^2}. \tag{6-17}$$

This is a consequence of the following geometrical observation: Say $\boldsymbol{j} = \sum_{i=1}^{n+1} z_i N_i$, where $z_i \in \mathbb{Z}$. We fix $z_1, z_{k+1}, \dots, z_{n+1}$ and note that as we vary $z_2, \dots, z_k$, the $\tilde{\pi}_{1,\alpha}(q(\boldsymbol{j}))$ are almost disjoint and, most importantly, the strips they generate, $\mathfrak{s}^\alpha(\tilde{\pi}_{1,\alpha} q(\boldsymbol{j})) \subset \mathfrak{S}_{1,\alpha}(R)$, are almost disjoint for each $\alpha \in J$ (given a point in $\mathcal{H}_{1,\alpha}$ there are finitely many $\boldsymbol{j}$ such that the point belongs to $\mathfrak{s}^\alpha(\tilde{\pi}_{1,\alpha} q(\boldsymbol{j}))$). This is due to the fact that the projections $\tilde{\pi}_{1,\alpha}$ onto the affine subspace $\widetilde{\mathcal{H}}_{1,\alpha}$ are taken along directions that are transversal to $N_2, \dots, N_k$ and the infinite sides of the strips are in directions that are transversal to $N_2, \dots, N_k$. Using this geometric observation, (6-17) follows from the equivalent of (6-8) for strips.

Then we apply (6-9) to conclude with

$$\left\| \left(\sum_\alpha |\mathcal{E}_{1,\alpha} f_{1,\alpha}| \right) \cdot H \right\|_{L^{2/(k-1)}(Q)} \lesssim A(R) \prod_{i=1}^k \|f_i\|_{L^2}.$$

Thus we obtain

$$A(\delta^{-1}R) \le CA(R)$$

for a constant C that is independent of δ and R. Iterating this gives $A(\delta^{-N} r) \le C^N A(r)$. Therefore $\max_{r\in[0,\delta^{-2}]} A(\delta^{-N} r) \le C^N \max_{r\in[0,\delta^{-2}]} A(r) = C^N C(\delta)\mu^{(n+1-k)/2}$ is obtained from the uniform pointwise bound

$$\begin{aligned} \left\| \left(\sum_\alpha |\mathcal{E}_{1,\alpha} f_{1,\alpha}| \right) \prod_{i=2}^k \mathcal{E}_i f_i \right\|_{L^\infty} &\lesssim \left\| \sum_\alpha |\mathcal{E}_{1,\alpha} f_{1,\alpha}| \right\|_{L^\infty} \prod_{i=2}^k \|\mathcal{E}_i f_i\|_{L^\infty} \\ &\lesssim \mu^{\frac{n+1-k}{2}} \left(\sum_\alpha \|f_{1,\alpha}\|_{L^2} \right) \prod_{i=2}^k \|f_i\|_{L^2}, \end{aligned} \tag{6-18}$$

which is integrated over arbitrary cubes of size $\le \delta^{-2}$. Note that we have used the support properties of $f_{1,\alpha}$ to obtain the improved bound.

For $R \in [\delta^{-N}, \delta^{-N-1}]$, the above implies

$$A(R) \le C^N C(\delta)\mu^{\frac{n+1-k}{2}} \le R^\epsilon C(\delta)\mu^{\frac{n+1-k}{2}}$$

provided that $C^N \le \delta^{-N\epsilon}$. Therefore choosing $\delta = C^{-1/\epsilon}$ leads to the desired result. $\square$

Proof of Corollary 6.2. In each $\mathcal{H}_i$, $i = 1, \dots, k$, $y_i \in \mathbb{R}$, we define $\mathcal{H}_i + y_i N_i$ to be the translation of $\mathcal{H}_i$ by $y_i N_i$. Also $\mathcal{CH}_i(r) + y_i N_i$ is the corresponding translation of $\mathcal{CH}_i(r)$ by $y_i N_i$.

Given any vector $y \in \mathbb{R}^{n+1}$ with $|y_i - c_i(q)| \le r$, $i = 1, \dots, k$, and $y_i = c_i(q)$, $k+1 \le i \le n+1$, we claim

$$\begin{aligned} &\left\| \left(\sum_\alpha |\mathcal{E}_{1,\alpha} f_{1,\alpha}| \right) \cdot H \right\|_{L^{2/(k-1)}(q)} \\ &\lesssim_N C(\epsilon) r^\epsilon \mu^{\frac{n-k+1}{2}} \prod_{i=2}^k \left(\sum_{q' \in \mathcal{CH}_i(r) + y_i N_i} \left\langle \frac{d(\pi_{N_i} q, q')}{r} \right\rangle^{-(2N-n^2)} \left\| \left\langle \frac{x - c(q')}{r} \right\rangle^N \chi_{q'} \mathcal{F}_i^{-1} f_i \right\|^2_{L^2(\mathcal{H}_i + y_i N_i)} \right)^{\frac12} \\ &\quad \cdot \sum_\alpha \left(\sum_{q' \in \mathcal{C}_{1,\alpha}(r) + y_1 N_1} \left\langle \frac{d(\tilde{\pi}_{1,\alpha}(q), q')}{r} \right\rangle^{-(2N-2k)} \left\| \left\langle \frac{x^\alpha - c(q')}{r} \right\rangle^N \chi_{q'} \mathcal{F}^{-1}_{1,\alpha} f_{1,\alpha} \right\|^2_{L^2(\mathcal{H}_{1,\alpha} + y_1 N_1)} \right)^{\frac12}. \end{aligned}$$

It suffices to prove this estimate for $y = 0$, in which case it is very similar to (6-15). Except that, for the $f_{1,\alpha}$ terms we do not use strips, but cubes. This should be a reason for concern, as the use of strips was necessary to keep the localization of the $f_{1,\alpha}$ at scale μ intact throughout the induction process. However, given that $\mu \approx r^{-1}$, the multiplication with $\chi_{q'}$ alters the localization by a factor of $r^{-1} \approx \mu$. A similar argument to the one used in the proof of (6-15) gives the above estimate.

Next we average the above estimate over the values of $(y_1, \dots, y_k)$ satisfying $|y_i - c_i(q)| \leq r$ (keeping $y_i = c_i(q),\ i \geq k+1$) to obtain

$$\begin{aligned}\Big\| \prod_{i=1}^k \mathcal{E}_i f_i \Big\|_{L^1(q)} &\lesssim C(\epsilon) r^{\epsilon} (r^{-1})^{\frac{n-k+1}{2}} r^{-\frac{k}{2}} \\ &\quad \cdot \prod_{i=1}^k \Big(\int_{|y_i| \leq r} \sum_{q' \in \mathcal{CH}_i(r) + y_i N_i} \Big\langle \frac{d(\pi_{N_i} q, q')}{r} \Big\rangle^{-N} \Big\| \Big\langle \frac{x - c(q')}{r} \Big\rangle^{N} \chi_{q'} \mathcal{E}_i f_i \Big\|^2_{L^2(\mathcal{H}_i + y_i N_i)} \Big)^{\frac{1}{2}} \\ &\lesssim C(\epsilon) r^{\epsilon} r^{-\frac{n+1}{2}} \prod_{i=1}^k \| \tilde{\chi}_q \mathcal{E}_i f_i \|_{L^2},\end{aligned}$$

which finishes the proof. □

7. A new multilinear estimate

In this section we address (4-10), the last supporting detail in the proof of Proposition 4.1. As described in Section 5, we are given k smooth hypersurfaces $S_i = \Sigma_i(U_i)$ with smooth parametrizations Σ_i obeying (5-1). These hypersurfaces can be thought of as living in the frequency space and generate the operators $\mathcal{E}_i$. In addition we are given another smooth surface S of dimension $n-k+1$, which should be thought of as living in the physical space, with the following properties:

(P1) S is uniformly transversal to $N_1(\zeta_1), \dots, N_k(\zeta_k)$ for all choices $\zeta_i \in S_i$. There exists $\nu > 0$ such that, for any $\zeta_i \in S_i,\ i = 1, \dots, k$, for any $y \in S$ and for any orthonormal basis $v_{k+1}, \dots, v_{n+1}$ of $T_y S$, the following holds true:

$$\mathrm{vol}(N_1(\zeta_1), \dots, N_k(\zeta_k), v_{k+1}, \dots, v_{n+1}) \geq \nu.$$

(P2) There exists $\nu > 0$ such that for any $P_1, P_2 \in S$, for any $\zeta_1 \in S_1$, for any $\zeta_i \in S_i,\ \zeta_j \in S_j$, $2 \leq i < j \leq k$ and for any $\alpha_i, \alpha_j \in \mathbb{R}$, the following holds true:

$$\mathrm{vol}(\overrightarrow{P_1 P_2}, N_1(\zeta_1), \overrightarrow{v}) \geq \nu |\overrightarrow{P_1 P_2}| \cdot |\overrightarrow{v}|, \tag{7-1}$$

where $\overrightarrow{v} = \alpha_i N_i(\zeta_i) - \alpha_j N_j(\zeta_j)$.

As already mentioned in Section 5, in this section we make no curvature assumptions on S_i. However, we note that property (P2) follows from curvature properties similar to those used in Theorem 1.2; in other words the curvature properties have been encoded in the structure of S.

Given $r > 0$, we recall that $\mathcal{C}(r)$ is the set of unit cubes in $\mathbb{R}^{n+1}$ with centers in the lattice $r\mathbb{Z}^{n+1}$. With S as above and for each $q \in \mathcal{C}(r)$, we define

$$S(q) = q + S \cap B(0, r^2).$$

Here $S \cap B(0, r^2)$ should be understood as follows: we cut the surface S at scale $\approx r^2$, and whether this is performed in a ball or cube, centered at the origin or somewhere else, is unimportant. The reason for doing this comes from the use of wave packets and their scales.

More generally, given a subset $A \subset \mathcal{C}(r)$, we define

$$S(A) = \bigcup_{q \in A} S(q).$$

The main result of this section is the following theorem.

Theorem 7.1. *Assume that S_i, $i = 1, \dots, k$, and S are as above. Then for any $\epsilon > 0$, there is $C(\epsilon)$ such that the following holds true*:

$$\left(\sum_{q \in \mathcal{C}(r) \cap B(0,r^2)} \left(\|\mathcal{E}_1 f_1\|_{L^2(S(q))} \prod_{i=2}^{k} \|\mathcal{E}_i f_i\|_{L^2(q)} \right)^{\frac{2}{k-1}} \right)^{\frac{k-1}{2}} \leq C(\epsilon) r^{\frac{k}{2}} r^{\epsilon} \prod_{i=1}^{k} \|f_i\|_{L^2(U_i)}. \tag{7-2}$$

The above result has a multilinear flavor to it. The factor $r^{k/2}$ has to appear because we consider the mass of $\mathcal{E}_i f_i$ in neighborhoods of size r of hypersurfaces across which we would have good energy estimates; see the proof of the theorem for details. Otherwise (7-2) is similar to a multilinear restriction estimate, see (6-2) (with $\mu = 1$), except that now, one of the objects $\mathcal{E}_1 f_1$ is measured in a more complex fashion.

The complexity of this estimate does not stem from the fact that we collect energy from various spatial regions; indeed if v_i are arbitrary vectors, then an estimate of the type

$$\left\| \prod_{i=1}^{k} \|\mathcal{E}_i f_i\|_{L^2(q+v_i)} \right\|_{l_q^{2/(k-1)}} \lesssim r^{\frac{k}{2}} r^{\epsilon} \prod_{i=1}^{k} \|f_i\|_{L^2(U_i)}$$

is similar to the one with $v_i = 0$, which in turn is similar to (6-2) (with $\mu = 1$).

The complexity has to do with the factor $\| \|\mathcal{E}_1 f_1\|_{L^2(U_i)} \|_{L^2(q+v_1)}$ being replaced with $\|\mathcal{E}_1 f_1\|_{L^2(q+S)}$, that is, with collecting the energy of $\mathcal{E}_1 f_1$ not only across a cube $q + v_1$, but across a thickened surface $q + S$. It is the dimensionality of the surface S being $n - k + 1$ versus that of v_1 being 0 that changes the character of the estimate. Another feature to point out is the following: the classical multilinear estimate improves under certain localization properties of the support of the interacting functions (see the μ-factor in (6-2)); (7-2) does not improve under such localizations.

In [Bejenaru 2017b] we provided an energy estimate of the type

$$\|\mathcal{E}_1 f_1\|_{L^2(\widetilde{S}+q)} \lesssim r^{\frac{1}{2}} \|f_1\|_{L^2(U_1)}, \tag{7-3}$$

where $\widetilde{S}$ is a hypersurface (i.e., of codimension 1) that is transversal to the propagation directions of $\mathcal{E}_1 f_1$, that is, to any $N_1(\zeta_1)$ with $\zeta_1 \in S_1$.

The starting point of the arguments in this section is a refinement of (7-3) in terms of wave packets. We use the result of Lemma 3.4 with $c = 1$ and $R = 4r^2$ to obtain the wave packet decomposition

$$\mathcal{E}_1 f_1 = \sum_{T_1 \in \mathcal{T}_1} \phi_{T_1}.$$

We also recall the definition of $c_N(T_1)$ from (3-12) and their property (3-13).

Lemma 7.2. *There exists $N \in \mathbb{N}$ such that for any $q \in \mathcal{C}(r)$ centered inside $B(0, r^2)$, the following holds true*:

$$\|\mathcal{E}_1 f_1\|_{L^2(S(q))} \lesssim r^{\frac{1}{2}} \left(\sum_{T_1 \in \mathcal{T}_1} \left\langle \frac{d(T_1, S(q))}{r} \right\rangle^{-N} c_{2N}(T_1)^2 \right)^{\frac{1}{2}}. \tag{7-4}$$

Proof. For every $q \in \mathcal{C}(r)$ whose center lies inside $B(0, r^2)$, it is obvious that $S(q) \subset B(0, 4r^2)$. We write

$$\begin{aligned}
\|\mathcal{E}_1 f_1\|^2_{L^2(S(q))} &\lesssim \sum_{q' \cap S(q) \neq \varnothing} \|\mathcal{E}_1 f_1\|^2_{L^2(q')} \\
&\lesssim \sum_{q' \cap S(q) \neq \varnothing} \sum_{T_1 \in \mathcal{T}_1} \|\phi_{T_1}\|^2_{L^2(q')} = \sum_{T_1 \in \mathcal{T}_1} \sum_{q' \cap S(q) \neq \varnothing} \|\phi_{T_1}\|^2_{L^2(q')} \\
&\lesssim \sum_{T_1 \in \mathcal{T}_1} \sum_{q' \cap S(q) \neq \varnothing} \tilde{\chi}_{T_1}(x_{q'}, t_{q'})^N \tilde{\chi}_{T_1}(x_{q'}, t_{q'})^{-N} \|\phi_{T_1}\|^2_{L^2(q')} \\
&\lesssim \sum_{T_1 \in \mathcal{T}_1} \left\langle \frac{d(T_1, S(q))}{r} \right\rangle^{-N} r c_{2N}(T_1)^2.
\end{aligned}$$

In justifying the last line we used the following two estimates: the obvious estimate

$$\mathrm{supp}_{q' \cap S(q) \neq \varnothing} \tilde{\chi}_{T_1}(x_{q'}, t_{q'})^N \lesssim \left\langle \frac{d(T_1, S(q))}{r} \right\rangle^{-N},$$

as well as

$$\sum_{q' \cap S(q) \neq \varnothing} \tilde{\chi}_{T_1}(x_{q'}, t_{q'})^{-N} \|\phi_{T_1}\|^2_{L^2(q')} \lesssim r c_{2N}(T_1)^2. \tag{7-5}$$

We justify (7-5) as follows: From (3-12) we obtain

$$\mathrm{supp}_{q'} \tilde{\chi}_T(x_{q'}, t_{q'})^{-2N} \|\phi_{T_1}\|^2_{L^2(q')} \lesssim r c_{2N}(T_1)^2.$$

Then (7-5) follows from

$$\sum_{q' \cap S(q) \neq \varnothing} \tilde{\chi}_{T_1}(x_{q'}, t_{q'})^N \lesssim 1.$$

But, choosing N large enough, this is a direct consequence of the transversality between T_1 and $S(q)$. □

Proof of Theorem 7.1. As we already explained in the proof of Theorem 6.1, it suffices to establish the result under the following assumption: given some $0 < \delta \ll 1$, the diameter of U_i is $\leq \delta$.

The setup is also similar to the one in Section 6. We pick $\zeta_i^0 \in \Sigma_i$, let $N_i = N_i(\zeta_i^0)$ be the normal to Σ_i and let $\mathcal{H}_i$ be the transversal hyperplane passing through the origin with normal $N_i(\zeta_i^0)$. Using a smooth change of coordinates, we can assume that $U_i \subset \mathcal{H}_i$ and that

$$\mathcal{E}_i f_i = \int_{U_i} e^{i(x'\xi' + x_i \varphi_i(\xi'))} f_i(\xi')\, d\xi', \tag{7-6}$$

where $x = (x_i, x')$, x_i is the coordinate in the direction of N_i and x' are the coordinates in the directions from $\mathcal{H}_i$. Since the diameter of U_i is $\lesssim \delta$, it follows that $|\nabla\varphi_i(x) - \nabla\varphi_i(y)| \lesssim \delta$ for any $x, y \in U_i$. Using the normals N_i we construct all entities described in Section 6A as well as the margin of a function $f:\mathcal{H}_i\to\mathbb{C}$ as defined in (6-10).

We complete the system of vectors by choosing $N_{k+1},\dots,N_{n+1}$ such that (6-1) is satisfied. We then construct the lattice $\mathcal{L}:=\{z_1N_1+\dots+z_{n+1}N_{n+1}:(z_1,\dots,z_{n+1})\in\mathbb{Z}^{n+1}\}$ and for a given $r>0$ we let $\mathcal{C}(r)$ be the set of parallelepipeds of size r in $\mathbb{R}^{n+1}$ relative to the lattice $\mathcal{L}$. The lattice $\mathcal{L}$ and the set of parallelepipeds $\mathcal{C}(r)$ obtained this way are "oblique". Thus this set is different than the one claimed in (7-2), which is built on the standard orthonormal basis. However, passing from results in terms of an oblique lattice to the ones in the standard basis is easy: it can be done by changing coordinates, or by direct estimates.

Our induction aims at quantifying the behavior of $A(R)$ defined below.

Definition 7.3. Given $r\le R\lesssim r^2$ we define $A(R)$ to be the best constant for which the estimate

$$\Bigl(\sum_{q\in\mathcal{C}(r)\cap Q}\Bigl(\|\mathcal{E}_1f_1\|_{L^2(S(q))}\prod_{i=2}^k\|\mathcal{E}_if_i\|_{L^2(q)}\Bigr)^{\frac{2}{k-1}}\Bigr)^{\frac{k-1}{2}} \le A(R)r^{\frac{k}{2}}\Bigl(\sum_{T_1\in\mathcal{T}_1}\Bigl\langle\frac{d(T_1,S(Q))}{R}\Bigr\rangle^{-N}c_{2N}(T_1)^2\Bigr)^{\frac12}\prod_{i=2}^k\|f_i\|_{L^2(U_i)} \tag{7-7}$$

holds true for all parallelepipeds $Q\in\mathcal{C}(R)$, and all $f_i\in L^2(U_i)$, $i=2,\dots,k$, obeying the margin requirement

$$\operatorname{margin}^i(f_i)\ge\delta-R^{-\frac12}. \tag{7-8}$$

Note that in the above definition $f_1\in L^2(U_1)$ is a fixed element, unlike $f_2,\dots,f_k$, which can vary.

The above estimate holds true for $R=r$ with $A(r)\approx 1$; indeed, it follows from (7-4) and the obvious estimate

$$\|\mathcal{E}_if_i\|_{L^2(q)}\lesssim r^{\frac12}\|f_i\|_{L^2(U_i)}.$$

Note also that we limit the range of the argument to $R\lesssim r^2$. This is important so as to be able to use the wave packet described above.

Next, we proceed with the induction step. We provide an estimate inside any cube $\boldsymbol{Q}\in\mathcal{C}(\delta^{-1}R)$ based on prior information on estimates inside cubes $Q\in\mathcal{C}(R)\cap\boldsymbol{Q}$. Without restricting the generality of the argument, we assume that $\boldsymbol{Q}$ is centered at the origin and recall that each $Q\in\mathcal{C}(R)\cap\boldsymbol{Q}$ has its center in $\mathcal{L}$. When such a Q is projected using π_{N_i} onto $\mathcal{H}_i$ one obtains $\pi_{N_i}Q\in\mathcal{CH}_i$. We let Q_0 be the cube in $\mathcal{C}(R)$ centered at the origin.

We strengthen the induction hypothesis (7-7) to

$$\Bigl(\sum_{q\in\mathcal{C}(r)\cap Q}\Bigl(\|\mathcal{E}_1f_1\|_{L^2(S(q))}\prod_{i=2}^k\|\mathcal{E}_if_i\|_{L^2(q)}\Bigr)^{\frac{2}{k-1}}\Bigr)^{\frac{k-1}{2}} \lesssim A(R)\Bigl(\sum_{T_1\in\mathcal{T}_1}\Bigl\langle\frac{d(T_1,S(Q))}{R}\Bigr\rangle^{-N}c_{2N}(T_1)^2\Bigr)^{\frac12} \cdot\prod_{i=2}^k\Bigl(\sum_{Q'\in\mathcal{CH}_i(R)}\Bigl\langle\frac{d(\pi_{N_i}Q,Q')}{R}\Bigr\rangle^{-(N-2n^2)}\Bigl\|\Bigl\langle\frac{x-c(Q')}{R}\Bigr\rangle^{N}\chi_{Q'}\mathcal{F}_i^{-1}f_i\Bigr\|_{L^2}^2\Bigr)^{\frac12}. \tag{7-9}$$

The improvement for the terms $\mathcal{E}_i f_i$ with $i \geq 2$ is standard by now, see (6-15) and the references to [Bejenaru 2017a]. Using (7-9) we conclude the argument using the discrete Loomis–Whitney inequality in (6-9). For $i = 2, \dots, n$, we define the functions $g_i : \mathcal{L}(\mathcal{H}_i) \to \mathbb{R}$ by

$$g_i(\boldsymbol{j}) = \left(\sum_{Q' \in \mathcal{CH}_i(R)} \left\langle \frac{d(Q(\boldsymbol{j}), Q')}{R} \right\rangle^{-(N-2n^2)} \left\| \left\langle \frac{x' - c(Q')}{R} \right\rangle^{N} \chi_{q'} \mathcal{F}_i^{-1} f_i \right\|_{L^2}^2 \right)^{\frac{1}{2}}, \quad \boldsymbol{j} \in \mathcal{L}(\mathcal{H}_i),$$

where we recall that $Q(\boldsymbol{j}) \in \mathcal{CH}_i(R)$ is the cube centered at $R\boldsymbol{j}$.

From (6-8), it is easy to see that for N large enough (depending only on n), $g_i \in l^2(\mathbb{Z}^n)$, $i = 2, \dots, k$, with

$$\|g_i\|_{l^2(\mathcal{L}(\mathcal{H}_i))} \lesssim \|f_i\|_{L^2}.$$

For $i = 1$ and $\boldsymbol{j} \in \mathcal{L}$, we recall that $Q(\boldsymbol{j}) = Q_0 + R\boldsymbol{j} \in \mathcal{C}(R)$ is the cube centered at $R\boldsymbol{j}$, and define

$$g_1(\boldsymbol{j}) = \left(\sum_{T_1 \in \mathcal{T}_1} \left\langle \frac{d(T_1, S(Q(\boldsymbol{j})))}{R} \right\rangle^{-N} c_{2N}(T_1)^2 \right)^{\frac{1}{2}}.$$

We claim that $g_1 \in l^\infty_{j_1, j_{k+1}, \dots, j_{n+1}} l^2_{j_2, \dots, j_k}(D)$, where $D = \{\boldsymbol{j} \in \mathcal{L} : \|\boldsymbol{j}\|_{l^\infty} \leq \delta^{-1}\}$ is the domain of interest, together with the estimate

$$\|g_1\|_{l^\infty_{j_1, j_{k+1}, \dots, j_{n+1}} l^2_{j_2, \dots, j_k}(D)} \lesssim \left(\sum_{T_1 \in \mathcal{T}_1} \left\langle \frac{d(T_1, S(\boldsymbol{Q}))}{\delta^{-1} R} \right\rangle^{-N} c_{2N}(T_1)^2 \right)^{\frac{1}{2}}. \tag{7-10}$$

We assume for a moment (7-10) to be true. Using (7-9), we invoke (6-9) and the above estimates on g_i to obtain

$$\begin{aligned}
&\left(\sum_{q \in \mathcal{C}(r) \cap \boldsymbol{Q}} \left(\|\mathcal{E}_1 f_1\|_{L^2(S(q))} \prod_{i=2}^{k} \|\mathcal{E}_i f_i\|_{L^2(q)} \right)^{\frac{2}{k-1}} \right)^{\frac{k-1}{2}} \\
&\qquad = \left(\sum_{Q \in \mathcal{C}(R) \cap \boldsymbol{Q}} \sum_{q \in \mathcal{C}(r) \cap Q} \left(\|\mathcal{E}_1 f_1\|_{L^2(S(q))} \prod_{i=2}^{k} \|\mathcal{E}_i f_i\|_{L^2(q)} \right)^{\frac{2}{k-1}} \right)^{\frac{k-1}{2}} \\
&\qquad \lesssim A(R) r^{\frac{k}{2}} \left(\sum_{T_1 \in \mathcal{T}_1} \left\langle \frac{d(T_1, S(\boldsymbol{Q}))}{R} \right\rangle^{-N} c_{2N}(T_1)^2 \right)^{\frac{1}{2}} \prod_{i=2}^{k} \|f_i\|_{L^2(U_i)}.
\end{aligned}$$

Thus we establish that

$$A(\delta^{-1} R) \lesssim A(R).$$

This implies (7-2) in a standard fashion, see for instance [Bejenaru 2017a], and concludes our proof.

We owe an argument for the claim (7-10). We fix $j_1, j_{k+1}, \dots, j_{n+1}$ with

$$\max\{|j_1|, |j_{k+1}|, \dots, |j_{n+1}|\} \leq \delta^{-1}.$$

Then (7-10) is a consequence of the following estimate:

$$\sum_{\substack{j_2, \dots, j_k \\ |j_l| \leq \delta^{-1}}} \sum_{T_1 \in \mathcal{T}_1} \left\langle \frac{d(T_1, S(Q(\boldsymbol{j})))}{R} \right\rangle^{-N} c_{2N}(T_1)^2 \lesssim \sum_{T_1 \in \mathcal{T}_1} \left\langle \frac{d(T_1, S(\delta^{-1} Q_0))}{\delta^{-1} R} \right\rangle^{-N} c_{2N}(T_1)^2,$$

which in turn follows from

$$\sum_{\substack{j_2,\dots,j_k\\ |j_l|\le\delta^{-1}}} \left\langle \frac{d(T_1, S(Q(\boldsymbol{j})))}{R} \right\rangle^{-N} \lesssim \left\langle \frac{d(T_1, S(\delta^{-1}Q_0))}{\delta^{-1}R} \right\rangle^{-N}. \tag{7-11}$$

The estimate (7-11) is easily derived from the following claim: given any $d \in \mathbb{N}$, there are $\lesssim d^{k-1}$ values of $\boldsymbol{j} \in D$ such that $d(T_1, S(Q(\boldsymbol{j}))) \le dR$.

Thus, the last thing we need to do is to establish the claim above. Let $\boldsymbol{j}_1, \boldsymbol{j}_2 \in D$ be such that $d(T_1, S(Q(\boldsymbol{j}_1))), d(T_1, S(Q(\boldsymbol{j}_2))) \approx dR$. Let L_1 be the center line of T_1; it has direction $N_1 = N_1(\zeta_1)$ for some $\zeta_1 \in S_1$. Using the fact that $R \ge r$, we conclude that there are points $P_1, P_2 \in T_1$, $\widetilde{P}_1 \in S(Q(\boldsymbol{j}_1))$, $\widetilde{P}_2 \in S(Q(\boldsymbol{j}_2))$ with the following properties:

- $P_1, P_2 \in L_1$.
- $\widetilde{P}_1 \in S + R\boldsymbol{j}_1$, $\widetilde{P}_2 \in S + R\boldsymbol{j}_2$.
- $d(P_1, \widetilde{P}_1), d(P_2, \widetilde{P}_2) \lesssim dR$.

From the vector identity

$$\overrightarrow{\widetilde{P}_1\widetilde{P}_2} = \overrightarrow{\widetilde{P}_1 P_1} + \overrightarrow{P_1 P_2} + \overrightarrow{P_2\widetilde{P}_2}$$

and the above properties, we obtain

$$|\overrightarrow{\widetilde{P}_1\widetilde{P}_2} - \overrightarrow{P_1 P_2}| \lesssim dR.$$

On the other hand, $\widetilde{P}_1 = Q_1 + R\boldsymbol{j}_1$, $\widetilde{P}_2 = Q_2 + R\boldsymbol{j}_2$ for some $Q_1, Q_2 \in S$, and therefore

$$\overrightarrow{\widetilde{P}_1\widetilde{P}_2} - \overrightarrow{P_1 P_2} = \overrightarrow{Q_1 Q_2} + R(\boldsymbol{j}_1 - \boldsymbol{j}_2) + \alpha N_1$$

for some $\alpha \in \mathbb{R}$. Now we bring in the transversality considerations, see (7-1), to conclude that

$$dR \gtrsim |\overrightarrow{Q_1 Q_2} + R(\boldsymbol{j}_1 - \boldsymbol{j}_2) + \alpha N_1| \gtrsim R|\boldsymbol{j}_1 - \boldsymbol{j}_2|;$$

here we use the structure of the lattice $\mathcal{L}$ to infer that $\boldsymbol{j}_1 - \boldsymbol{j}_2 = \alpha_i N_i(\zeta_i) - \alpha_j N_j(\zeta_j)$ for some $i, j \in \{2, \dots, k\}$ and some $\alpha_i, \alpha_j \in \mathbb{R}$.

Thus $d \gtrsim |\boldsymbol{j}_1 - \boldsymbol{j}_2|$, and, as a consequence, there are about d^{k-1} values of $\boldsymbol{j}$ with the property that $d(T_1, S(Q(\boldsymbol{j}))) \le dR$. □

Acknowledgements

Part of this work was supported by a grant from the Simons Foundation (#359929, Ioan Bejenaru). Part of this work was supported by the National Science Foundation under grant DMS-160444.

References

[Bejenaru 2017a] I. Bejenaru, "The multilinear restriction estimate: a short proof and a refinement", *Math. Res. Lett.* **24**:6 (2017), 1585–1603. MR Zbl

[Bejenaru 2017b] I. Bejenaru, "Optimal bilinear restriction estimates for general hypersurfaces and the role of the shape operator", *Int. Math. Res. Not.* **2017**:23 (2017), 7109–7147. MR

[Bejenaru 2017c] I. Bejenaru, "The optimal trilinear restriction estimate for a class of hypersurfaces with curvature", *Adv. Math.* **307** (2017), 1151–1183. MR Zbl

[Bennett et al. 2006] J. Bennett, A. Carbery, and T. Tao, "On the multilinear restriction and Kakeya conjectures", *Acta Math.* **196**:2 (2006), 261–302. MR Zbl

[Bourgain 1995] J. Bourgain, "Estimates for cone multipliers", pp. 41–60 in *Geometric aspects of functional analysis* (Israel, 1992–1994), edited by J. Lindenstrauss and V. Milman, Oper. Theory Adv. Appl. **77**, Birkhäuser, Basel, 1995. MR Zbl

[Bourgain 2013] J. Bourgain, "On the Schrödinger maximal function in higher dimension", *Tr. Mat. Inst. Steklova* **280** (2013), 53–66. MR Zbl

[Bourgain and Demeter 2015] J. Bourgain and C. Demeter, "The proof of the l^2 decoupling conjecture", *Ann. of Math.* (2) **182**:1 (2015), 351–389. MR Zbl

[Bourgain and Guth 2011] J. Bourgain and L. Guth, "Bounds on oscillatory integral operators based on multilinear estimates", *Geom. Funct. Anal.* **21**:6 (2011), 1239–1295. MR Zbl

[Bourgain et al. 2016] J. Bourgain, C. Demeter, and L. Guth, "Proof of the main conjecture in Vinogradov's mean value theorem for degrees higher than three", *Ann. of Math.* (2) **184**:2 (2016), 633–682. MR Zbl

[Du et al. 2017] X. Du, L. Guth, and X. Li, "A sharp Schrödinger maximal estimate in $\mathbb{R}^2$", *Ann. of Math.* (2) **186**:2 (2017), 607–640. MR Zbl

[Foschi and Klainerman 2000] D. Foschi and S. Klainerman, "Bilinear space-time estimates for homogeneous wave equations", *Ann. Sci. École Norm. Sup.* (4) **33**:2 (2000), 211–274. MR Zbl

[Guth 2010] L. Guth, "The endpoint case of the Bennett–Carbery–Tao multilinear Kakeya conjecture", *Acta Math.* **205**:2 (2010), 263–286. MR Zbl

[Guth 2016a] L. Guth, "A restriction estimate using polynomial partitioning", *J. Amer. Math. Soc.* **29**:2 (2016), 371–413. MR Zbl

[Guth 2016b] L. Guth, "A restriction estimate using polynomial partitioning", *J. Amer. Math. Soc.* **29**:2 (2016), 371–413. MR Zbl

[Kenig and Merle 2006] C. E. Kenig and F. Merle, "Global well-posedness, scattering and blow-up for the energy-critical, focusing, non-linear Schrödinger equation in the radial case", *Invent. Math.* **166**:3 (2006), 645–675. MR Zbl

[Lee 2003] S. Lee, "Endpoint estimates for the circular maximal function", *Proc. Amer. Math. Soc.* **131**:5 (2003), 1433–1442. MR Zbl

[Lee 2006] S. Lee, "Bilinear restriction estimates for surfaces with curvatures of different signs", *Trans. Amer. Math. Soc.* **358**:8 (2006), 3511–3533. MR Zbl

[Lee and Vargas 2010] S. Lee and A. Vargas, "Restriction estimates for some surfaces with vanishing curvatures", *J. Funct. Anal.* **258**:9 (2010), 2884–2909. MR Zbl

[Merle and Vega 1998] F. Merle and L. Vega, "Compactness at blow-up time for L^2 solutions of the critical nonlinear Schrödinger equation in 2D", *Internat. Math. Res. Notices* **1998**:8 (1998), 399–425. MR Zbl

[Stein 1993] E. M. Stein, *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*, Princeton Mathematical Series **43**, Princeton University Press, 1993. MR Zbl

[Tao 2001] T. Tao, "Endpoint bilinear restriction theorems for the cone, and some sharp null form estimates", *Math. Z.* **238**:2 (2001), 215–268. MR Zbl

[Tao 2003] T. Tao, "A sharp bilinear restrictions estimate for paraboloids", *Geom. Funct. Anal.* **13**:6 (2003), 1359–1384. MR Zbl

[Tao 2006] T. Tao, *Nonlinear dispersive equations: local and global analysis*, CBMS Regional Conference Series in Mathematics **106**, Amer. Math. Soc., Providence, RI, 2006. MR Zbl

[Tao and Vargas 2000a] T. Tao and A. Vargas, "A bilinear approach to cone multipliers, I: Restriction estimates", *Geom. Funct. Anal.* **10**:1 (2000), 185–215. MR Zbl

[Tao and Vargas 2000b] T. Tao and A. Vargas, "A bilinear approach to cone multipliers, II: Applications", *Geom. Funct. Anal.* **10**:1 (2000), 216–258. MR Zbl

[Wolff 2001] T. Wolff, "A sharp bilinear cone restriction estimate", *Ann. of Math.* (2) **153**:3 (2001), 661–698. MR Zbl

Received 28 Feb 2018. Revised 25 May 2018. Accepted 29 Jun 2018.

IOAN BEJENARU: ibejenaru@math.ucsd.edu
Department of Mathematics, University of California, San Diego, La Jolla, CA, United States

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at msp.org/apde.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in APDE are usually in English, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with details about how your graphics were generated.

White space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

ANALYSIS & PDE

Volume 12 No. 4 2019